

**ESTIMATION OF  
VARIANCE-COVARIANCE  
COMPONENTS FOR GEODETIC  
OBSERVATIONS AND  
IMPLICATIONS ON  
DEFORMATION TREND  
ANALYSIS**

**J. GRODECKI**

**October 1997**



**TECHNICAL REPORT  
NO. 186**

## PREFACE

In order to make our extensive series of technical reports more readily available, we have scanned the old master copies and produced electronic versions in Portable Document Format. The quality of the images varies depending on the quality of the originals. The images have not been converted to searchable text.

**ESTIMATION OF VARIANCE-COVARIANCE  
COMPONENTS FOR GEODETIC  
OBSERVATIONS AND IMPLICATIONS ON  
DEFORMATION TREND ANALYSIS**

Jacek Grodecki

Department of Geodesy and Geomatics Engineering  
University of New Brunswick  
P.O. Box 4400  
Fredericton, N.B.  
Canada  
E3B 5A3

August 1997

© Jacek Grodecki, 1997

## PREFACE

This technical report is a reproduction of a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Geodesy and Geomatics Engineering, April 1997. The research was supervised by Dr. Adam Chrzanowski, and funding was provided by the Natural Sciences and Engineering Research Council of Canada, Natural Resources Canada, and the University of New Brunswick.

As with any copyrighted material, permission to reprint or quote extensively from this report must be received from the author. The citation to this work should appear as follows:

Grodecki, J. (1997). *Estimation of Variance-Covariance Components for Geodetic Observations and Implications on Deformation Trend Analysis*. Ph.D. dissertation, Department of Geodesy and Geomatics Engineering Technical Report No. 186, University of New Brunswick, Fredericton, New Brunswick, Canada, 243 pp.

## **ABSTRACT**

The statistical methods for the estimation of variance-covariance components for unbalanced data are reviewed in this thesis. Computational aspects of the presented methods are compared and their applicability to geodetic data is discussed.

Prior information about the unknown variance components is introduced within the framework of the Generalized Maximum Likelihood (GML) methodology. The inverted gamma prior is used to introduce prior information about the variance components, and the noninformative prior is used when no prior information is available. The Fisher scoring method is applied to the resulting posterior probability density functions and the estimating equations are derived. Prior information is also introduced by means of the weighted constraints on the unknown variance-covariance components in the dispersion-mean model. The estimating equations of the dispersion-mean model with weighted constraints are derived, and conditions for equivalence between the dispersion-mean model with weighted constraints and the GML estimation are formulated.

The effect of neglecting the errors of the estimated variance-covariance components, in the least squares adjustment, on the covariance matrix of the estimated location parameters is discussed.

The influence of different aspects of the estimation of variance components on the results of spatial deformation trend analysis is investigated, based on practical examples. These

include the amount of prior information, the choice of the method of estimation, and the choice of the error model.

An efficient numerical algorithm for detecting influential observations, in terms of their influence on the results of variance components estimation, is developed and tested on geodetic survey data.

All numerical procedures and algorithms developed in the thesis are demonstrated on practical examples.

## TABLE OF CONTENTS

<b>ABSTRACT</b> .....	ii
<b>TABLE OF CONTENTS</b> .....	iv
<b>LIST OF FIGURES</b> .....	vii
<b>LIST OF TABLES</b> .....	ix
<b>ACKNOWLEDGMENTS</b> .....	xi
<b>1. INTRODUCTION</b> .....	1
<b>1.1 Motivation and Objectives of the Research</b> .....	1
<b>1.2 Previous Studies</b> .....	6
<b>1.3 Methodology</b> .....	10
<b>1.4 Organization of Contents of the Dissertation</b> .....	11
<b>1.5 Summary of the Contributions</b> .....	14
<b>2. REVIEW OF STATISTICAL METHODS FOR ESTIMATION OF VARIANCE-COVARIANCE COMPONENTS</b> .....	16
<b>2.1 Estimation of Variance-Covariance Components - General Formulation</b> .....	16
2.1.1 ANOVA Methods .....	17
2.1.2 Distribution-Based Methods .....	23
2.1.3 Criteria-Based Methods .....	25
<b>2.2 Maximum Likelihood Methods</b> .....	25
2.2.1 Maximum Likelihood (ML).....	26
2.2.2 Restricted Maximum Likelihood (REML).....	35
<b>2.3 Bayesian Methods</b> .....	44
2.3.1 Bayesian Approach to the Estimation of Variance-Covariance Components: General Formulation .....	44
2.3.2 Bayes Estimation .....	51
2.3.3 Interval Estimation and Hypothesis Testing .....	53
2.3.4 Generalized Maximum Likelihood (GML) Estimation .....	55
<b>2.4 Criteria-Based Methods</b> .....	58
2.4.1 Minimum Norm Quadratic Estimation (MINQE) .....	59
2.4.2 Best Quadratic Estimation (BQE) .....	66
2.4.3 Minimum Variance Quadratic Unbiased Estimation (MIVQUE).....	69
2.4.4 Almost Unbiased Estimation (AUE).....	70
<b>2.5 Comparison of the Reviewed Methods</b> .....	72

<b>3. INTRODUCTION OF PRIOR INFORMATION: GML ESTIMATION OF VARIANCE COMPONENTS AND THE DISPERSION-MEAN MODEL WITH WEIGHTED CONSTRAINTS .....</b>	<b>75</b>
<b>3.1 GML Estimation with Inverted Gamma Prior .....</b>	<b>78</b>
3.1.1 Derivation of the Vector of Scores .....	79
3.1.2 Derivation of the Information Matrix .....	81
3.1.3 Application of the Fisher Scoring Method .....	84
<b>3.2 GML Estimation with Noninformative Prior .....</b>	<b>87</b>
3.2.1 Derivation of the Vector of Scores .....	88
3.2.2 Derivation of the Information Matrix .....	101
3.2.3 Application of the Fisher Scoring Method .....	111
<b>3.3 Application of the Principle of Mixed Estimation to the     Dispersion-Mean Model .....</b>	<b>113</b>
3.3.1 Principle of the Dispersion-Mean Model .....	113
3.3.2 The Dispersion-Mean Model with Weighted Constraints .....	116
3.3.3 Conditions for Equivalence with GML Estimation .....	118
<b>4. INFLUENCE OF ESTIMATION OF VARIANCE-COVARIANCE COMPONENTS ON THE COVARIANCE MATRIX OF THE ESTIMATED LOCATION PARAMETERS .....</b>	<b>122</b>
<b>5. ANALYSIS OF HOMOGENEITY OF THE DATA: DETECTION OF INFLUENTIAL OBSERVATIONS .....</b>	<b>126</b>
<b>5.1 Influence Function .....</b>	<b>128</b>
<b>5.2 Computational Aspects of Detection of Influential Observations .....</b>	<b>131</b>
5.2.1 The Effect of Omitting an Observation on the Inverse of the Matrix of Normal Equations .....	133
5.2.2 The Effect of Omitting an Observation on the $R_0$ Matrix .....	134
<b>5.3 Influence of a Single Observation on the Asymptotic Covariance     Matrix of the Vector of the Estimated Variance-Covariance     Components .....</b>	<b>139</b>
<b>6. INFLUENCE OF ESTIMATION OF VARIANCE-COVARIANCE COMPONENTS ON SPATIAL DEFORMATION TREND ANALYSIS .....</b>	<b>141</b>
<b>6.1 Principle of Spatial Deformation Trend Analysis .....</b>	<b>141</b>



<b>6.2 Influence of Estimation of Variance-Covariance Components on the Iterative Weighted Similarity Transformation .....</b>	<b>149</b>
<b>7. NUMERICAL EXAMPLES .....</b>	<b>153</b>
<b>7.1 Estimation of Variance Components for the Mactaquac Monitoring Network: Campaigns of 1986, 1989, 1991 and 1993 .....</b>	<b>154</b>
7.1.1 Conformity with the Hartley-Rao Conditions .....	158
7.1.2 Discussion of the Results .....	165
<b>7.2 Influence of the Estimation of Variance Components on the Spatial Deformation Trend Analysis Process .....</b>	<b>172</b>
7.2.1 Influence of the Error Model .....	172
7.2.2 Influence of the Method of Estimation .....	177
7.2.3 Influence of the Amount of Prior Information .....	180
7.2.4 Discussion of the Results .....	184
<b>7.3 Analysis of Homogeneity of the Data: Mactaquac Monitoring Network (Campaign of 1989) .....</b>	<b>185</b>
<b>8. SUMMARY, CONCLUSIONS AND RECOMMENDATIONS .....</b>	<b>190</b>
<b>8.1 Recommended Strategy for Estimation of Variance or Variance-Covariance Components for Geodetic Observations.....</b>	<b>190</b>
<b>8.2 Summary of Results and Conclusions .....</b>	<b>191</b>
<b>8.3 Recommendations for Related Future Research .....</b>	<b>194</b>
<b>REFERENCES .....</b>	<b>197</b>
 <b>APPENDICES</b>	
<b>APPENDIX A NOTATION .....</b>	<b>205</b>
<b>APPENDIX B ASYMPTOTIC PROPERTIES OF GML ESTIMATORS .....</b>	<b>207</b>
<b>APPENDIX C RESULTS OF THE ESTIMATION OF VARIANCE COMPONENTS .....</b>	<b>217</b>
<b>C.1 REML Estimation .....</b>	<b>217</b>
<b>C.2 GML Estimation with Inverted Gamma Prior .....</b>	<b>219</b>
<b>C.3 GML Estimation with Noninformative Prior .....</b>	<b>223</b>
<b>C.4 The Dispersion-Mean Model with Weighted Constraints .....</b>	<b>225</b>
<b>C.5 Comparison of the Estimated Variance Components .....</b>	<b>231</b>
<b>C.6 Ratios of the Estimated Variance Components to their Asymptotic Standard Deviations .....</b>	<b>236</b>

## LIST OF FIGURES

<b>Figure 7.1</b>	<b>Mactaquac Horizontal Monitoring Network .....</b>	<b>155</b>
<b>Figure 7.2</b>	<b>Displacement Field and Error Ellipses (95%): 2 Components Model Used for the 1993 Data, REML Estimation Method .....</b>	<b>174</b>
<b>Figure 7.3</b>	<b>Displacement Field and Error Ellipses (95%): 3 Components Model Used for the 1993 Data, REML Estimation Method .....</b>	<b>175</b>
<b>Figure 7.4</b>	<b>Displacement Field and Error Ellipses (95%): 2 Components Model versus 3 Components Model .....</b>	<b>176</b>
<b>Figure 7.5</b>	<b>Displacement Field and Error Ellipses (95%): 3 Components Model Used for the 1993 Data; GML with Inverted Gamma Prior Estimation Method, Ratio of the Standard Deviations of the Priors to the Prior Values = 2 .....</b>	<b>178</b>
<b>Figure 7.6</b>	<b>Displacement Field and Error Ellipses (95%): REML versus GML with Inverted Gamma Prior .....</b>	<b>179</b>
<b>Figure 7.7</b>	<b>Displacement Field and Error Ellipses (95%): 3 Components Model Used for the 1993 Data, Dispersion - Mean Model with Weighted Constraints, Ratio of the Standard Deviations of the Priors to the Prior Values = 2 .....</b>	<b>181</b>
<b>Figure 7.8</b>	<b>Displacement Field and Error Ellipses (95%): 3 Components Model Used for the 1993 Data, Dispersion - Mean Model with Weighted Constraints, Ratio of the Standard Deviations of the Priors to the Prior Values = 16 .....</b>	<b>182</b>
<b>Figure 7.9</b>	<b>Displacement Field and Error Ellipses (95%): Ratio of 2 versus Ratio of 16 .....</b>	<b>183</b>
<b>Figure C.1</b>	<b>Campaign of 1986 (<math>a_S</math> component) .....</b>	<b>231</b>
<b>Figure C.2</b>	<b>Campaign of 1986 (<math>b_S</math> component) .....</b>	<b>232</b>
<b>Figure C.3</b>	<b>Campaign of 1989 (<math>a_S</math> component) .....</b>	<b>232</b>
<b>Figure C.4</b>	<b>Campaign of 1989 (<math>b_S</math> component) .....</b>	<b>233</b>
<b>Figure C.5</b>	<b>Campaign of 1991 (<math>a_S</math> component) .....</b>	<b>233</b>
<b>Figure C.6</b>	<b>Campaign of 1991 (<math>b_S</math> component) .....</b>	<b>234</b>
<b>Figure C.7</b>	<b>Campaign of 1991 (<math>a_D</math> component) .....</b>	<b>234</b>
<b>Figure C.8</b>	<b>Campaign of 1993 (<math>a_S</math> component) .....</b>	<b>235</b>
<b>Figure C.9</b>	<b>Campaign of 1993 (<math>b_S</math> component) .....</b>	<b>235</b>
<b>Figure C.10</b>	<b>Campaign of 1993 (<math>a_D</math> component) .....</b>	<b>236</b>
<b>Figure C.11</b>	<b>Campaign of 1986 (<math>a_S^2</math> component) .....</b>	<b>237</b>
<b>Figure C.12</b>	<b>Campaign of 1986 (<math>b_S^2</math> component) .....</b>	<b>237</b>
<b>Figure C.13</b>	<b>Campaign of 1989 (<math>a_S^2</math> component) .....</b>	<b>238</b>
<b>Figure C.14</b>	<b>Campaign of 1989 (<math>b_S^2</math> component) .....</b>	<b>238</b>

<b>Figure C.15 Campaign of 1991 (<math>a_S^2</math> component)</b> .....	239
<b>Figure C.16 Campaign of 1991 (<math>b_S^2</math> component)</b> .....	239
<b>Figure C.17 Campaign of 1991 (<math>a_D^2</math> component)</b> .....	240
<b>Figure C.18 Campaign of 1993 (<math>a_S^2</math> component)</b> .....	240
<b>Figure C.19 Campaign of 1993 (<math>b_S^2</math> component)</b> .....	241
<b>Figure C.20 Campaign of 1993 (<math>a_D^2</math> component)</b> .....	241

## LIST OF TABLES

<b>Table 7.1</b>	<b>Summary of the Mactaquac monitoring network: campaigns of 1986, 1989, 1991 and 1993 .....</b>	<b>154</b>
<b>Table 7.2</b>	<b>The empirical, a priori, values of variance components .....</b>	<b>157</b>
<b>Table 7.3</b>	<b>Rejected residuals (tau-max criterion used) .....</b>	<b>186</b>
<b>Table 7.4</b>	<b>Influence of formal outliers on the <math>b_S^2</math> term .....</b>	<b>188</b>
<b>Table C.1</b>	<b>REML estimation: 1986 campaign, 2 parameter model .....</b>	<b>217</b>
<b>Table C.2</b>	<b>REML estimation: 1989 campaign, 2 parameter model .....</b>	<b>217</b>
<b>Table C.3</b>	<b>REML estimation: 1989 campaign, 1 parameter models .....</b>	<b>218</b>
<b>Table C.4</b>	<b>REML estimation: 1991 campaign, 3 parameter model .....</b>	<b>218</b>
<b>Table C.5</b>	<b>REML estimation: 1993 campaign, 3 parameter model .....</b>	<b>218</b>
<b>Table C.6</b>	<b>REML estimation: 1993 campaign, 2 parameter models .....</b>	<b>218</b>
<b>Table C.7</b>	<b>GML estimation with inverted gamma prior: 1986 campaign, 2 parameter model .....</b>	<b>219</b>
<b>Table C.8</b>	<b>GML estimation with inverted gamma prior: 1989 campaign, 2 parameter model .....</b>	<b>220</b>
<b>Table C.9</b>	<b>GML estimation with inverted gamma prior: 1991 campaign, 3 parameter model .....</b>	<b>221</b>
<b>Table C.10</b>	<b>GML estimation with inverted gamma prior: 1993 campaign, 3 parameter model .....</b>	<b>222</b>
<b>Table C.11</b>	<b>GML estimation with inverted gamma prior: 1993 campaign, 2 parameter model .....</b>	<b>223</b>
<b>Table C.12</b>	<b>GML estimation with noninformative prior: 1986 campaign, 2 parameter model .....</b>	<b>224</b>
<b>Table C.13</b>	<b>GML estimation with noninformative prior: 1989 campaign, 2 parameter model .....</b>	<b>224</b>
<b>Table C.14</b>	<b>GML estimation with noninformative prior: 1989 campaign, 1 parameter models .....</b>	<b>224</b>
<b>Table C.15</b>	<b>GML estimation with noninformative prior: 1991 campaign, 3 parameter model .....</b>	<b>224</b>
<b>Table C.16</b>	<b>GML estimation with noninformative prior: 1993 campaign, 3 parameter model .....</b>	<b>225</b>
<b>Table C.17</b>	<b>GML estimation with noninformative prior: 1993 campaign, 2 parameter models .....</b>	<b>225</b>
<b>Table C.18</b>	<b>Dispersion-mean model with weighted constraints: 1986 campaign, 2 parameter model .....</b>	<b>226</b>
<b>Table C.19</b>	<b>Dispersion-mean model with weighted constraints: 1989 campaign, 2 parameter model .....</b>	<b>227</b>
<b>Table C.20</b>	<b>Dispersion-mean model with weighted constraints: 1991 campaign, 3 parameter model .....</b>	<b>228</b>

<b>Table C.21 Dispersion-mean model with weighted constraints: 1993 campaign, 3 parameter model .....</b>	<b>229</b>
<b>Table C.22 Dispersion-mean model with weighted constraints: 1993 campaign, 2 parameter models .....</b>	<b>230</b>

## **ACKNOWLEDGMENTS**

I wish to express my deep gratitude and warmest thanks to my supervisor Dr. Adam Chrzanowski for his continuous guidance and support during the course of my graduate studies.

The financial support from the University of New Brunswick, the Natural Science and Engineering Council of Canada, and the Natural Resources of Canada is gratefully acknowledged.

I would also like to express my appreciation to the members of the Examining Board for their valued comments and suggestions. A special thanks goes to Dr. William R. Knight of the Department of Mathematics and Statistics at UNB for his many hours spent on reviewing the draft manuscript, his constructive criticism and generous assistance. Appreciation is also due to Dr. Yong-Qi Chen of the Department of Land Surveying and Geoinformatics at the Hong Kong Polytechnical University, to Dr. R. Dayal Gupta of the Department of Mathematics, Statistics and Computer Science at UNB Saint John and to Dr. James M. Secord of the Department of Geodesy and Geomatics Engineering at UNB for their valued recommendations and comments on the draft manuscript.

I owe a great deal to my friends: Mark Caissy for helping me to master the “MINQE” software, Bruce Chaplin for many hours spent on proofreading the draft manuscript, and Vladimir Argeseanu for his valuable comments and advice.

I dedicate this thesis to my wife Barbara for her immeasurable support and continuous encouragement.

# CHAPTER 1

## INTRODUCTION

This chapter outlines motivation and objectives of the research, provides a comprehensive summary of previous studies, describes the methodology of the research, summarizes the contents of the dissertation, and lists the major contributions of the research.

### **1.1 Motivation and Objectives of the Research**

The results of the least squares adjustment of geodetic observations are directly influenced by the weights assigned to the observations. Therefore, proper estimation of the variance-covariance components, and hence the weights of observations, and utilization of all available information is of the utmost importance.

The process of estimating the variance-covariance components can be divided into the following three steps:

- (1) formulation of the variance-covariance (error) model,
- (2) estimation of the unknown variance-covariance components, and
- (3) statistical testing of the estimated components.

This thesis investigates aspects of step (2), namely, efficient methods of introducing prior information within the framework of Bayesian methodology, influence of the estimation of variance-covariance components on the covariance matrix of the estimated location



parameters, and the problem of detecting influential observations. The formulation of the optimal variance-covariance model, step (1), is not investigated. As far as step (3) is concerned, it is assumed throughout the thesis that the asymptotic covariance matrix of the estimated variance or variance-covariance components adequately approximates the real covariance matrix, and that the sample is of a sufficiently large size. The problem of finding an exact (non-asymptotic) probability distribution function of the estimated variance or variance-covariance components is not considered in the thesis.

The estimation of variance-covariance components for geodetic observations differs, in many respects, from the estimation of variance-covariance components for other types of data.

Firstly, geodetic data are inherently unbalanced. The implication of this fact is that, in general, only methods suitable for the estimation of variance-covariance components for unbalanced data can be applied to geodetic observations. Most of these methods are computationally equivalent, or similar to a large degree, even though they are derived from different underlying principles. The first objective of this research is thus to provide a comprehensive summary of these methods, with an emphasis on their similarities, and the merits and demerits of each of them in the context of their application to geodetic data.

Secondly, the estimation of variance-covariance components for geodetic data generally suffers from small redundancy problems, which, for the most part, is not the case with other types of data. This has two important implications.

The first implication is that, in order to strengthen the estimation process, the available prior information concerning the unknown variance-covariance components should be utilized. The sources of prior information could be the manufacturer's specifications of measurement accuracy, or analyses of physical sources of errors of observations (see, e.g., [Chrzanowski, 1974 and 1975] and [Blachut et al., 1979]). Bayesian estimation methods make provision for the introduction of prior information. Bayes estimation and interval estimation (see, e.g., [Koch, 1987], [Koch, 1988], and [Koch, 1990]) require the application of numerical integration techniques and thus, even for a moderate number of variance or variance-covariance components, are computationally impractical [Harville, 1977]. The classical maximum likelihood methods, on the other hand, do not make any provision for introduction of prior information.

There exists, however, an interesting alternative. Harville [1977] proposes to apply maximum likelihood methodology to the posterior probability density function, being the product of the likelihood function of the data and the prior probability density function of the variance components. Harville [1977, p. 336] calls the method the pseudo-Bayesian procedure and states that, "The pseudo-Bayesian procedure that estimates  $\theta$  by maximizing the expression ... [for the natural logarithm of the posterior probability density

function with noninformative prior] would seem to be worth investigating". Harville [1977] does not, however, investigate the matter any further and does not propose any numerical algorithms. Koch [1990] describes the Generalized Maximum Likelihood estimation method, the principle of which, as in Harville's pseudo-Bayesian approach, is to maximize the posterior probability density function. The resulting estimate of the vector of variance-covariance components is called the Generalized Maximum Likelihood (GML) or the Maximum A Posteriori (MAP) estimate. Koch [1990] does not go beyond describing the principle of GML estimation, and provides no numerical algorithms for finding the solution.

As the GML method promises to offer a viable and computationally efficient alternative to the Bayes and interval estimation methods, one of the main goals of this research is to derive computational algorithms (estimating equations) for finding the GML estimates of the vector of variance or variance-covariance components under miscellaneous types of prior probability density functions.

The second implication of small redundancy of geodetic data on the estimation of variance-covariance components is that the undetected outliers can have a very large influence on the estimated components. The definition of outliers is broadened here to include such observations that do not conform to the postulated error model, i.e., the outliers in the variance-covariance components space are more important than the outliers in the location parameters space. The degree of conformity of observations to the

postulated variance-covariance components model can be quantified by means of their influence on the vector of estimated variance-covariance components. The problem of detecting influential observations for the case of estimating the *a posteriori* variance factor is discussed in [Chatterjee and Hadi, 1988]. As the problem of detecting influential observations for the more general case of estimating the variance or variance-covariance components does not appear to have been solved yet, this forms the next goal of this investigation.

It is an obvious fact that the weights of observations, computed through the process of estimation of variance-covariance components, exert direct influence on the results of the least squares adjustment. Both the vector of adjusted coordinates (location parameters), and its covariance matrix, are affected by the choice of the variance-covariance components estimation method, the amount of prior information, and the choice of the error model. Resulting changes in the vector of estimated location parameters and its covariance matrix affect, in turn, the outcome of spatial deformation trend analysis, and hence influence the deformation modelling process. The ensuing objective of this research is thus to investigate the magnitude of these influences, based on numerical examples.

When the weights of observations are computed from the estimated variance-covariance components, the implication is that they are stochastic quantities estimated with a certain accuracy. The usual approach is, however, to treat the computed weight matrix of observations as fixed and errorless [Searle et al., 1992]. Whereas the vector of estimated

location parameters, resulting from the least squares adjustment process, remains largely unaffected by this assumption [Kackar and Harville, 1981], the covariance matrix of the estimated vector of location parameters is, in general, underestimated [Searle et al., 1992]. The problem of underestimation of the covariance matrix of the vector of estimated location parameters has, in general, not been solved yet. Searle et al. [1992, p. 320] state that, “There are ... a number of classical methods available to correct this variance underestimation problem, but unfortunately they can be difficult to implement.” Searle et al. [1992] propose the use of either the Taylor series expansion method, or the bootstrap method to correct the problem. Searle et al. [1992] do not, however, provide any solutions and moreover, state [1992, p. 320] that, “Both these methods may lead to implementation difficulties: the Taylor series may be an extremely involved calculation, while the bootstrap may require enormous computing power.”

The last objective of this research is therefore to investigate the problem of underestimation of the covariance matrix of the vector of estimated location parameters.

## **1.2 Previous Studies**

As far as the methods of estimation of variance-covariance components are concerned, the existing literature is exhaustive. A very extensive bibliography of this subject is given in [Sahai, 1979] and [Sahai et al., 1985], while a comprehensive review of the existing methods can be found in [Khuri and Sahai, 1985] and [Searle et al., 1992].

The literature relating particularly to geodetic applications of estimation of variance-covariance components is also vast, as depicted by the following examples:

1. The Analysis of Variance (ANOVA) method is applied to geodetic levelling networks by Kelly [1991].
2. Grafarend et al. [1980] and Grafarend [1984] propose the Helmert-type method of estimation of variance-covariance components for geodetic observations. The method constitutes an extension to the method proposed by Helmert [1924].
3. Grafarend and Kleusberg [1980] use the Helmert-type method to estimate variance components for gyrotheodolite observations. The method is also applied to satellite ranging data by Sahin et al. [1992].
4. Conditions for equivalence between the Helmert-type method and the Minimum Norm Quadratic Unbiased Estimation (MINQUE) technique are discussed in [Kelm, 1978], [Welsch, 1978] and [Chen, 1983].
5. Grafarend [1978] discusses the Best Quadratic Unbiased Estimation (BQUE) and MINQUE methods, and formulates conditions for their equivalence.
6. Application of the Best Invariant Quadratic Unbiased Estimation (BIQUE) method to electronic distance measurements on calibration lines is shown in [Koch, 1981]. Application of BIQUE to the heterogeneous deformation observations is discussed by Schaffrin [1981] and Caspary [1987].
7. The problem of estimation of variance-covariance components for repeated observations is analyzed in [Schaffrin, 1983].

8. The BQUE method is extended to the condition adjustment with constraints by Yu [1992].
9. Application of MINQUE to analysis of deformations is discussed by Chen [1983]. Chen and Chrzanowski [1985] give an example of the Minimum Norm Quadratic Estimation (MINQE) of weights for geodetic levelling networks. Chen et al. [1990a] apply the MINQE principle to triangulation and GPS networks.
10. An extension of MINQUE theory to the case with singular covariance matrix is given in [Sjöberg, 1985]
11. An extension of MINQUE to the condition adjustment is shown in [Sjöberg, 1983a], and to the condition adjustment with unknowns in [Sjöberg, 1983b].
12. The maximum likelihood methods are introduced to the geodetic community by Kubik [1967 and 1970].
13. The Marginal Maximum Likelihood (MML), also called the Restricted Maximum Likelihood (REML), method is compared with BIQUE by Koch [1986].
14. Application of the Maximum Likelihood (ML) method to repeated EDM observations can be found in [Sjöberg, 1980].
15. The REML method and its comparison with the Helmert-type estimation method and the Iterated Almost Unbiased Estimation (IAUE) procedure of Förstner [1979a and 1979b] is presented in [Ou, 1989].
16. Application of the IAUE algorithm to analysis of levelling data for the final adjustment of the North American Vertical Datum of 1988 is presented by Lucas et al. [1985].

17. Bayesian methods of estimation of variance-covariance components, and in particular the Bayes and interval estimation methods, are introduced for geodetic applications by Koch [1987, 1988 and 1990].
18. Ou [1991] discusses the Bayes estimation with the, so called, approximative likelihood function, maximization of which results in an estimator equivalent to IAUE.

The principle of the Generalized Maximum Likelihood (GML) estimation of variance-covariance components, for geodetic data, is described in [Koch, 1990]. Furthermore, GML estimation with noninformative prior, for any type of data, is discussed in the paper of Harville [1977], which reviews the maximum likelihood methods of estimation of variance components. In both cases, however, only the principle of GML estimation is described and the GML estimating equations are not given.

There exist a large number of publications concerned with detection of influential observations (see, e.g., [Belsley et al., 1980], [Chatterjee and Hadi, 1988] and [Fox, 1991]) for the case of least squares regression analysis. The concept of influence functions, introduced by Hampel [1968 and 1974] is discussed, for example, in [Huber, 1981], [Hampel et al., 1986] and [Chatterjee and Hadi, 1988]. However, the problem of detecting influential observations in the estimation of the variance-covariance components, with the exception of the trivial case of one variance component being the *a posteriori* variance factor [Chatterjee and Hadi, 1988], has not yet been treated.



The problem of the omission of errors of the estimated variance-covariance components in the least-squares estimation process, and thus the problem of underestimation of the covariance matrix of the estimated location parameters, is discussed in [Searle et al., 1992]. Searle et al. [1992] propose two potential methods for solving the underestimation problem: the bootstrap method and the Taylor series expansion method. No solution is given, and the practicality of the proposed methods is questioned.

### **1.3 Methodology**

The problem of computationally efficient introduction of prior information into the estimation of variance components, within the framework of Bayesian methodology, is approached by means of the Generalized Maximum Likelihood (GML) estimation method. Two cases are considered. In the first case, the inverted gamma prior (see, e.g., [Koch, 1990]) is used for introducing the prior information about the variance components. In order to find a solution to the resulting optimization problem, the Fisher scoring method is applied, and the GML estimating equations are derived. In the second case, the noninformative prior is used: to express the fact that there exists no prior information about the unknown variance components. As in the first case, the Fisher scoring method is applied to the resulting optimization problem and the GML estimating equations are derived. In addition, the weighted constraints on the unknown variance-covariance components are introduced into the dispersion-mean model, first proposed by Pukelsheim [1974]. The least squares procedure is applied to the resulting dispersion-mean model

with weighted constraints, and conditions for equivalence with the GML method are formulated.

The problem of analysis of influential observations is approached from the computational perspective. The possibility of finding an efficient algorithm which quantifies the influence of each observation, without the need to repeat the whole variance-covariance components estimation process, is investigated. Investigations are limited to the case involving the variance components only.

The degree to which the covariance matrix of the estimated location parameters is affected by the estimation of variance-covariance components is investigated for the case when the difference between the true and the estimated covariance matrices of observations is a fixed quantity.

The task of comparing the influence of different factors of the variance components estimation process on the results of the spatial deformation trend analysis process is approached by analyzing the effect of such factors on the transformed displacement components and their confidence regions.

#### **1.4 Organization of Contents of the Dissertation**

Chapter 2 provides a comprehensive summary of existing statistical methods for the estimation of variance-covariance components for unbalanced data. The presented

methods are grouped according to their estimation principles, and their computational similarities and differences are analyzed.

Chapter 3 deals with the problem of introduction of prior information. The GML methodology is applied for two choices of the priors, namely, the inverted gamma prior and the noninformative prior. For both cases, the estimating equations are derived by applying the Fisher scoring method to the resulting posterior probability density functions. In addition, prior information is introduced into the dispersion-mean model by means of weighted constraints on the unknown variance-covariance components. The estimating equations, resulting from the application of the least squares principle, are subsequently derived. Conditions for equivalence between this method and the GML estimation are formulated.

In Chapter 4, the influence of the estimation of variance-covariance components on the covariance matrix of the estimated location parameters is investigated.

Chapter 5 contains a description of measures of the influence of observations on the vector of estimated variance-covariance components and its asymptotic covariance matrix. An algorithm for efficient computation of the influence of a single observation on the estimated variance components and their asymptotic covariance matrix is derived.

In Chapter 6 the process of spatial deformation trend analysis is described. The influence of the estimation of variance-covariance components on the results of spatial deformation trend analysis is also discussed.

Chapter 7 implements the proposed methods and procedures. A horizontal monitoring network at the Mactaquac hydroelectric generating station near Fredericton, New Brunswick, [Chrzanowski and Secord, 1987 and 1990], [Chrzanowski et al., 1989] is used to illustrate and test the algorithms and methods derived in this thesis. The GML estimating equations, derived in Chapter 3 under noninformative and inverted gamma priors, are applied to periodic observations of the Mactaquac network. Estimating equations resulting from application of the least squares principle to the dispersion-mean model with weighted constraints, derived in Chapter 3, are also applied to the same data. The results are compared with the results of REML estimation. In addition, the influence of the choice of the estimation method, the amount of prior information, and the choice of the error model, on the results of spatial deformation trend analysis process is presented - based on an analysis of the 1991 and 1993 campaigns of the Mactaquac monitoring network. Finally, the procedure for detecting influential observations, proposed in Chapter 5, is applied to the 1989 Mactaquac data set.

Chapter 8 summarizes results of the research, outlines the strategy advocated by the author, and presents conclusions and recommendations.

## **1.5 Summary of the Contributions**

The contributions of this research are:

1. A comprehensive summary of existing methods of estimation of variance-covariance components for unbalanced data, with emphasis on their computational similarities and differences.
2. Discussion of the effects of the estimation of variance-covariance components on the covariance matrix of the estimated location parameters, in the least squares adjustment.
3. Derivation of the Generalized Maximum Likelihood estimating equations with the inverted gamma prior.
4. Derivation of the Generalized Maximum Likelihood estimating equations with the noninformative prior.
5. Development of the dispersion-mean model with weighted constraints and derivation of the estimating equations.
6. Development of an efficient computational algorithm for analyzing the influence of a single observation on the estimated variance components and their covariance matrix.
7. Analysis of the contribution of various aspects of the estimation of variance components - such as the method of estimation, the amount of prior information and the error model - on the process of spatial deformation trend analysis.
8. Analysis of asymptotic properties of the GML estimators.

9. Verification of posterior normality of the posterior probability density function with either the inverted gamma or the noninformative prior, for the Gauss-Markov model with the unknown variance factor.

## CHAPTER 2

### REVIEW OF STATISTICAL METHODS FOR ESTIMATION OF VARIANCE-COVARIANCE COMPONENTS

This chapter concentrates on statistical methods of estimation of variance-covariance components for unbalanced data. As a result, the ANOVA (Analysis of Variance) methods will get limited attention since they are, in general, not suited to estimation of variance-covariance components in unbalanced cases. Emphasis will therefore be placed on:

- (1) maximum likelihood estimation methods, such as ML (Maximum Likelihood) and REML (Restricted Maximum Likelihood),
- (2) Bayesian methods, such as GML (Generalized Maximum Likelihood), the Bayes estimation, and the interval estimation, and
- (3) criteria-based estimation methods, such as MINQE (Minimum Norm Quadratic Estimation), BQE (Best Quadratic Estimation), MIVQUE (Minimum Variance Quadratic Unbiased Estimation) and AUE (Almost Unbiased Estimation) methods.

#### **2.1 Estimation of the Variance-Covariance Components - General Formulation**

Given an extended Gauss-Markov model [Kleffe, 1977], [Schaffrin, 1983], [Rao and Kleffe, 1988], [Oswald, 1992], also called a general mixed model:

$$\mathbf{l} = \mathbf{Ax} + \mathbf{v}, \quad \mathbf{v} \sim (\mathbf{0}, \mathbf{C}_\theta = \theta_1 \mathbf{T}_1 + \theta_2 \mathbf{T}_2 + \dots + \theta_r \mathbf{T}_r), \quad \theta \in \Theta, \quad (2.1)$$

where:  $\mathbf{l}$  is a vector of  $m$  observations,

$\mathbf{A}$  is a given first order design matrix,

$\mathbf{x}$  is a vector of  $u$  unknown location parameters,

$\mathbf{v}$  is a vector of  $m$  unobservable random errors,

$\mathbf{T}_1, \dots, \mathbf{T}_r$  are given, linearly independent, symmetric, matrices,

$\theta = (\theta_1, \dots, \theta_r)$  is a vector of  $r$  unknown variance-covariance components, and

$\Theta$  is the subset of the  $r$ -dimensional Euclidean space such that  $\mathbf{C}_\theta$  is a positive definite matrix,

the problem at hand is to find an estimate of the vector of variance-covariance components:

$$\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_r). \quad (2.2)$$

Generally, there are three distinct ways to get the solution:

### ***2.1.1 ANOVA Methods***

The ANOVA methods are the methods that work by equating a quadratic form of the observations to its expectation, and solving that equation ; they are:

(1) ANOVA methods for balanced data,



(2) general ANOVA methods for unbalanced data, including among others the Henderson's methods I, II and III (see, e.g., [Searle et al., 1992]), and the Helmert-type estimation method [Grafarend et al., 1980].

Even though these methods are inherently heuristic in nature, they have some desirable properties, e.g., unbiasedness, or in the case of balanced data, the minimum variance property.

For the balanced data, the quadratic forms are the sums of squares, and the resulting estimators have the properties of minimum variance and unbiasedness [Hultquist, 1985]. The required computations are always straightforward and simple. The ANOVA estimators are best quadratic unbiased (BQUE), and under the assumption that the observations are normally distributed, the estimators are best unbiased (BUE). Searle [1987] points out that under the assumption of normality of the data, the ANOVA estimators are numerically identical to (1) the Restricted Maximum Likelihood (REML) estimators, (2) the Iterated Best Invariant Quadratic Unbiased Estimators (Iterated BIQUE), and (3) the Iterated Minimum Norm Quadratic Estimators, under the conditions of unbiasedness and invariance (IMINQE(U,I)).

For the unbalanced data, the quadratic forms of observations are no longer simple sums of squares, and the estimators no longer have the property of minimum variance. From all the general ANOVA methods for unbalanced data the Helmert-type estimation method,

which was introduced by Helmert in 1924, is especially popular among geodetic community. Its applications to various types of geodetic data can be found, among others, in [Welsch, 1978], [Grafarend et al., 1980], [Grafarend and Kleusberg, 1980], [Grafarend, 1984], and [Sahin et al., 1992]. Kelm [1978], Grafarend et al. [1980], and Chen [1983] show that the iterated Helmert-type estimation of the variance components is numerically equivalent to REML estimation, Iterated BIQUÉ, and IMINQE(U,I). It can be shown [Chen, 1983], however, that this is no longer the case when the iterated Helmert-type estimation is applied to the general mixed model, eqn. (2.1).

As reported by Searle [1987] and Searle et al. [1992], regardless of whether the data are balanced or unbalanced, ANOVA may produce negative estimates of variances. Furthermore, the distributional properties of ANOVA estimators - even under the assumption that the data are normally distributed - are generally not known.

Data are classified as being balanced by referring to properties of the following variance components model (see, e.g., [Hartley and Rao, 1967], [Searle and Henderson, 1979] or [Searle et al., 1992]):

$$\mathbf{l} = \mathbf{Ax} + \mathbf{U}_1\mathbf{b}_1 + \mathbf{U}_2\mathbf{b}_2 + \dots + \mathbf{U}_{r-1}\mathbf{b}_{r-1} + \mathbf{e}, \quad (2.3)$$

where:  $\mathbf{l}$  is a vector of  $m$  observations,

$\mathbf{A}$  is a given first order design matrix,

$\mathbf{x}$  is a vector of  $u$  unknown location parameters,

$\mathbf{U}_i$  is an  $m \times m_i$  matrix of known fixed numbers,  $m_i \leq m$ ,

$\mathbf{b}_i$  is a vector of  $m_i$  independent random variables from  $N(0, \sigma_i^2)$ ,  $m_i \leq m$ ,

$\mathbf{e}$  is a vector of  $m$  independent random variables from  $N(0, \sigma^2)$ , and

the random vectors  $\mathbf{b}_i$  and  $\mathbf{e}$  are mutually independent.

Since  $\mathbf{e}$  is a vector of random variables just like each  $\mathbf{b}_i$ , one can define  $\mathbf{U}_r = \mathbf{I}_m$ ,  $\mathbf{b}_r = \mathbf{e}$  and  $\sigma_r = \sigma$ . In such case the variance components model of eqn. (2.3) reads:

$$\mathbf{l} = \mathbf{A}\mathbf{x} + \sum_{i=1}^r \mathbf{U}_i \mathbf{b}_i, \quad \text{with} \quad \mathbf{C}_\theta = \sum_{i=1}^r \sigma_i^2 \mathbf{U}_i \mathbf{U}_i^T, \quad (2.4)$$

The variance components model given by eqn. (2.4) can be considered as being a special case of the general mixed model, eqn. (2.1), in which case:

$$\mathbf{v} = \sum_{i=1}^r \mathbf{U}_i \mathbf{b}_i, \quad (2.5)$$

$$\mathbf{T}_i = \mathbf{U}_i \mathbf{U}_i^T, \quad \text{for } i = 1, 2, \dots, r, \quad (2.6)$$

$$\theta_i = \sigma_i^2, \quad \text{for } i = 1, 2, \dots, r, \quad (2.7)$$

Searle and Henderson [1979] and Searle et al. [1992] formulate the variance components model for balanced data, valid for any multi-factored model. For a linear model of  $p - 1$  main effect factors this model is given as:

$$\mathbf{y} = \sum_{\mathbf{i}=0}^{\mathbf{1}_p} \left( \mathbf{1}_{n_p}^{i_p} \otimes \mathbf{1}_{n_{p-1}}^{i_{p-1}} \otimes \dots \otimes \mathbf{1}_{n_1}^{i_1} \right) \mathbf{a}_i, \quad (2.8)$$

where:  $\mathbf{y}$  is the vector of observations (equivalent to the vector  $\mathbf{l}$  in eqn. (2.4)),

$\mathbf{a}_i$  are the vectors of effects (corresponding either to the vector  $\mathbf{x}$  or vectors  $\mathbf{b}_i$  in eqn. (2.4)),

$\mathbf{1}_{n_j}$  is a summing vector of  $n_j$  elements equal to unity, with  $\mathbf{1}_{n_j}^0 = \mathbf{I}_{n_j}$ , and

$\mathbf{i} = [i_p, \dots, i_1]$  with  $i_j = 0$  or  $1$  for  $j = 1, 2, \dots, p$ .

There exist  $2^p$  possible terms in eqn. (2.8); however, not all of them have to be present in a given model. The dispersion matrix is in turn given by

$$\mathbf{C}_\theta = \sum_{\mathbf{i}=0}^{\mathbf{1}_p} \theta_i \left( \mathbf{J}_{n_p}^{i_p} \otimes \mathbf{J}_{n_{p-1}}^{i_{p-1}} \otimes \dots \otimes \mathbf{J}_{n_1}^{i_1} \right), \quad (2.9)$$

where:  $\theta_i = \text{var}(\mathbf{a}_i)$ , and

$\mathbf{J}_{n_j}$  is a square matrix of order  $n_j$  with every element unity, with  $\mathbf{J}_{n_j}^0 = \mathbf{I}_{n_j}$ .

Searle and Henderson [1979] and Searle et al. [1992] give examples of some linear models for balanced data formulated in terms of eqn. (2.8) and eqn. (2.9). For example, the 1-way classification model with  $n$  observations in each of  $m$  classes, which in classical notation is given by

$$y_{ij} = \mu + a_i + e_{ij}, \text{ with } i = 1, \dots, m, \text{ and } j = 1, \dots, n \quad (2.10)$$

where:  $\mu$  is a general mean,

$\mathbf{a}$  is the vector of effects,

$\mathbf{e}$  is the vector of residual errors, and

$$\text{var}(\mathbf{a}_i) = \sigma_a^2, \text{var}(\mathbf{e}_{ij}) = \sigma_e^2, \text{ and the vectors } \mathbf{e} \text{ and } \mathbf{a} \text{ are uncorrelated,}$$

is given by Searle et al. [1992] as:

$$\mathbf{y} = (\mathbf{1}_m \otimes \mathbf{1}_n)\mu + (\mathbf{I}_m \otimes \mathbf{1}_n)\mathbf{a} + \mathbf{e}, \quad (2.11)$$

with the dispersion matrix

$$\mathbf{C}_\theta = \sigma_a^2(\mathbf{I}_m \otimes \mathbf{J}_n) + \sigma_e^2(\mathbf{I}_m \otimes \mathbf{I}_n). \quad (2.12)$$

Taking  $p = 2$  in the variance components model for balanced data, given by eqn. (2.8) and eqn. (2.9), we get

$$\mathbf{y} = (\mathbf{1}_{n_2} \otimes \mathbf{1}_{n_1})\mathbf{a}_{11} + (\mathbf{1}_{n_2} \otimes \mathbf{I}_{n_1})\mathbf{a}_{10} + (\mathbf{I}_{n_2} \otimes \mathbf{1}_{n_1})\mathbf{a}_{01} + (\mathbf{I}_{n_2} \otimes \mathbf{I}_{n_1})\mathbf{a}_{00}, \quad (2.13)$$

with the dispersion matrix

$$\mathbf{C}_\theta = \theta_{11}(\mathbf{J}_{n_2} \otimes \mathbf{J}_{n_1}) + \theta_{10}(\mathbf{J}_{n_2} \otimes \mathbf{I}_{n_1}) + \theta_{01}(\mathbf{I}_{n_2} \otimes \mathbf{J}_{n_1}) + \theta_{00}(\mathbf{I}_{n_2} \otimes \mathbf{I}_{n_1}). \quad (2.14)$$

Assuming that  $\mathbf{a}_{11}$  is fixed and dropping  $\mathbf{a}_{10}$  from the model we arrive at the 1-way classification model of eqn. (2.11) and eqn. (2.12), where:  $\mathbf{a}_{11} = \mu$ ,  $\mathbf{a}_{01} = \mathbf{a}$ ,  $\mathbf{a}_{00} = \mathbf{e}$ ,

$$\theta_{01} = \sigma_a^2, \theta_{00} = \sigma_e^2, n_1 = n, \text{ and } n_2 = m.$$

If the data are unbalanced then they can no longer be expressed by the variance components model of eqn. (2.8) and eqn. (2.9). In general, geodetic data are unbalanced. Kelly [1991] gives levelling measurements as an example of unbalanced data: if the levelling measurements are grouped according to the levelling lines along which they are

made, then the uneven number of sections in each levelling line implies that the data are unbalanced.

Since, in general, geodetic data are unbalanced, the simplest and the most popular method of the estimation of variances - the ANOVA method for balanced data - cannot be applied in most of the cases. Therefore, only the methods for estimation of variance-covariance components for unbalanced data, such as the methods listed in Subsection 2.1.2 and Subsection 2.1.3 and the general ANOVA methods, are universally applicable to geodetic data.

### *2.1.2. Distribution-Based Methods*

The distribution-based methods approach the problem of estimating variance-covariance components by attributing a distribution to the data. Having defined the distribution function, the estimates are sought in the form of either the expected values, or the modal values. The family of maximum likelihood methods and the Bayesian methods belong to this category.

The family of maximum likelihood methods comprises the Maximum Likelihood (ML) and the Restricted Maximum Likelihood (REML) methods. The ML estimator of the vector of variance-covariance components,  $\theta$ , is defined (see, e.g., [Rao and Kleffe, 1988]) as such value,  $\hat{\theta}$ , that maximizes the likelihood function of the vector of observations:

$$L_M(\hat{\theta}) = \sup_{\theta} L_M, \text{ subject to } \theta \in \Theta. \quad (2.15)$$

As far as the estimation of variance-covariance components is concerned, because of the numerical complexity of the task, the observations are assumed to be normally distributed and as a result the likelihood function  $L_M$  is of the multivariate normal type. In the case of the REML estimation, the likelihood function of the projected vector of observations (onto the orthogonal complement of the space defined by the columns of the first order design matrix  $A$ ) is solved for maximum. Again, the observations are assumed to be normally distributed.

In the case of the Bayesian methods, either the so called Bayes rule is used to find the estimated variance-covariance components, or solution is sought in form of the mode of the posterior density function. The principle of the Bayes rule (see, e.g., [Koch, 1990]) is to minimize the posterior expected loss under a chosen loss function. The resulting estimator is called the Bayes estimator. If the loss function is quadratic, then the estimate obtained under the Bayes rule is identical to the expected value of the vector of variance-covariance components,  $\theta$ , computed with the posterior probability density function,  $p(\theta|\mathbf{l})$ . This is given by Koch [1990] as:

$$\hat{\theta}_B = \int_{\Theta} \theta p(\theta|\mathbf{l}) d\theta \quad \text{with } \Theta \text{ being the parameter space.} \quad (2.16)$$

If the maximum of the posterior density function is sought, then the resulting estimator is given by Koch [1990] as such value,  $\hat{\theta}$ , that satisfies:

$$p(\hat{\theta}|\mathbf{l}) = \sup_{\theta} p(\theta|\mathbf{l}), \quad (2.17)$$

and is called the Generalized Maximum Likelihood (GML) estimator.

### ***2.1.3. Criteria-Based Methods.***

The criteria-based methods approach the problem of estimating variance-covariance components by seeking a quadratic estimator (QE),  $\mathbf{I}^T \mathbf{M} \mathbf{l}$ , of a linear combination of the variance-covariance components,  $\mathbf{p}^T \theta$ , that will satisfy certain optimality criteria. Criteria such as the minimum norm (MINQE), the minimum variance (MIVQUE), or the minimum mean squared error (BQE) are utilized. The estimator is derived by minimizing some objective function (the norm, the variance, or the mean squared error) subject to certain constraints imposed on the estimator, such as translation invariance or unbiasedness.

## **2.2 Maximum Likelihood Methods**

The Maximum Likelihood (ML) method of estimation (see, e.g., [Scholz, 1985]) was introduced by Fisher in 1925. The method was first applied to the problem of estimating variance-covariance components by Hartley and Rao [1967]. The Restricted Maximum Likelihood (REML) method, also called the Marginal Maximum Likelihood (MML) method, is a variation of the original method of Maximum Likelihood, and was first applied in context of estimation of variance-covariance components by Patterson and Thompson [1971].



### 2.2.1 Maximum Likelihood (ML)

The maximum likelihood (ML) method requires that the probability density function of the vector of observations is specified. The principle of the method is to find the maximum of this probability density function. The ML method can be used for any type of the probability density function, but in the case of estimation of variance-covariance components, it is customary to assume the multivariate normal distribution for the vector of observations. Under this assumption, the probability density function, called the likelihood function, is a function of the vector of location parameters,  $\mathbf{x}$ , and the vector of variance-covariance components,  $\theta$ , and is given by Searle et al. [1992] as:

$$L_M = L_M(\mathbf{x}, \theta | \mathbf{l}) = p(\mathbf{l} | \mathbf{x}, \theta) = (2\pi)^{-\frac{m}{2}} |\mathbf{C}_\theta|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(\mathbf{l} - \mathbf{Ax})^T \mathbf{C}_\theta^{-1}(\mathbf{l} - \mathbf{Ax})\right]. \quad (2.18)$$

Maximization of the likelihood function is equivalent to maximization of the natural logarithm of the likelihood function (log-likelihood function):

$$l_M = \ln(L_M) = -\frac{m}{2} \ln(2\pi) - \frac{1}{2} \ln|\mathbf{C}_\theta| - \frac{1}{2}(\mathbf{l} - \mathbf{Ax})^T \mathbf{C}_\theta^{-1}(\mathbf{l} - \mathbf{Ax}). \quad (2.19)$$

The ML estimator of the vector of location parameters,  $\mathbf{x}$ , and the vector of variance-covariance components,  $\theta$ , is then defined as the value,  $\hat{\theta}$ , that maximizes the log-likelihood function, subject to the condition that the vector  $\theta$  belongs to the parameter space  $\Theta$  - defined by the requirement of positive definiteness of the covariance matrix  $\mathbf{C}_\theta$ , and is given by Rao and Kleffe [1988] as:

$$l_M(\hat{\mathbf{x}}, \hat{\boldsymbol{\theta}}) = \sup_{\mathbf{x}, \boldsymbol{\theta}} l_M, \quad \text{subject to } \boldsymbol{\theta} \in \Theta. \quad (2.20)$$

Ignoring the constraints on the parameter space,  $\Theta$ , the maximum of the log-likelihood function can be solved for by applying one of the numerical methods for the nonlinear, unconstrained optimization. The two methods that have been most widely used in the context of estimation of variance-covariance components are:

(1) The Newton-Raphson method (see, e.g., [Jennrich and Sampson, 1976], [Harville, 1977] and [Searle et al., 1992]), which gives the  $(i+1)$ -th iterate of the vector of location parameters,  $\mathbf{x}$ , and of the vector of variance-covariance components,  $\boldsymbol{\theta}$ , as:

$$\begin{aligned} \begin{bmatrix} \hat{\mathbf{x}}_{(i+1)} \\ \hat{\boldsymbol{\theta}}_{(i+1)} \end{bmatrix} &= \begin{bmatrix} \hat{\mathbf{x}}_{(i)} \\ \hat{\boldsymbol{\theta}}_{(i)} \end{bmatrix} - \begin{bmatrix} \frac{\partial^2 l_M}{\partial \mathbf{x} \partial \mathbf{x}} \Big|_{\mathbf{x}=\hat{\mathbf{x}}_{(i)}, \boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_{(i)}} & \frac{\partial^2 l_M}{\partial \mathbf{x} \partial \boldsymbol{\theta}} \Big|_{\mathbf{x}=\hat{\mathbf{x}}_{(i)}, \boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_{(i)}} \\ \frac{\partial^2 l_M}{\partial \boldsymbol{\theta} \partial \mathbf{x}} \Big|_{\mathbf{x}=\hat{\mathbf{x}}_{(i)}, \boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_{(i)}} & \frac{\partial^2 l_M}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}} \Big|_{\mathbf{x}=\hat{\mathbf{x}}_{(i)}, \boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_{(i)}} \end{bmatrix}^{-1} \begin{bmatrix} \frac{\partial l_M}{\partial \mathbf{x}} \Big|_{\mathbf{x}=\hat{\mathbf{x}}_{(i)}, \boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_{(i)}} \\ \frac{\partial l_M}{\partial \boldsymbol{\theta}} \Big|_{\mathbf{x}=\hat{\mathbf{x}}_{(i)}, \boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_{(i)}} \end{bmatrix} \\ &= \begin{bmatrix} \hat{\mathbf{x}}_{(i)} \\ \hat{\boldsymbol{\theta}}_{(i)} \end{bmatrix} - \begin{bmatrix} \hat{\mathbf{x}}_{(i)} \\ \hat{\boldsymbol{\theta}}_{(i)} \end{bmatrix}^{\mathbf{H}_M} \begin{bmatrix} \hat{\mathbf{x}}_{(i)} \\ \hat{\boldsymbol{\theta}}_{(i)} \end{bmatrix}^{-1} \mathbf{s}_M \begin{bmatrix} \hat{\mathbf{x}}_{(i)} \\ \hat{\boldsymbol{\theta}}_{(i)} \end{bmatrix}, \end{aligned} \quad (2.21)$$

where:

$$\mathbf{H}_M \begin{pmatrix} \mathbf{x} \\ \boldsymbol{\theta} \end{pmatrix} = \begin{bmatrix} -\mathbf{A}^T \mathbf{C}_\theta^{-1} \mathbf{A} & \left\{ \mathbf{r} \mathbf{A}^T \mathbf{C}_\theta^{-1} \mathbf{T}_k \mathbf{C}_\theta^{-1} (\mathbf{I} - \mathbf{A}\mathbf{x}) \right\}_{k=1}^r \\ \text{symmetric} & \left\{ \frac{1}{2} \text{tr}(\mathbf{C}_\theta^{-1} \mathbf{T}_k \mathbf{C}_\theta^{-1} \mathbf{T}_m) - (\mathbf{I} - \mathbf{A}\mathbf{x})^T \mathbf{C}_\theta^{-1} \mathbf{T}_k \mathbf{C}_\theta^{-1} \mathbf{T}_m \mathbf{C}_\theta^{-1} (\mathbf{I} - \mathbf{A}\mathbf{x}) \right\}_{k,m=1}^r \end{bmatrix}$$

$$= \begin{bmatrix} \mathbf{H}_{M_{\mathbf{x}\mathbf{x}}} & \mathbf{H}_{M_{\mathbf{x}\theta}} \\ \mathbf{H}_{M_{\theta\mathbf{x}}} & \mathbf{H}_{M_{\theta\theta}} \end{bmatrix} \quad (2.22)$$

is the Hessian matrix, and

$$\mathbf{s}_M \begin{pmatrix} \mathbf{x} \\ \theta \end{pmatrix} = \begin{pmatrix} \mathbf{s}_{M_{\mathbf{x}}} \\ \mathbf{s}_{M_{\theta}} \end{pmatrix} = \begin{bmatrix} \mathbf{A}^T \mathbf{C}_{\theta}^{-1} (\mathbf{I} - \mathbf{A}\mathbf{x}) \\ \left\{ c - \frac{1}{2} \text{tr}(\mathbf{C}_{\theta}^{-1} \mathbf{T}_k) + \frac{1}{2} (\mathbf{I} - \mathbf{A}\mathbf{x})^T \mathbf{C}_{\theta}^{-1} \mathbf{T}_k \mathbf{C}_{\theta}^{-1} (\mathbf{I} - \mathbf{A}\mathbf{x}) \right\}_{k=1}^r \end{bmatrix} \quad (2.23)$$

is the vector of likelihood scores.

(2) The Fisher scoring method (see, e.g., [Jennrich and Sampson, 1976] and [Searle et al., 1992]), which in the (i+1)-th iteration produces the following vector of location parameters,  $\mathbf{x}$ , and the vector of the variance-covariance components,  $\theta$ :

$$\begin{bmatrix} \hat{\mathbf{x}}_{(i+1)} \\ \hat{\theta}_{(i+1)} \end{bmatrix} = \begin{bmatrix} \hat{\mathbf{x}}_{(i)} \\ \hat{\theta}_{(i)} \end{bmatrix} - \left\{ \mathbf{E} \begin{bmatrix} \frac{\partial^2 l_M}{\partial \mathbf{x} \partial \mathbf{x}} \Big|_{\mathbf{x}=\hat{\mathbf{x}}_{(i)}, \theta=\hat{\theta}_{(i)}} & \frac{\partial^2 l_M}{\partial \mathbf{x} \partial \theta} \Big|_{\mathbf{x}=\hat{\mathbf{x}}_{(i)}, \theta=\hat{\theta}_{(i)}} \\ \frac{\partial^2 l_M}{\partial \theta \partial \mathbf{x}} \Big|_{\mathbf{x}=\hat{\mathbf{x}}_{(i)}, \theta=\hat{\theta}_{(i)}} & \frac{\partial^2 l_M}{\partial \theta \partial \theta} \Big|_{\mathbf{x}=\hat{\mathbf{x}}_{(i)}, \theta=\hat{\theta}_{(i)}} \end{bmatrix} \right\}^{-1} \begin{bmatrix} \frac{\partial l_M}{\partial \mathbf{x}} \Big|_{\mathbf{x}=\hat{\mathbf{x}}_{(i)}, \theta=\hat{\theta}_{(i)}} \\ \frac{\partial l_M}{\partial \theta} \Big|_{\mathbf{x}=\hat{\mathbf{x}}_{(i)}, \theta=\hat{\theta}_{(i)}} \end{bmatrix} \\ = \begin{bmatrix} \hat{\mathbf{x}}_{(i)} \\ \hat{\theta}_{(i)} \end{bmatrix} + \left[ \mathbf{J}_M \begin{pmatrix} \hat{\mathbf{x}}_{(i)} \\ \hat{\theta}_{(i)} \end{pmatrix} \right]^{-1} \mathbf{s}_M \begin{pmatrix} \hat{\mathbf{x}}_{(i)} \\ \hat{\theta}_{(i)} \end{pmatrix}, \quad (2.24)$$

where:

$$\mathbf{J}_M \begin{pmatrix} \mathbf{x} \\ \theta \end{pmatrix} = \begin{bmatrix} \mathbf{A}^T \mathbf{C}_{\theta}^{-1} \mathbf{A} & \mathbf{0} \\ \mathbf{0} & \left\{ \frac{1}{2} \text{tr}(\mathbf{C}_{\theta}^{-1} \mathbf{T}_k \mathbf{C}_{\theta}^{-1} \mathbf{T}_m) \right\}_{k,m=1}^r \end{bmatrix} = \begin{bmatrix} \mathbf{J}_{M_{\mathbf{x}}}(\theta) & \mathbf{0} \\ \mathbf{0} & \mathbf{J}_{M_{\theta\theta}}(\theta) \end{bmatrix} \quad (2.25)$$

is the information matrix.

Taking into account that [Searle et al., 1992]:

$$\left\{ \frac{1}{2} \text{tr}(\mathbf{C}_{\theta}^{-1} \mathbf{T}_k) \right\}_{k=1}^r = \left\{ \frac{1}{2} \text{tr}(\mathbf{C}_{\theta}^{-1} \mathbf{T}_k \mathbf{C}_{\theta}^{-1} \mathbf{T}_m) \right\}_{k,m=1}^r \boldsymbol{\theta} = \mathbf{J}_{\mathbf{M}_{\theta\theta}}(\boldsymbol{\theta}) \boldsymbol{\theta}, \quad (2.26)$$

and using the  $(i+1)$ -th ML estimate of the vector of location parameters,

$$\hat{\mathbf{x}}_{(i+1)} = \left( \mathbf{A}^T \mathbf{C}_{\hat{\theta}_{(i)}}^{-1} \mathbf{A} \right)^{-1} \mathbf{A}^T \mathbf{C}_{\hat{\theta}_{(i)}}^{-1} \mathbf{l} = \left[ \mathbf{J}_{\mathbf{M}_{\mathbf{x}\mathbf{x}}}(\hat{\boldsymbol{\theta}}_{(i)}) \right]^{-1} \mathbf{w}_{\hat{\theta}_{(i)}}, \quad (2.27)$$

with  $\mathbf{w}_{\hat{\theta}_{(i)}} = \mathbf{A}^T \mathbf{C}_{\hat{\theta}_{(i)}}^{-1} \mathbf{l}$ , to compute the second term of the vector of likelihood scores,

$\mathbf{s}_{\mathbf{M}_{\theta}}$ , (see eqn. (2.23)) one gets the following iterative equation for estimating the vector of variance-covariance components:

$$\begin{aligned} \hat{\boldsymbol{\theta}}_{(i+1)} &= \left[ \left\{ \frac{1}{2} \text{tr}(\mathbf{C}_{\hat{\theta}_{(i)}}^{-1} \mathbf{T}_k \mathbf{C}_{\hat{\theta}_{(i)}}^{-1} \mathbf{T}_m) \right\}_{k,m=1}^r \right]^{-1} \left\{ \frac{1}{2} (\mathbf{l} - \mathbf{A} \hat{\mathbf{x}}_{(i+1)})^T \mathbf{C}_{\hat{\theta}_{(i)}}^{-1} \mathbf{T}_k \mathbf{C}_{\hat{\theta}_{(i)}}^{-1} (\mathbf{l} - \mathbf{A} \hat{\mathbf{x}}_{(i+1)}) \right\}_{k=1}^r \\ &= \left[ \left\{ \frac{1}{2} \text{tr}(\mathbf{C}_{\hat{\theta}_{(i)}}^{-1} \mathbf{T}_k \mathbf{C}_{\hat{\theta}_{(i)}}^{-1} \mathbf{T}_m) \right\}_{k,m=1}^r \right]^{-1} \left\{ \frac{1}{2} \mathbf{l}^T \mathbf{R}_{\hat{\theta}_{(i)}} \mathbf{T}_k \mathbf{R}_{\hat{\theta}_{(i)}} \mathbf{l} \right\}_{k=1}^r = \left[ \mathbf{J}_{\mathbf{M}_{\theta\theta}}(\hat{\boldsymbol{\theta}}_{(i)}) \right]^{-1} \mathbf{q}_{\mathbf{R}}(\hat{\boldsymbol{\theta}}_{(i)}), \end{aligned} \quad (2.28)$$

with  $\mathbf{q}_{\mathbf{R}}(\hat{\boldsymbol{\theta}}_{(i)}) = \left\{ \frac{1}{2} \mathbf{l}^T \mathbf{R}_{\hat{\theta}_{(i)}} \mathbf{T}_k \mathbf{R}_{\hat{\theta}_{(i)}} \mathbf{l} \right\}_{k=1}^r$ , where:

$$\mathbf{R}_{\theta} = \mathbf{C}_{\theta}^{-1} \left[ \mathbf{I} - \mathbf{A} (\mathbf{A}^T \mathbf{C}_{\theta}^{-1} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{C}_{\theta}^{-1} \right]. \quad (2.29)$$

Equation (2.27) is identical to the least squares equation with the covariance matrix of the vector of observations given by  $\mathbf{C}_{\hat{\theta}_{(i)}}$ . In the limit, eqn. (2.27) gives the ML estimator of the vector of location parameters  $\hat{\mathbf{x}}$ .

Equation (2.28) forms the basis for the ML estimation of variance-covariance components. In the limit, eqn. (2.28) gives the ML estimator of the vector of variance-covariance components  $\hat{\theta}$ .

Hartley and Rao [1967] establish asymptotic properties of the ML estimates of variance components for the variance components model identical - with the exception that all matrices  $U_i$  are required to have in each row precisely one element equal to 1 and the remaining  $m_i-1$  elements equal to 0, which in turn implies that the matrices  $U_i^T U_i$  are diagonal - to the one given by eqn. (2.3).

Hartley and Rao [1967] assume that the following assumptions (the Hartley-Rao conditions) hold true:

1. The experiment is to be regarded as one of a series of experiments for which  $m \rightarrow \infty$  and  $m_i \rightarrow \infty$  ( $i = 1, 2, \dots, r-1$ ), in such a way that all positive elements of the diagonal matrices  $U_i^T U_i$  are smaller than a constant  $R$ .
2. As  $m \rightarrow \infty$ , all elements of the inverse of the matrix of normal equations,  $(\mathbf{A}^T \mathbf{C}_\theta^{-1} \mathbf{A})^{-1}$ , are in absolute value smaller than  $R/m$ .

3. The base  $\mathbf{W}$  of the adjoined  $m$  by  $(u+n)$  matrix  $\mathbf{M}=[\mathbf{A}|\mathbf{U}_1|\dots|\mathbf{U}_r]$  is of the form

$$\mathbf{W}=[\mathbf{A}|\mathbf{U}^*] \text{ (m by c), where } \mathbf{U}^* \text{ contains at least one column from each } \mathbf{U}_i, n = \sum_{i=1}^r m_i,$$

and  $u+r \leq c \leq u+n$ .

Under these assumptions Hartley and Rao [1967] establish that:

1. If the maximum likelihood estimate of the vector of variance components provides the global maximum of the likelihood, it is weakly consistent.
2. MLE of the vector of variance components is asymptotically efficient, and it is asymptotically normally distributed with the mean  $\theta_0$  (the true parameter) and covariance matrix equal to the inverse of the information matrix computed at  $\theta_0$ .

Applicability of the Hartley - Rao conditions to geodetic networks, and, in particular, to the deformation monitoring network at the Mactaquac hydroelectric generating station near Fredericton, New Brunswick [Chrzanowski and Secord, 1987 and 1990] and [Chrzanowski et al., 1989], is discussed in Chapter 7.

Miller [1977] considers a more general situation, where the Hartley-Rao condition (1) requiring that all positive elements of the diagonal matrices  $\mathbf{U}_i^T \mathbf{U}_i$  are smaller than a universal constant  $R$ , is dropped.

Rao and Kleffe [1988] consider asymptotic properties of MINQE and ML estimators in a sequence of the following random effects models:

$$\mathbf{l}_n = \mathbf{A}_n \mathbf{x}_n + \mathbf{v}_n, \quad \mathbf{v}_n = \mathbf{U}_n \mathbf{b}_n \sim (\mathbf{0}, \mathbf{C}_\theta = \theta_1 \mathbf{T}_{1n} + \theta_2 \mathbf{T}_{2n} + \dots + \theta_r \mathbf{T}_{rn}), \quad n = 1, 2, \dots, \quad (2.30)$$

where  $\mathbf{b}_n$  has independently distributed subsets,  $\mathbf{b}_n = (\mathbf{b}_{1n}^T, \dots, \mathbf{b}_{k(n)n}^T)^T$ , with

$$\mathbf{E}(\mathbf{b}_{in}) = \mathbf{0}, \quad \mathbf{E}(\mathbf{b}_{in} \mathbf{b}_{in}^T) = \Sigma_{in}(\theta), \quad \mathbf{E}\|\mathbf{b}_{in}\|^4 < \infty, \quad (2.31)$$

$$\sup_{i,n} \lambda_{\max} \Sigma_{in}(\theta) < \infty \quad \forall \theta \in E^r, \quad \text{where} \quad (2.32)$$

$\lambda_{\max}$  denotes a maximum eigenvalue of a matrix and  $\|\cdot\|$  stands for the Euclidean norm of a matrix. The covariance matrices  $\Sigma_{in}(\theta)$  are linear functions of  $\theta \in \Theta$  such that they are semi-positive definite for each  $i$  and  $n$ . The vector of variance-covariance components,  $\theta$ , is said to belong to the interior of the parameter space  $\Theta$  if

$$\inf_{i,n} \lambda_{\min} \Sigma_{in}(\theta) > 0. \quad (2.33)$$

Assuming multivariate normality of  $\mathbf{l}$  they prove that if the iterative process given by eqn. (2.28) converges then MLE is weakly consistent. Rao and Kleffe [1988] also prove that, if it exists, MLE is asymptotically efficient, and asymptotically normal with the asymptotic covariance matrix expressed by:

$$\text{var}(\hat{\theta}) \approx [\mathbf{J}_{M_{\theta\theta}}(\hat{\theta})]^{-1} = 2 \left[ \left\{ \text{tr}(\mathbf{C}_{\hat{\theta}}^{-1} \mathbf{T}_k \mathbf{C}_{\hat{\theta}}^{-1} \mathbf{T}_m) \right\}_{k,m=1}^r \right]^{-1}. \quad (2.34)$$

If the matrix of normal equations,  $\mathbf{A}^T \mathbf{C}_{\hat{\theta}}^{-1} \mathbf{A}$ , is rank deficient, then its regular inverse may be replaced by any generalized inverse. In particular, the generalized inverse may be computed by replacing the first order design matrix,  $\mathbf{A}$ , by a matrix  $\mathbf{A}^*$  of dimension  $m$  by  $u^*$ , with  $u^* = \text{rank}(\mathbf{A})$ , composed of any  $u^*$  linearly independent columns of the matrix  $\mathbf{A}$ , taking a regular inverse of the normal equations matrix [Harville, 1977], and augmenting the resulting matrix with  $u - u^*$  rows and column of zeros.

As reported by Jennrich and Sampson [1976], while the Newton-Raphson algorithm converges faster in the neighbourhood of the maximum, the Fisher scoring algorithm performs better farther away from the solution point, and shows better robustness toward poor starting values. If the starting values are very poor, the Newton-Raphson algorithm may converge to a point that is not the global maximum, and in some instances it may not converge at all [Harville, 1977]. Since the expected values of second-order partial derivatives of the log-likelihood function  $l_M$  are easier to compute than the derivatives themselves [Harville, 1977], the Fisher scoring method requires less computer time than the Newton-Raphson procedure. This is achieved, however, at the expense of a slower rate of convergence. In addition, the Fisher scoring method provides the user directly with the asymptotic covariance matrix of the estimates, without the need for additional computations, as is the case with the Newton-Raphson method.



If the constraints on the parameter space  $\Theta$  (see eqn. (2.1)), requiring that the covariance matrix  $C_{\theta}$  is positive definite, are imposed, then the maximization problem to be solved is of the nonlinear constrained optimization type. Harville [1977] proposes the following methods for the solution of the maximum likelihood problem with constraints:

1. The interior methods, belonging to the family of the penalty techniques, that force each iterate to stay within the constraint space by subtracting a penalty function (the penalty increases to infinity at the boundary of the constraint space) from the original likelihood function.
2. The gradient projection technique that projects the original search direction (the direction obtained by the unconstrained gradient method) onto the space defined by the active constraints.
3. The transformation technique that, by the transformation of the variables, transforms the constrained maximization problem into the unconstrained one.

With any of the above given numerical methods (1, 2 or 3), poor choice of the approximate starting values may adversely affect the final results, resulting in estimates that are not the ML estimates of the vector of variance-covariance components [Koch, 1986]. Even if the starting values are carefully chosen [Harville, 1977] there is, in general, no assurance that the solution to the maximization problem will yield the ML estimates. Harville [1977] suggests that one of the possible solutions to this problem would be to perform several computations using different starting points. If all of these computations

result in the same estimates, then one can be reasonably sure that the ML estimates have been found.

The ML estimator of the vector of variance-covariance components,  $\hat{\theta}$ , is translation invariant (see [Harville, 1977] and [Rao and Kleffe, 1988]); however, it is biased as it does not take into account the loss of the degrees of freedom resulting from the estimation of the vector of location parameters  $\mathbf{x}$ .

Not much is known about the small sample properties of ML estimators [Scholz, 1985], and the asymptotic covariance matrix may be unrealistic for samples of small size.

Numerically, the ML estimation is identical to the iterated MINQE with the condition of invariance: IMINQE(I) [Rao, 1979].

### ***2.2.2 Restricted Maximum Likelihood (REML)***

The Restricted Maximum Likelihood (REML) method of estimation of variance-covariance components constitutes an extension of the ML method. As in the ML estimation method, the REML method requires that the probability density function of the vector of observations is specified. Again, the vector of observations is customarily assumed to have the multivariate normal distribution. In contrast to the ML method, however, the REML estimators are obtained by maximizing the likelihood function of the vector of linearly independent error contrasts, as opposed to the maximization of the

likelihood function of the vector of observations in the ML method. The vector of error contrasts is defined by Searle [1979] as a vector of linear combinations of observations that are invariant to  $\mathbf{Ax}$ . This may be thought of as a projection of the vector of observations,  $\mathbf{l}$ , onto the orthogonal complement of a space defined by the column vectors of the first order design matrix  $\mathbf{A}$  [Koch, 1986]. It can be proven (see [Harville, 1977] or [Searle et al., 1992]) that, if  $\mathbf{K}^T\mathbf{l}$  is such a vector, then:

$$\mathbf{K}^T\mathbf{A} = \mathbf{0}, \quad (2.35)$$

which is equivalent to  $\mathbf{K}$  being a product of matrices  $\mathbf{U}$  and  $\mathbf{N}$ , where:

$$\mathbf{N} = \mathbf{I} - \mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T \quad (2.36)$$

is a projection matrix, and  $\mathbf{U}$  is any matrix of full row rank,

$$\text{rank}(\mathbf{U}) = m - \text{rank}(\mathbf{A}) = m - u^*. \quad (2.37)$$

It can be proven that the REML method is invariant with respect to the choice of the error contrasts. The resulting likelihood function differs only by a constant when an alternative set of linearly independent error contrasts is chosen [Harville, 1977].

The resulting (orthogonal complement) likelihood function  $L_R^*$ , which is a likelihood function of the vector of error contrast, is a function of the vector of variance-covariance components,  $\theta$ , and does not depend on the vector of location parameters  $\mathbf{x}$ . It is given by Searle et al. [1992] as:

$$L_R^* = L_R^*(\theta|\mathbf{l}) = (2\pi)^{-\frac{1}{2}(m-u)} |\mathbf{K}^T \mathbf{C}_\theta \mathbf{K}|^{-\frac{1}{2}} \exp\left[-\frac{1}{2} \mathbf{l}^T \mathbf{K} (\mathbf{K}^T \mathbf{C}_\theta \mathbf{K})^{-1} \mathbf{K}^T \mathbf{l}\right], \quad (2.38)$$

which differs only by a proportionality factor [Harville, 1977] from an alternative expression:

$$L_R^{**} = L_R^{**}(\theta|\mathbf{l}) = |\mathbf{C}_\theta|^{-\frac{1}{2}} |\mathbf{A}^T \mathbf{C}_\theta^{-1} \mathbf{A}|^{-\frac{1}{2}} \exp\left(-\frac{1}{2} \mathbf{l}^T \mathbf{R}_\theta \mathbf{l}\right), \quad (2.39)$$

where (see [Searle et al., 1992] and [Rao and Kleffe, 1988]):

$$\mathbf{R}_\theta = \mathbf{K} (\mathbf{K}^T \mathbf{C}_\theta \mathbf{K})^{-1} \mathbf{K}^T, \text{ and,} \quad (2.40)$$

under the assumption that  $\mathbf{K}^T \mathbf{K} = \mathbf{I}$  and  $\mathbf{K} \mathbf{K}^T = \mathbf{N}$  (see eqn. (3.43),

$$|\mathbf{K}^T \mathbf{C}_\theta \mathbf{K}| = |\mathbf{C}_\theta| |\mathbf{A}^T \mathbf{C}_\theta^{-1} \mathbf{A}| |\mathbf{A}^T \mathbf{A}|^{-1}. \quad (2.41)$$

As reported by Harville [1977], all statistical inferences concerning REML estimation are based on  $m-u^*$  linearly independent error contrasts. As a result, it makes no difference which subset of error contrasts is used for estimation, as the likelihood function for any subset differs only by a proportionality factor. In particular, if the matrix of normal equations,  $\mathbf{A}^T \mathbf{C}_\theta^{-1} \mathbf{A}$ , is rank deficient, then its regular inverse may be replaced by any generalized inverse without affecting the results of estimation. One of the possible choices for a generalized inverse is the regular inverse of the dimension  $u^*$  by  $u^*$ , with  $u^* = \text{rank}(\mathbf{A})$ , of the reduced normal equations matrix  $(\mathbf{A}^*)^T \mathbf{C}_\theta^{-1} \mathbf{A}^*$  - with  $\mathbf{A}^*$  being an  $m$  by  $u^*$  matrix composed of any  $u^*$  linearly independent columns of the matrix  $\mathbf{A}$  - augmented by  $u-u^*$  rows and columns of zeros. In subsequent formulae, whenever the regular inverse is

used, it will be implicitly understood that in the case of  $\text{rank}(\mathbf{A}) < u$ , the matrix  $\mathbf{A}$  will be replaced by  $\mathbf{A}^*$  and the number of unknown location parameters,  $u$ , will be replaced by  $u^*$ .

Instead of maximizing the orthogonal complement likelihood function, its natural logarithm can be maximized:

$$l_R^* = \ln(L_R^*) = -\frac{1}{2}(m-u)\ln(2\pi) - \frac{1}{2}\ln|\mathbf{K}^T\mathbf{C}_\theta\mathbf{K}| - \frac{1}{2}\mathbf{1}^T\mathbf{K}(\mathbf{K}^T\mathbf{C}_\theta\mathbf{K})^{-1}\mathbf{K}^T\mathbf{l}, \quad (2.42)$$

or, alternatively:

$$l_R^{**} = \ln(L_R^{**}) = -\frac{1}{2}\ln|\mathbf{C}_\theta| - \frac{1}{2}\ln|\mathbf{A}^T\mathbf{C}_\theta^{-1}\mathbf{A}| - \frac{1}{2}\mathbf{1}^T\mathbf{R}_\theta\mathbf{l}. \quad (2.43)$$

The REML estimation method can also be viewed as an application of the method of maximum likelihood to the marginal likelihood function (see, e.g., [Koch, 1988], [Koch, 1990] or [Searle et al., 1992]) obtained from the original likelihood function  $L_M$  by integrating the vector of location parameters,  $\mathbf{x}$ , in the following way:

$$\begin{aligned} L_R(\theta|\mathbf{l}) &= p(\mathbf{l}|\theta) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} L_M(\theta, \mathbf{x}|\mathbf{l}) d\mathbf{x} = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} p(\mathbf{l}|\theta, \mathbf{x}) d\mathbf{x} \\ &\propto |\mathbf{C}_\theta|^{-\frac{1}{2}} |\mathbf{A}^T\mathbf{C}_\theta^{-1}\mathbf{A}|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\mathbf{1}^T\mathbf{R}_\theta\mathbf{l}\right). \end{aligned} \quad (2.44)$$

The marginal log-likelihood function resulting from the natural logarithm of the marginal likelihood function  $L_R$ ,

$$l_R = \ln(L_R) = c - \frac{1}{2}\ln|\mathbf{C}_\theta| - \frac{1}{2}\ln|\mathbf{A}^T\mathbf{C}_\theta^{-1}\mathbf{A}| - \frac{1}{2}\mathbf{1}^T\mathbf{R}_\theta\mathbf{l}, \quad (2.45)$$

is identical, up to a constant  $c$ , to the orthogonal complement log-likelihood function shown in eqn. (2.43).

The first derivative of the log-likelihood function  $l_{\mathbf{R}}^*$ , eqn. (2.42), with respect to the  $i$ -th variance-covariance component is computed as follows:

$$\frac{\partial l_{\mathbf{R}}^*}{\partial \theta_i} = -\frac{1}{2} \frac{\partial \ln |\mathbf{K}^T \mathbf{C}_\theta \mathbf{K}|}{\partial \theta_i} - \frac{1}{2} \frac{\partial \mathbf{l}^T \mathbf{K} (\mathbf{K}^T \mathbf{C}_\theta \mathbf{K})^{-1} \mathbf{K}^T \mathbf{l}}{\partial \theta_i}, \quad (2.46)$$

which on using eqn. (2.40) becomes

$$\frac{\partial l_{\mathbf{R}}^*}{\partial \theta_i} = -\frac{1}{2} \frac{\partial \ln |\mathbf{K}^T \mathbf{C}_\theta \mathbf{K}|}{\partial \theta_i} - \frac{1}{2} \frac{\partial \mathbf{l}^T \mathbf{R}_\theta \mathbf{l}}{\partial \theta_i}. \quad (2.47)$$

Then, using (see, e.g., [Searle et al., 1992])

$$\frac{\partial \ln |\mathbf{A}|}{\partial a_i} = \text{tr} \left( \mathbf{A}^{-1} \frac{\partial \mathbf{A}}{\partial a_i} \right), \quad (2.48)$$

which holds for a regular symmetric matrix  $\mathbf{A}$  being a function of  $a_i$ , and using (see, e.g., [Koch, 1990])

$$\frac{\partial \mathbf{l}^T \mathbf{R}_\theta \mathbf{l}}{\partial \theta_i} = -\mathbf{l}^T \mathbf{R}_\theta \mathbf{T}_i \mathbf{R}_\theta \mathbf{l} \quad (2.49)$$

we get

$$\frac{\partial l_{\mathbf{R}}^*}{\partial \theta_i} = -\frac{1}{2} \text{tr} \left[ (\mathbf{K}^T \mathbf{C}_\theta \mathbf{K})^{-1} \mathbf{K}^T \mathbf{T}_i \mathbf{K} \right] + \frac{1}{2} \mathbf{l}^T \mathbf{R}_\theta \mathbf{T}_i \mathbf{R}_\theta \mathbf{l}, \quad (2.50)$$

and because of eqn. (2.40)

$$\frac{\partial l_{\mathbf{R}}^*}{\partial \theta_i} = -\frac{1}{2} \text{tr}(\mathbf{R}_\theta \mathbf{T}_i) + \frac{1}{2} \mathbf{1}^T \mathbf{R}_\theta \mathbf{T}_i \mathbf{R}_\theta \mathbf{1}. \quad (2.51)$$

Similarly, the first derivative of the log-likelihood function  $l_{\mathbf{R}}^{**}$ , eqn. (2.43), with respect to the  $i$ -th variance-covariance component can be derived by computing the following derivatives:

$$\frac{\partial l_{\mathbf{R}}^{**}}{\partial \theta_i} = -\frac{1}{2} \frac{\partial \ln |\mathbf{C}_\theta|}{\partial \theta_i} - \frac{1}{2} \frac{\partial \ln |\mathbf{A}^T \mathbf{C}_\theta^{-1} \mathbf{A}|}{\partial \theta_i} - \frac{1}{2} \frac{\partial \mathbf{1}^T \mathbf{R}_\theta \mathbf{1}}{\partial \theta_i}. \quad (2.52)$$

Using eqn. (2.48) and eqn. (2.49), we get

$$\frac{\partial l_{\mathbf{R}}^{**}}{\partial \theta_i} = -\frac{1}{2} \text{tr}(\mathbf{C}_\theta^{-1} \mathbf{T}_i) - \frac{1}{2} \text{tr} \left[ (\mathbf{A}^T \mathbf{C}_\theta^{-1} \mathbf{A})^{-1} \mathbf{A}^T \frac{\partial \mathbf{C}_\theta^{-1}}{\partial \theta_i} \mathbf{A} \right] + \frac{1}{2} \mathbf{1}^T \mathbf{R}_\theta \mathbf{T}_i \mathbf{R}_\theta \mathbf{1}. \quad (2.53)$$

This becomes, on using (see, e.g., [Searle et al., 1992])

$$\frac{\partial \mathbf{A}^{-1}}{\partial a_i} = -\mathbf{A}^{-1} \frac{\partial \mathbf{A}}{\partial a_i} \mathbf{A}^{-1}, \quad (2.54)$$

$$\begin{aligned} \frac{\partial l_{\mathbf{R}}^{**}}{\partial \theta_i} &= -\frac{1}{2} \text{tr}(\mathbf{C}_\theta^{-1} \mathbf{T}_i) + \frac{1}{2} \text{tr} \left[ \mathbf{C}_\theta^{-1} \mathbf{A} (\mathbf{A}^T \mathbf{C}_\theta^{-1} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{C}_\theta^{-1} \mathbf{T}_i \right] + \frac{1}{2} \mathbf{1}^T \mathbf{R}_\theta \mathbf{T}_i \mathbf{R}_\theta \mathbf{1} \\ &= -\frac{1}{2} \text{tr}(\mathbf{R}_\theta \mathbf{T}_i) + \frac{1}{2} \mathbf{1}^T \mathbf{R}_\theta \mathbf{T}_i \mathbf{R}_\theta \mathbf{1}, \end{aligned} \quad (2.55)$$

which is identical to eqn. (2.51).

Thus, the vector of likelihood scores is for all three log-likelihood functions given as:

$$\mathbf{s}_{\mathbf{R}}(\theta) = \frac{\partial l_{\mathbf{R}}}{\partial \theta} = \frac{\partial l_{\mathbf{R}}^*}{\partial \theta} = \frac{\partial l_{\mathbf{R}}^{**}}{\partial \theta} = \left\{ c - \frac{1}{2} \text{tr}(\mathbf{R}_\theta \mathbf{T}_k) + \frac{1}{2} \mathbf{1}^T \mathbf{R}_\theta \mathbf{T}_k \mathbf{R}_\theta \mathbf{1} \right\}_{k=1}^r. \quad (2.56)$$

The Hessian matrix is given by Searle et al. [1992] as:

$$\mathbf{H}_R(\boldsymbol{\theta}) = \frac{\partial^2 l_R}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}} = \left\{ \frac{1}{2} \text{tr}(\mathbf{R}_\theta \mathbf{T}_k \mathbf{R}_\theta \mathbf{T}_m) - \mathbf{l}^T \mathbf{R}_\theta \mathbf{T}_k \mathbf{R}_\theta \mathbf{T}_m \mathbf{R}_\theta \mathbf{l} \right\}_{k,m=1}^r. \quad (2.57)$$

The Restricted Maximum Likelihood method is also known as the Marginal Maximum Likelihood (MML) method (see, e.g., [Rao, 1979] or [Rao and Kleffe, 1988]). The REML (MML) estimator of the vector of variance-covariance components  $\boldsymbol{\theta}$  is defined as the value,  $\hat{\boldsymbol{\theta}}$ , that maximizes the orthogonal complement (marginal) log-likelihood function, subject to the condition that the vector  $\boldsymbol{\theta}$  belongs to the parameter space  $\Theta$ , and is given by Harville [1977] as:

$$l_R(\hat{\boldsymbol{\theta}}) = \sup_{\boldsymbol{\theta} \in \Theta} l_R, \quad \text{subject to } \boldsymbol{\theta} \in \Theta. \quad (2.58)$$

If the constraints imposed by the definition of the parameter space  $\Theta$  are ignored, then the maximization problem to be solved is of the nonlinear, unconstrained optimization type. As in the ML estimation method, the two numerical methods that have been most widely used in the context of the REML estimation of variance-covariance components are:

(1) The Newton-Raphson method (see, e.g., [Jennrich and Sampson, 1976], [Harville, 1977] and [Searle et al., 1992]) which gives the  $(i+1)$ -th iterate of the vector of variance-covariance components,  $\boldsymbol{\theta}$ , as:



$$\hat{\theta}_{(i+1)} = \hat{\theta}_{(i)} - [\mathbf{H}_R(\hat{\theta}_{(i)})]^{-1} \mathbf{s}_R(\hat{\theta}_{(i)}), \quad (2.59)$$

where:

$\mathbf{H}_R(\theta)$  is the Hessian matrix, and

$\mathbf{s}_R(\theta)$  is the vector of likelihood scores.

(2) The Fisher scoring method (see [Jennrich and Sampson, 1976] and [Searle et al., 1992]), which in the (i+1)-th iteration produces the following estimate of the vector of variance-covariance components  $\theta$ :

$$\hat{\theta}_{(i+1)} = \hat{\theta}_{(i)} - \left\{ E[\mathbf{H}_R(\hat{\theta}_{(i)})] \right\}^{-1} \mathbf{s}_R(\hat{\theta}_{(i)}) = \hat{\theta}_{(i)} + [\mathbf{J}_R(\hat{\theta}_{(i)})]^{-1} \mathbf{s}_R(\hat{\theta}_{(i)}), \quad (2.60)$$

where:

$$\mathbf{J}_R(\theta) = \left\{ \frac{1}{2} \text{tr}(\mathbf{R}_\theta \mathbf{T}_k \mathbf{R}_\theta \mathbf{T}_m) \right\}_{k,m=1}^r \quad (2.61)$$

is the information matrix.

Taking into account that [Searle et al., 1992]:

$$\left\{ \frac{1}{2} \text{tr}(\mathbf{R}_\theta \mathbf{T}_k) \right\}_{k=1}^r = \left\{ \frac{1}{2} \text{tr}(\mathbf{R}_\theta \mathbf{T}_k \mathbf{R}_\theta \mathbf{T}_m) \right\}_{k,m=1}^r \theta = \mathbf{J}_R(\theta) \theta, \quad (2.62)$$

one gets the following set of scoring equations:

$$\begin{aligned}\hat{\theta}_{(i+1)} &= \left[ \left\{ \frac{1}{m} \text{tr} \left( \mathbf{R}_{\hat{\theta}_{(i)}} \mathbf{T}_k \mathbf{R}_{\hat{\theta}_{(i)}} \mathbf{T}_m \right) \right\}_{k,m=1}^r \right]^{-1} \left\{ \frac{1}{c} \mathbf{1}^T \mathbf{R}_{\hat{\theta}_{(i)}} \mathbf{T}_k \mathbf{R}_{\hat{\theta}_{(i)}} \mathbf{1} \right\}_{k=1}^r \\ &= \left[ \mathbf{J}_R(\hat{\theta}_{(i)}) \right]^{-1} \mathbf{q}_R(\hat{\theta}_{(i)}).\end{aligned}\quad (2.63)$$

In the limit, eqn. (2.63) gives the REML estimator of the vector of variance-covariance components  $\hat{\theta}$ .

Under the Hartley - Rao conditions (see Subsection 2.2.1) the asymptotic properties of the REML estimates of variance components are the same as the asymptotic properties of the ML estimates, since REML essentially constitutes only a reparametrization of ML [Silvey, 1975].

A proof that, under certain conditions, REMLE is weakly consistent, asymptotically efficient, and asymptotically normal, with the asymptotic covariance matrix :

$$\text{var}(\hat{\theta}) \approx \left[ \mathbf{J}_R(\hat{\theta}) \right]^{-1} = 2 \left[ \left\{ m \text{tr} \left( \mathbf{R}_{\hat{\theta}} \mathbf{T}_k \mathbf{R}_{\hat{\theta}} \mathbf{T}_m \right) \right\}_{k,m=1}^r \right]^{-1}. \quad (2.64)$$

can be found in Rao and Kleffe [1988].

If the constraints on the parameter space  $\Theta$  are not ignored, then the maximization problem given by eqn. (2.58) is of the nonlinear, constrained optimization type. The nonlinear constrained optimization techniques, listed in Subsection 2.2.1 in the context of ML estimation, may also be applied to such estimation problem.

In general, the REML method suffers from the same problems as the ML method, i.e., the problems concerned with the appropriate choice of the approximate starting values of variance-covariance components, and the problems with estimates falling outside of the parameter space  $\Theta$ .

As the REML estimators account for the loss of degrees of freedom resulting from estimation of  $\mathbf{x}$ , they are unbiased. For balanced data, the REML solutions are identical to the ANOVA estimates [Searle, 1987]. Numerically, the REML estimation is identical to the iterated MINQE with the conditions of invariance and unbiasedness: IMINQE(U,I) [Rao, 1979]. REML is also numerically equivalent to the iterated BIQUE.

## **2.3 Bayesian Methods**

### ***2.3.1 Bayesian Approach to the Estimation of Variance-Covariance Components - General Formulation***

As in the maximum likelihood methods, the Bayesian methods require that the distribution function of the vector of observations is specified. In contrast to the other methods, however, the Bayesian methods treat the unknown variance-covariance components as random quantities. The Bayesian methodology requires, therefore, that some distribution function (prior probability density function) is assigned to the vector of variance-covariance components, expressing the existing prior knowledge about this vector.

Application of the Bayes' Theorem results in the posterior density function, that is later used to find estimates of the vector of variance-covariance components.

The Bayes' Theorem (as applied to the problem of estimation of variance-covariance components) states that the posterior probability density function,  $p(\theta|\mathbf{l})$ , of the vector of variance-covariance components,  $\theta$ , given the vector of observations,  $\mathbf{l}$ , is proportional to the product of the prior probability density function,  $p(\theta)$ , of the vector of variance-covariance components,  $\theta$ , and the likelihood function,  $p(\mathbf{l}|\theta)$ , (probability density function of the vector of observations,  $\mathbf{l}$ , given the vector of variance-covariance components  $\theta$ ):

$$p(\theta|\mathbf{l}) = c p(\theta) p(\mathbf{l}|\theta). \quad (2.65)$$

The proportionality factor,  $c$ , is given as:

$$c = \frac{1}{p(\mathbf{l})}, \quad (2.66)$$

where the density function of the vector of observations,  $p(\mathbf{l})$ , may be obtained from the joint density function of  $\mathbf{l}$  and  $\theta$ ,  $p(\mathbf{l},\theta)$ , as a marginal density:

$$p(\mathbf{l}) = \int_{\Theta} p(\mathbf{l},\theta) d\theta = \int_{\Theta} p(\theta) p(\mathbf{l}|\theta) d\theta, \quad \theta \in \Theta. \quad (2.67)$$

A linear model with unknown variance-covariance components, which is equivalent to the general mixed model, eqn. (2.1), of the classical (non-Bayesian) approach, is defined by Koch [1987] as:

$$E(\mathbf{l}|\mathbf{x}) = \mathbf{Ax}, \quad D(\mathbf{l}|\theta) = \mathbf{C}_\theta = \theta_1\mathbf{T}_1 + \theta_2\mathbf{T}_2 + \dots + \theta_r\mathbf{T}_r, \quad \theta \in \Theta, \quad (2.68)$$

where:  $E(\mathbf{l}|\mathbf{x})$  is the expected value of the vector of observations,  $\mathbf{l}$ , under the condition

that the vector of location parameters,  $\mathbf{x}$ , is given,

$\mathbf{A}$  is a given first order design matrix,

$\mathbf{x}$  is the random vector of location parameters with prior probability density

function  $p(\mathbf{x})$ ,

$\mathbf{T}_1, \dots, \mathbf{T}_r$  are given, linearly independent, symmetric matrices,

$\theta$  is the random vector of variance-covariance components with prior probability

density function  $p(\theta)$ ,

$\Theta$  is the set of all  $\theta \in E^r$  such that

$D(\mathbf{l}|\theta)$  - the covariance matrix of the vector of observations  $\mathbf{l}$  - is positive definite.

Under the linear model with unknown variance-covariance components, eqn. (2.68), and under the assumption that the vector of observations,  $\mathbf{l}$ , is normally distributed, the likelihood function (the density function of the vector of observations,  $\mathbf{l}$ , given the parameters  $\mathbf{x}$  and  $\theta$ ), is identical to the ML likelihood function  $L_M$ , eqn. (2.18). This is given by Koch [1987] as:

$$p(\mathbf{l}|\mathbf{x},\theta) = L_M(\mathbf{x},\theta|\mathbf{l}) = (2\pi)^{-\frac{m}{2}} |\mathbf{C}_\theta|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(\mathbf{l} - \mathbf{Ax})^T \mathbf{C}_\theta^{-1}(\mathbf{l} - \mathbf{Ax})\right]. \quad (2.69)$$

As indicated by Harville [1977], the preferable approach is to avoid dealing with the full likelihood function, eqn. (2.69), and use a marginal likelihood function instead, i.e., a function that does not depend on the vector of location parameters  $\mathbf{x}$ . To obtain the likelihood function,  $p(\mathbf{l}|\theta)$ , that depends only on the vector of variance-covariance components,  $\theta$ , the vector of location parameters,  $\mathbf{x}$ , is removed through integration of the likelihood function  $p(\mathbf{l}|\mathbf{x},\theta)$ . The resulting marginal likelihood function:

$$p(\mathbf{l}|\theta) = L_R(\theta|\mathbf{l}) \propto |\mathbf{C}_\theta|^{-\frac{1}{2}} |\mathbf{A}^T \mathbf{C}_\theta^{-1} \mathbf{A}|^{-\frac{1}{2}} \exp\left(-\frac{1}{2} \mathbf{l}^T \mathbf{R}_\theta \mathbf{l}\right), \quad (2.70)$$

is identical to the marginal likelihood function  $L_R$ , eqn. (2.44), and differs only by a proportionality constant from the orthogonal complement likelihood function  $L_R^{**}$ , eqn. (2.39), which serves as a basis for the REML estimation.

As in the REML estimation method, the problem of rank deficiency of the matrix of normal equations,  $\mathbf{A}^T \mathbf{C}_\theta^{-1} \mathbf{A}$ , may be dealt with by replacing the first order design matrix,  $\mathbf{A}$ , by an  $m$  by  $u^*$  matrix  $\mathbf{A}^*$  comprised of any  $u^*$  linearly independent columns of the matrix  $\mathbf{A}$ , with  $u^* = \text{rank}(\mathbf{A})$ .

Koch [1990] proposes two approaches to the problem of the choice of the prior probability density function for the vector of variance-covariance components  $\theta$ .

In the first approach, it is assumed that there is no prior information concerning the vector  $\theta$ . The so-called noninformative prior (Jeffrey's prior) probability density function (see, e.g., [Harville, 1977] or [Koch 1990]) is chosen for the prior probability density function, which is proportional to the square root of the determinant of the information matrix  $\mathbf{J}_R(\theta)$  - associated with the marginal likelihood function  $L_R$ :

$$p(\theta) \propto |\mathbf{J}_R(\theta)|^{\frac{1}{2}} \quad (2.71)$$

where:

$$\mathbf{J}_R(\theta) = \left\{ \frac{1}{2} \text{tr}(\mathbf{R}_\theta \mathbf{T}_k \mathbf{R}_\theta \mathbf{T}_m) \right\}_{k,m=1}^r. \quad (2.72)$$

The posterior probability density function,  $p(\theta|\mathbf{l})$ , of the vector of variance-covariance components,  $\theta$ , given the observations,  $\mathbf{l}$ , resulting from the application of Bayes' Theorem, is then:

$$p(\theta|\mathbf{l}) \propto |\mathbf{J}_R(\theta)|^{\frac{1}{2}} |\mathbf{C}_\theta|^{-\frac{1}{2}} |\mathbf{A}^T \mathbf{C}_\theta^{-1} \mathbf{A}|^{-\frac{1}{2}} \exp\left(-\frac{1}{2} \mathbf{l}^T \mathbf{R}_\theta \mathbf{l}\right). \quad (2.73)$$

The posterior probability density function, eqn. (2.73), is later used to find the Bayes estimate of the vector of variance-covariance components. It is also used for the purpose of interval estimation and testing of statistical hypotheses.

In the second approach, Koch [1990] assumes that there exists some prior information about the vector of variance-covariance components  $\theta$ , and that the vector of variance-covariance components is composed of the sub-vector of variance components being strictly positive, and the sub-vector of covariance components which can take both positive and negative values.

Koch [1990] asserts that, since the variance components admit only positive values, the choice of the prior probability density function for these components must reflect this fact. Two alternatives for the choice of the prior distribution function for the variance components are given:

(1) the inverted gamma distribution, which, assuming that the  $o$  variance components are independent, has the following probability density function:

$$p_1(\theta^I) \propto \prod_{i=1}^o \left[ \left( \frac{1}{\theta_i} \right)^{p_i+1} \exp\left(-\frac{b_i}{\theta_i}\right) \right] \text{ for } \theta^I = \{\theta_1, \theta_2, \dots, \theta_o\} \quad (2.74)$$

with:  $b_i > 0$ ,  $p_i > 0$ , and  $0 < \theta_i < \infty$ , for  $i = 1, 2, \dots, o$ ,

$$p_i = \frac{\theta_{0_i}^2}{V_{\theta_{0_i}}} + 2, \text{ and}$$

$$b_i = (p_i - 1)\theta_{0_i},$$

where:  $\theta_{0_i}$  is the prior value (mean of the prior distribution) of the  $i$ -th variance component, and



$V_{\theta_{0_i}}$  is the variance of the prior distribution for the  $i$ -th variance component,

and (2) the truncated normal distribution, which, assuming that the  $o$  variance components are independent, has the probability density function of the following form:

$$p_1(\theta^I) \propto \prod_{i=1}^o \exp \left[ -k_{1i}\theta_i - k_{2i}(\theta_i - \theta_{0_i})^2 \right] \text{ for } \theta^I = \{\theta_1, \theta_2, \dots, \theta_o\} \quad (2.75)$$

with:  $k_{2i} > 0$ , and  $0 < \theta_i < \infty$ , for  $i = 1, 2, \dots, o$ ,

where:  $\theta_{0_i}$  is the prior value of the  $i$ -th variance component,

$V_{\theta_{0_i}}$  is the variance of the prior information for the  $i$ -th variance component, and

$k_{1i}$  and  $k_{2i}$  are the parameters of the truncated normal distribution for  $i$ -th variance component.

The truncated normal distribution is a valid choice for the prior probability density function provided the relationship  $V_{\theta_{0_i}} < \theta_{0_i}^2$  holds.

Arguing that since the  $(r-o)$  covariance components, comprising the sub-vector of covariance components may take both positive and negative values, Koch [1990] proposes the multivariate normal distribution for a prior. The probability density function of this form of prior, assuming that the  $(r - o)$  covariance components are independent, is:

$$p_2(\theta^{II}) \propto \prod_{i=o+1}^r \exp \left[ -\frac{(\theta_i - \theta_{0_i})^2}{V_{\theta_{0_i}}} \right] \text{ for } \theta^{II} = \{\theta_{o+1}, \theta_{o+2}, \dots, \theta_r\}. \quad (2.76)$$

In his choice of informative priors Koch [1990] does not take into account constraints on the parameter space,  $\Theta$ , implied by the requirement of positive definiteness of the covariance matrix  $C_\theta$ . The only constraint on the parameter space assumed by Koch [1990] is that the variance components are strictly positive. There are no constraints on the sign of covariance components. As a result, it is possible to find such a set of variance and covariance components (with variance components being positive numbers) for which the covariance matrix  $C_\theta$  is negative definite [Gupta, 1995]. Moreover, the normal distribution, used as a prior for the covariance components, assigns positive probability to impossible situations, such as the case when the covariance matrix is negative definite [Knight, 1995].

### ***2.3.2 Bayes Estimation***

As indicated before, estimation of the vector of variance-covariance components is based on the posterior probability density function. The estimation process may be viewed as a statistical decision problem (see, e.g., [Silvey, 1975] or [Koch, 1990]), with an associated loss function  $L(\theta, \hat{\theta})$ . The posterior expected loss of the estimation is defined as:

$$E[L(\theta, \hat{\theta})] = \int_{\Theta} L(\theta, \hat{\theta}) p(\theta|\mathbf{1}) d\theta . \quad (2.77)$$

A Bayes estimator is defined as the value  $\hat{\theta}$  that minimizes the posterior expected loss. The statistical decision that minimizes the posterior expected loss is called the Bayes rule [Koch, 1990].

If the quadratic loss function,

$$L(\theta, \hat{\theta}) = (\theta - \hat{\theta})^T \mathbf{P}(\theta - \hat{\theta}), \quad (2.78)$$

with  $\mathbf{P}$  being a positive definite matrix of constants, is chosen, then it can be proven [Koch, 1990] that the Bayes estimator is given by the expected value of the vector of variance-covariance components,  $\theta$ , computed with the posterior probability density function,  $p(\theta|\mathbf{I})$ , as:

$$\hat{\theta}_B = E(\theta) = \int_{\Theta} \theta p(\theta|\mathbf{I}) d\theta . \quad (2.79)$$

The covariance matrix of the Bayes estimate,  $\hat{\theta}_B$ , is given by Koch [1990] as:

$$\text{var}(\theta) = E\{[\theta - E(\theta)][\theta - E(\theta)]^T\} = \int_{\Theta} [\theta - E(\theta)][\theta - E(\theta)]^T p(\theta|\mathbf{I}) d\theta . \quad (2.80)$$

Regardless of whether the informative or noninformative priors are used, the integral in eqn. (2.79), required to get the Bayes estimate, the integral in eqn. (2.80), required to get its covariance matrix, and the integral in eqn. (2.67), required to get the proportionality constant, cannot be solved analytically [Koch, 1990]. Numerical integration techniques, such as the Monte Carlo methods, have to be applied to solve for these quantities.

### 2.3.3 Interval Estimation and Hypothesis Testing

The posterior probability density function,  $p(\theta|\mathbf{l})$ , may also serve for interval estimation of the vector of variance-covariance components,  $\theta$ , or in other words, for the computation of the confidence regions. A Bayesian confidence region of content  $1-\alpha$ , also called a region of highest posterior density (see, e.g., [Koch, 1990]), is defined as the subspace  $B$  of the parameter space  $\Theta$  that satisfies the following conditions:

$$P(\theta \in B|\mathbf{l}) = \int_B p(\theta|\mathbf{l}) d\theta = 1-\alpha, \text{ and}$$
$$p(\theta_1|\mathbf{l}) \geq p(\theta_2|\mathbf{l}) \text{ for } \theta_1 \in B, \theta_2 \notin B. \quad (2.81)$$

Regardless of whether the informative or noninformative priors are used, the integral in eqn. (2.81), required to get the Bayesian confidence region of the vector of variance-covariance components, cannot be solved analytically [Koch, 1990]. Numerical integration techniques have to be applied to find the solution.

The posterior probability density function  $p(\theta|\mathbf{l})$  can be used for the hypotheses testing. A statistical testing of a null hypothesis  $H_0: \theta \in \Theta_0$  versus an alternative hypothesis  $H_1: \theta \in \Theta_1$ , can be viewed as a statistical decision problem, with an associated loss function  $L$ :

$$L(\theta \in \Theta_i, H_j) \begin{cases} = 0 & \text{for } i = j \\ \neq 0 & \text{for } i \neq j \end{cases} \quad \text{with } i, j \in \{0, 1\}. \quad (2.82)$$

Applying the Bayes rule to the posterior expected losses for accepting  $H_0$  versus  $H_1$ , under the assumption that  $\Theta_1$  is the complement of  $\Theta_0$  one gets the following rule [Koch, 1990]:

$$\text{if } \frac{P(H_0|\mathbf{1})L(\theta \in \Theta_0, H_1)}{P(H_1|\mathbf{1})L(\theta \in \Theta_1, H_0)} > 1 \text{ then accept } H_0, \quad (2.83)$$

where:

$$P(H_i|\mathbf{1}) = \int_{\Theta_i} p(\theta|\mathbf{1}) d\theta, \quad \text{for } i \in \{0, 1\}. \quad (2.84)$$

If the same loss is assigned to both incorrect decisions, i.e., rejection of the null hypothesis if it is true, or acceptance of the null hypothesis if it is not true, then the ratio given by eqn. (2.83) is reduced to the ratio of posterior probabilities, called the posterior odds, resulting in the following rule:

$$\text{if } \frac{P(H_0|\mathbf{1})}{P(H_1|\mathbf{1})} > 1 \text{ then accept } H_0. \quad (2.85)$$

If the simple hypothesis is tested, i.e.,  $H_0: \theta = \theta_0$  versus  $H_1: \theta = \theta_1$ , and the same loss is assigned to both incorrect decisions then the posterior odds are reduced to the ratio of the

values of the posterior probability density function computed at  $\theta_0$  and  $\theta_1$ , resulting in the following rule:

$$\text{if } \frac{p(\theta_0|\mathbf{I})}{p(\theta_1|\mathbf{I})} > 1 \text{ then accept } H_0 . \quad (2.86)$$

Regardless of whether the informative or noninformative priors are used, the integral in eqn. (2.84), required to compute the posterior odds, cannot be solved analytically [Koch, 1990]. Numerical integration techniques have to be applied to find the solution. The simple hypotheses can, however, be tested without any additional computational effort, using eqn. (2.86).

#### ***2.3.4 Generalized Maximum Likelihood (GML) Estimation***

It may be argued (see, e.g., [Lindley and Smith, 1972] or [Harville, 1977]) that, with the exception of trivial cases, the computational effort involved in solving for the proportionality factor, eqn. (2.66), and the Bayes estimate of the variance-covariance components, eqn. (2.79), makes the task impractical. Because of this fact, it is sometimes proposed (see, e.g., [Harville, 1977] or [Koch, 1990]) that the expected values are to be replaced by the modal values, which are obtained by applying the principle of maximum likelihood estimation to the posterior probability density function  $p(\theta|\mathbf{I})$ . The resulting estimator is known as the Generalized Maximum Likelihood (GML) estimator, or the Maximum a Posteriori (MAP) estimator [Koch, 1990].

Similar to the maximum likelihood approach, instead of maximizing the posterior probability density function,  $p(\theta|\mathbf{l})$ , itself, one may maximize its natural logarithm:

$$l_P = \ln[p(\theta|\mathbf{l})] . \quad (2.87)$$

The GML estimator of the vector of variance-covariance components,  $\theta$ , is defined [Koch, 1990] as the value,  $\hat{\theta}$ , that maximizes the natural logarithm of the posterior probability density function  $l_P$ , subject to the condition that the vector  $\theta$  belongs to the parameter space  $\Theta$  - defined by the requirement of positive definiteness of the covariance matrix (see eqn. (2.68)):

$$l_P(\hat{\theta}) = \sup_{\theta} l_P , \quad \text{subject to } \theta \in \Theta . \quad (2.88)$$

If the constraints on the parameter space are ignored, then one of the numerical methods for non-linear unconstrained optimization, such as the Newton-Raphson or the Fisher scoring method (see Section 2.2), may be applied to solve eqn. (2.88).

It can be proven (see Appendix B) that, under certain regularity conditions GML estimators are asymptotically normal with the covariance matrix given by the inverse of the information matrix,  $\mathbf{J}_P$ , computed at  $\hat{\theta}$ :

$$\text{var}(\hat{\theta}) \approx [\mathbf{J}_P(\hat{\theta})]^{-1} , \quad (2.89)$$

where,

$$\mathbf{J}_P(\boldsymbol{\theta}) = -\mathbf{E} \left( \left\{ \frac{\partial^2 l_P}{\partial \theta_i \partial \theta_j} \right\}_{i,j=1}^r \right)^m. \quad (2.90)$$

Asymptotic properties of GML estimators are analyzed in Appendix B.

To obtain the exact confidence regions, e.g., the highest posterior density regions (see eqn. (2.81)), numerical integration techniques have to be applied. Similarly, numerical integration has to be applied to compute the posterior odds, eqn. (2.85), required for testing statistical hypotheses involving the estimated vector of variance-covariance components.

If the noninformative prior is used, then, as reported by Harville [1977], for the ordinary fixed ANOVA model, the GML estimation method produces estimated variance components having uniformly smaller mean squared error than both the ML and the REML estimates.

The GML estimated variance-covariance components are not, in general, unbiased.

With the exception of simple ANOVA models [Harville, 1977], the GML method has not yet been applied to the estimation of variance-covariance components. As the problem of GML estimation for the more general models has not been solved yet, it will be the subject of investigation in Chapter 3 of this thesis.



## 2.4 Criteria-Based Methods

All criteria-based methods approach the problem of estimation of the vector of variance-covariance components by seeking a quadratic estimator,  $\hat{\gamma} = \mathbf{l}^T \mathbf{M} \mathbf{l}$  (where  $\mathbf{M}$  is a symmetric matrix to be determined), of a linear combination of the variance-covariance components,  $\gamma = \mathbf{p}^T \boldsymbol{\theta}$  (where  $\mathbf{p}$  is a known vector), that minimizes a certain optimality criterion, such as the norm, the variance or the mean squared error, subject to some constraint, such as translation invariance or unbiasedness.

The quadratic estimator,  $\hat{\gamma} = \mathbf{l}^T \mathbf{M} \mathbf{l}$ , is said to be translation invariant if (see, e.g., [Rao and Kleffe, 1988]):

$$\mathbf{l}^T \mathbf{M} \mathbf{l} = (\mathbf{l} - \mathbf{A} \mathbf{x}_0)^T \mathbf{M} (\mathbf{l} - \mathbf{A} \mathbf{x}_0), \quad \forall \mathbf{x}_0 \in \mathbf{E}^m, \quad (2.91)$$

which is equivalent to:

$$\mathbf{M} \mathbf{A} = \mathbf{0} . \quad (2.92)$$

The translation invariance property is especially desirable in the case of geodetic networks where it implies that, regardless of the definition of the origin of the coordinate system in which the network is coordinated, the results of the estimation of variance-covariance components do not change.

The quadratic estimator,  $\hat{\gamma} = \mathbf{l}^T \mathbf{M} \mathbf{l}$ , is said to be unbiased for  $\gamma = \mathbf{p}^T \boldsymbol{\theta}$  if and only if (see, e.g., [Rao and Kleffe, 1988]):

$$\mathbf{E}(\mathbf{l}^T \mathbf{M} \mathbf{l}) = \mathbf{p}^T \boldsymbol{\theta}, \quad (2.93)$$

which is equivalent to:

$$\begin{cases} \mathbf{A}^T \mathbf{M} \mathbf{A} = \mathbf{0} \\ \text{tr}(\mathbf{M} \mathbf{T}_i) = p_i, \quad i = 1, \dots, r. \end{cases} \quad (2.94)$$

#### 2.4.1 Minimum Norm Quadratic Estimation (MINQE)

The Minimum Norm Quadratic Estimation method (MINQE) was developed by Rao [1971a]. Given the general mixed model, eqn. (2.1), the principle of the method is to minimize the norm of the difference between a quadratic estimator,  $\hat{\boldsymbol{\gamma}} = \mathbf{l}^T \mathbf{M} \mathbf{l}$ , of a linear combination of the variance-covariance components,  $\boldsymbol{\gamma} = \mathbf{p}^T \boldsymbol{\theta}$ , and the "natural estimator",  $\hat{\boldsymbol{\gamma}}_n = \mathbf{v}^T \Delta \mathbf{v}$ , of  $\boldsymbol{\gamma}$ . The difference between the quadratic estimator,  $\hat{\boldsymbol{\gamma}}$ , and the "natural estimator",  $\hat{\boldsymbol{\gamma}}_n$ , is given by Rao and Kleffe [1988] as:

$$\begin{aligned} \hat{\boldsymbol{\gamma}} - \hat{\boldsymbol{\gamma}}_n &= \mathbf{l}^T \mathbf{M} \mathbf{l} - \mathbf{v}^T \Delta \mathbf{v} = (\mathbf{A} \mathbf{x} + \mathbf{v})^T \mathbf{M} (\mathbf{A} \mathbf{x} + \mathbf{v}) - \mathbf{v}^T \Delta \mathbf{v} \\ &= \mathbf{x}^T \mathbf{A}^T \mathbf{M} \mathbf{A} \mathbf{x} + \mathbf{x}^T \mathbf{A}^T \mathbf{M} \mathbf{v} + \mathbf{v}^T \mathbf{M} \mathbf{A} \mathbf{x} + \mathbf{v}^T \mathbf{M} \mathbf{v} - \mathbf{v}^T \Delta \mathbf{v} \quad (2.95) \\ &= \begin{bmatrix} \mathbf{v} \\ \mathbf{x} \end{bmatrix}^T \begin{bmatrix} \mathbf{M} - \Delta & \mathbf{M} \mathbf{A} \\ \mathbf{A}^T \mathbf{M} & \mathbf{A}^T \mathbf{M} \mathbf{A} \end{bmatrix} \begin{bmatrix} \mathbf{v} \\ \mathbf{x} \end{bmatrix}. \end{aligned}$$

Minimization of the norm of the difference between the quadratic estimator,  $\hat{\boldsymbol{\gamma}}$ , and the "natural estimator",  $\hat{\boldsymbol{\gamma}}_n$ , eqn. (2.95), is equivalent [Rao and Kleffe, 1988] to minimization of the norm:

$$\left\| \begin{array}{cc} \mathbf{M} - \Delta & \mathbf{MA} \\ \mathbf{A}^T \mathbf{M} & \mathbf{A}^T \mathbf{MA} \end{array} \right\|. \quad (2.96)$$

The MINQE of  $\gamma = \mathbf{p}^T \theta$  is therefore a quadratic form  $\hat{\gamma} = \mathbf{l}^T \mathbf{M} \mathbf{l}$ , where  $\mathbf{M}$  is chosen to minimize the norm given by eqn. (2.96), for a suitable choice of norm (subsequently assumed to be the Euclidean norm).

By imposing constraints on the estimator  $\hat{\gamma}$ , we get the following types of MINQEs:

1. Unbiased: MINQE(U),
2. Invariant: MINQE(I),
3. Unbiased and invariant: MINQE(U,I).

For other types of MINQEs see [Rao and Kleffe, 1988].

The concept of the "natural estimator" is, in the general case, somewhat intuitive. Rao and Kleffe [1988] give the following development for the definition of the "natural estimator", starting from the simplest model being the Gauss-Markov model with the unknown variance factor  $\theta = \sigma_0^2$ :

$$\mathbf{l} = \mathbf{A} \mathbf{x} + \mathbf{v}, \quad \mathbf{v} \sim (\mathbf{0}, \mathbf{C}_\theta = \theta \mathbf{I}). \quad (2.97)$$

Under the condition that the vector of location parameters,  $\mathbf{x}$ , is known, the natural estimator of  $\theta$  is then given as:

$$\hat{\gamma}_n = \frac{\mathbf{v}^T \mathbf{v}}{m} = \mathbf{v}^T \left( \frac{1}{m} \mathbf{I} \right) \mathbf{v} = \mathbf{v}^T (\boldsymbol{\mu} \mathbf{I}) \mathbf{v} = \mathbf{v}^T \Delta \mathbf{v}. \quad (2.98)$$

The concept of the “natural estimator” can be easily extended to the following variance components model:

$$\mathbf{l} = \mathbf{A}\mathbf{x} + \mathbf{v}, \quad \text{with}$$

$$\mathbf{v} = (\mathbf{v}_1^T, \mathbf{v}_2^T, \dots, \mathbf{v}_r^T)^T, \quad \mathbf{v} \sim (\mathbf{0}, \mathbf{C}_\theta = \theta_1 \mathbf{T}_1 + \theta_2 \mathbf{T}_2 + \dots + \theta_r \mathbf{T}_r), \quad (2.99)$$

where  $\mathbf{T}_i$  is a diagonal matrix that has its  $m_i$  elements being equal to one on the diagonal, in positions whose indices correspond to the indices of the  $m_i$  elements of the sub-vector  $\mathbf{v}_i$  of the vector of unobservable random errors,  $\mathbf{v}$ , and zeros elsewhere. Under the condition that the vector of location parameters,  $\mathbf{x}$ , is known, the “natural estimator” of

the linear combination of the variance components,  $\gamma = \sum_{i=1}^r p_i \theta_i = \mathbf{p}^T \boldsymbol{\theta}$ , is

$$\hat{\gamma}_n = \sum_{i=1}^r \left( p_i \frac{\mathbf{v}_i^T \mathbf{v}_i}{m_i} \right) = \sum_{i=1}^r \left( \frac{p_i}{m_i} \mathbf{v}^T \mathbf{T}_i \mathbf{v} \right) = \sum_{i=1}^r (\boldsymbol{\mu}_i \mathbf{v}^T \mathbf{T}_i \mathbf{v}) = \mathbf{v}^T \left[ \sum_{i=1}^r (\boldsymbol{\mu}_i \mathbf{T}_i) \right] \mathbf{v} = \mathbf{v}^T \Delta \mathbf{v}, \quad (2.100)$$

where the vector  $\boldsymbol{\mu} = (\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \dots, \boldsymbol{\mu}_r)^T$  can be thought of as a solution to the following linear equation:

$$\left\{ \sum_{i,j=1}^r \text{tr}(\mathbf{T}_i \mathbf{T}_j) \right\} \boldsymbol{\mu} = \mathbf{p}, \quad (2.101)$$

or, taking into account eqn.(2.99),

$$\left\{ \sum_{i=1}^r m_i \right\} \boldsymbol{\mu} = \mathbf{p}. \quad (2.102)$$

By analogy to the variance components model given by eqn. (2.99), the concept of the “natural estimator” is finally extended to the general mixed model (eqn. (2.1)).

Under the assumption that the vector of location parameters,  $\mathbf{x}$ , is known, the “natural estimator” of the linear combination of the variance-covariance components,

$\gamma = \sum_{i=1}^r p_i \theta_i = \mathbf{p}^T \boldsymbol{\theta}$ , is given as:

$$\hat{\gamma}_n = \sum_{i=1}^r (\mu_i \mathbf{v}^T \mathbf{T}_i \mathbf{v}) = \mathbf{v}^T \left[ \sum_{i=1}^r (\mu_i \mathbf{T}_i) \right] \mathbf{v} = \mathbf{v}^T \Delta \mathbf{v}, \quad (2.103)$$

where the vector  $\boldsymbol{\mu} = (\mu_1, \mu_2, \dots, \mu_r)^T$  is a solution to the following linear equation:

$$\left\{ \sum_{i,j=1}^r \text{tr}(\mathbf{T}_i \mathbf{T}_j) \right\} \boldsymbol{\mu} = \mathbf{p}. \quad (2.104)$$

The MINQE(U,I) - Unbiased and Invariant - of  $\gamma$  is obtained by minimizing the Euclidean norm given by eqn. (2.96) subject to the condition of unbiasedness, eqn. (2.94), and invariance, eqn. (2.92). Minimization of this norm results in the set of linear equations, given by Rao [1979] as:

$$\mathbf{S}_{\text{UI}\tilde{\boldsymbol{\theta}}} \boldsymbol{\theta} = \mathbf{q}_{\text{UI}\tilde{\boldsymbol{\theta}}}, \quad (2.105)$$

where:

$$\mathbf{S}_{\text{UI}\tilde{\boldsymbol{\theta}}} = \left\{ \sum_{k,m=1}^r \text{tr}(\mathbf{R}_{\tilde{\boldsymbol{\theta}}_k} \mathbf{T}_k \mathbf{R}_{\tilde{\boldsymbol{\theta}}_m} \mathbf{T}_m) \right\}, \quad (2.106)$$

$$\mathbf{q}_{U|\tilde{\theta}} = \left\{ \mathbf{c}^T \mathbf{R}_{\tilde{\theta}}^{-1} \mathbf{T}_k \mathbf{R}_{\tilde{\theta}}^{-1} \mathbf{1} \right\}_{k=1}^r, \text{ and} \quad (2.107)$$

$$\mathbf{R}_{\tilde{\theta}} = \mathbf{C}_{\tilde{\theta}}^{-1} \left[ \mathbf{I} - \mathbf{A} \left( \mathbf{A}^T \mathbf{C}_{\tilde{\theta}}^{-1} \mathbf{A} \right)^{-1} \mathbf{A}^T \mathbf{C}_{\tilde{\theta}}^{-1} \right], \quad (2.108)$$

with  $\tilde{\theta}$  being the initial vector of approximate variance-covariance components. The MINQE(U,I) of  $\gamma = \mathbf{p}^T \theta$  is  $\hat{\gamma} = \mathbf{p}^T \hat{\theta}$ , where  $\hat{\theta}$  is a solution to eqn. (2.105).

The MINQE(U) - Unbiased - of  $\gamma$  is obtained by minimizing the Euclidean norm given by eqn. (2.96) subject to the condition of unbiasedness, eqn. (2.94). This minimization results in the set of linear equations, given by Rao [1979] as:

$$\mathbf{S}_{U\tilde{\theta}} \theta = \mathbf{q}_{U\tilde{\theta}}, \quad (2.109)$$

where:

$$\mathbf{S}_{U\tilde{\theta}} = \left\{ \sum_{m=1}^r \text{tr} \left[ \left( \mathbf{C}_{\tilde{\theta}} + \mathbf{A} \mathbf{K} \mathbf{A}^T \right)^{-1} \left( \mathbf{T}_k - \mathbf{P}_{\tilde{\theta}} \mathbf{T}_k \mathbf{P}_{\tilde{\theta}}^T \right) \left( \mathbf{C}_{\tilde{\theta}} + \mathbf{A} \mathbf{K} \mathbf{A}^T \right)^{-1} \mathbf{T}_m \right] \right\}_{k,m=1}^r, \quad (2.110)$$

$$\mathbf{q}_{U\tilde{\theta}} = \left\{ \mathbf{c}^T \left( \mathbf{I} - \mathbf{A} \mathbf{x}_0 \right)^T \left( \mathbf{C}_{\tilde{\theta}} + \mathbf{A} \mathbf{K} \mathbf{A}^T \right)^{-1} \left( \mathbf{T}_k - \mathbf{P}_{\tilde{\theta}} \mathbf{T}_k \mathbf{P}_{\tilde{\theta}}^T \right) \left( \mathbf{C}_{\tilde{\theta}} + \mathbf{A} \mathbf{K} \mathbf{A}^T \right)^{-1} \left( \mathbf{I} - \mathbf{A} \mathbf{x}_0 \right) \right\}_{k=1}^r, \text{ and} \quad (2.111)$$

$$\mathbf{P}_{\tilde{\theta}} = \mathbf{A} \left[ \mathbf{A}^T \left( \mathbf{C}_{\tilde{\theta}} + \mathbf{A} \mathbf{K} \mathbf{A}^T \right)^{-1} \mathbf{A} \right]^{-1} \mathbf{A}^T \left( \mathbf{C}_{\tilde{\theta}} + \mathbf{A} \mathbf{K} \mathbf{A}^T \right)^{-1}, \quad (2.112)$$

with  $\mathbf{x}_0$  and  $\mathbf{K}$  being the prior value and dispersion of  $\mathbf{x}$ , respectively. The MINQE(U) of  $\gamma = \mathbf{p}^T \theta$  is  $\hat{\gamma} = \mathbf{p}^T \hat{\theta}$ , where  $\hat{\theta}$  is a solution to eqn. (2.109).

The MINQE(I) - Invariant - of  $\gamma$  is obtained by minimizing the Euclidean norm given by eqn (2.96) subject to the condition of invariance, eqn. (2.92). Minimization of this norm results in the set of linear equations, given by Rao [1979] as:

$$\mathbf{S}_{I\hat{\theta}}\theta = \mathbf{q}_{UI\hat{\theta}}, \quad (2.113)$$

where:

$$\mathbf{S}_{I\hat{\theta}} = \left\{ \sum_{k,m=1}^r \text{tr} \left( \mathbf{C}_{\hat{\theta}}^{-1} \mathbf{T}_k \mathbf{C}_{\hat{\theta}}^{-1} \mathbf{T}_m \right) \right\}^r. \quad (2.114)$$

The MINQE(I) of  $\gamma = \mathbf{p}^T \theta$  is  $\hat{\gamma} = \mathbf{p}^T \hat{\theta}$ , where  $\hat{\theta}$  is a solution to eqn. (2.113).

The MINQE estimates, which are computed with the initial vector of approximate variance-covariance components, may be used as a new set of approximate values for recomputing MINQE. The limiting estimator, if it exists, is called IMINQE (Iterated MINQE). The iterative procedure yields estimators that are, in general, not dependent on the choice of initial values. The  $(i+1)$ -th iterated MINQE is a solution of

$$\mathbf{S}_{\hat{\theta}_{(i)}} \hat{\theta}_{(i+1)} = \mathbf{q}_{\hat{\theta}_{(i)}}, \quad (2.115)$$

where:  $\hat{\theta}_{(i)}$  is the MINQE from the  $i$ -th iteration,

$\mathbf{S}$  is identical to either  $\mathbf{S}_{UI}$ , eqn. (2.106),  $\mathbf{S}_U$ , eqn. (2.110), or  $\mathbf{S}_T$ , eqn. (2.114), and

$\mathbf{q}$  is identical to either  $\mathbf{q}_{UI}$ , eqn. (2.107),  $\mathbf{q}_U$ , eqn. (2.111), or  $\mathbf{q}_{UI}$ , respectively.

The limiting value, if it exists, satisfies the equation:

$$\mathbf{S}_{\hat{\theta}} \hat{\theta} = \mathbf{q}_{\hat{\theta}}, \quad (2.116)$$

which is called the IMINQE equation. All three types of MINQE: (U,I), (U), or (I), can be solved for by applying the iterative procedure shown in eqn. (2.115). The limiting estimators are called IMINQE(U,I), IMINQE(U), and IMINQE(I), respectively.

The IMINQE(U,I) may exist as a solution to eqn. (2.116), although it does not have to be approached through the sequence of estimates following the iterative procedure shown in eqn. (2.115). An alternative computational scheme exists (see [Rao, 1979] and [Rao and Kleffe, 1988]), which preserves the sign of the variance-covariance components at each iteration step. This procedure, called the Iterated Almost Unbiased Estimation (IAUE) (cf. Subsection 2.4.4), gives the estimate of the k-th variance-covariance component, resulting from the (i+1)-th iterative step, as:

$$\hat{\theta}_{(i+1),k} = \frac{\hat{\theta}_{(i),k}}{\text{tr} \left( \mathbf{R}_{\hat{\theta}_{(i)}} \mathbf{T}_k \right)} \mathbf{I}^T \mathbf{R}_{\hat{\theta}_{(i)}} \mathbf{T}_k \mathbf{R}_{\hat{\theta}_{(i)}} \mathbf{I}. \quad (2.117)$$

The IMINQE(U,I) estimators are globally best, invariant, almost unbiased, asymptotically normal, consistent (provided MINQE(U,I) itself is consistent), and efficient in the class of invariant asymptotically normal estimators. The asymptotic covariance matrix of the IMINQE(U,I) of the vector of variance-covariance components is given by Rao and Kleffe [1988] as:



$$\text{var}(\hat{\theta}) = 2\mathbf{S}_{UI\hat{\theta}}^{-1} . \quad (2.118)$$

Numerically, IMINQE(U,I) is identical to the REML estimation and to the Iterated BIQUE. IMINQE(I) is numerically equivalent to the ML estimation of the variance-covariance components.

#### ***2.4.2 Best Quadratic Estimation (BQE)***

The general algorithms for the Best Quadratic Estimation (BQE) of variance-covariance components in the general mixed model, eqn. (2.1), were developed by LaMotte [1973]. The Best Quadratic Estimation method is a generalization of the concept of Best Quadratic Unbiased Estimation (BQUE) by Townsend and Searle [1971].

The problem at hand is to find a quadratic estimator,  $\hat{\gamma} = \mathbf{I}^T \mathbf{M} \mathbf{I}$ , (where  $\mathbf{M}$  is a symmetric matrix to be determined) of a linear combination of variance-covariance components,  $\gamma = \mathbf{p}^T \theta$ , (where  $\mathbf{p}$  is a known vector) that has minimum mean squared error. The mean squared error of an estimator  $\hat{\gamma}$  of  $\gamma$  is defined (see, e.g., Read [1985]) as:

$$\text{MSE}(\hat{\gamma}) = \text{var}(\hat{\gamma}) + [\mathbf{E}(\hat{\gamma}) - \gamma]^2 = \text{var}(\hat{\gamma}) + b^2, \quad (2.119)$$

where  $b$  is the bias.

Under the assumption that the vector of observations,  $\mathbf{l}$ , is normally distributed [LaMotte, 1973], the objective function, being the mean squared error of  $\hat{\gamma}$ , has the following form:

$$\text{MSE}(\mathbf{1}^T \mathbf{M} \mathbf{I}) = 2\text{tr}(\mathbf{C}_\theta \mathbf{M})^2 + 4\mathbf{x}^T \mathbf{A}^T \mathbf{M} \mathbf{C}_\theta \mathbf{M} \mathbf{A} \mathbf{x} + [\text{tr}(\mathbf{C}_\theta \mathbf{M}) + \mathbf{x}^T \mathbf{A}^T \mathbf{M} \mathbf{A} \mathbf{x} - \mathbf{p}^T \boldsymbol{\theta}]^2. \quad (2.120)$$

The mean squared error of  $\hat{\gamma}$ , eqn. (2.120), is subsequently minimized subject to the conditions of unbiasedness, eqn. (2.94), or both invariance, eqn. (2.92), and unbiasedness.

Under the condition of unbiasedness (eqn.(2.94)) the bias term in eqn. (2.120) disappears and the mean squared error of  $\hat{\gamma}$  becomes identical to the variance of  $\hat{\gamma}$ :

$$\text{MSE}(\mathbf{1}^T \mathbf{M} \mathbf{I}) = \text{var}(\mathbf{1}^T \mathbf{M} \mathbf{I}) = 2\text{tr}(\mathbf{C}_\theta \mathbf{M})^2 + 4\mathbf{x}^T \mathbf{A}^T \mathbf{M} \mathbf{C}_\theta \mathbf{M} \mathbf{A} \mathbf{x}. \quad (2.121)$$

The Best Quadratic Unbiased Estimator (BQUE) of  $\gamma$  is obtained by minimizing the mean squared error, eqn. (2.120), subject to the condition of unbiasedness, eqn. (2.94).

Minimization results in the following set of linear equations, given by Searle et al. [1992] as:

$$\mathbf{S}_{\text{BU}\tilde{\theta}} \boldsymbol{\theta} = \mathbf{q}_{\text{BU}\tilde{\theta}}, \quad (2.122)$$

where:

$$\mathbf{S}_{\text{BU}\tilde{\theta}} = \left\{ \text{tr}(\mathbf{M}_{k\tilde{\theta}} \mathbf{T}_m) \right\}_{k,m=1}^r, \quad (2.123)$$

$$\mathbf{q}_{\text{BU}\tilde{\theta}} = \left\{ \mathbf{1}^T \mathbf{M}_{k\tilde{\theta}} \tilde{\mathbf{1}} \right\}_{k=1}^r, \quad (2.124)$$

$$\mathbf{M}_{k\tilde{\theta}} = \mathbf{R}_{\tilde{\theta}} \mathbf{T}_k \mathbf{R}_{\tilde{\theta}} + \mathbf{R}_{\tilde{\theta}} \mathbf{T}_k \mathbf{B}_{\tilde{\theta}}^- + \mathbf{B}_{\tilde{\theta}}^- \mathbf{T}_k \mathbf{R}_{\tilde{\theta}}, \quad \text{and} \quad (2.125)$$

$$\mathbf{B}_{\tilde{\theta}}^- = \mathbf{C}_{\tilde{\theta}}^{-1} - \mathbf{R}_{\tilde{\theta}} - \frac{\mathbf{C}_{\tilde{\theta}}^{-1} \mathbf{A} \mathbf{x}_0 \mathbf{C}_{\tilde{\theta}}^{-1} \mathbf{x}_0^T \mathbf{A}^T \mathbf{C}_{\tilde{\theta}}^{-1}}{1 + \mathbf{x}_0^T \mathbf{A}^T \mathbf{A} \mathbf{x}_0}, \quad (2.126)$$

with  $\tilde{\theta}$  and  $\mathbf{x}_0$  being the pre-assigned values for the vector of variance-covariance components and the vector of location parameters, respectively. The BQUE of  $\gamma = \mathbf{p}^T \theta$  is  $\hat{\gamma} = \mathbf{p}^T \hat{\theta}$ , where  $\hat{\theta}$  is a solution to eqn. (2.122).

The Best Invariant Quadratic Unbiased Estimator (BIQUE) of  $\gamma$  is obtained by minimizing the mean squared error, eqn. (2.120), subject to the condition of unbiasedness, eqn. (2.94), and invariance, eqn. (2.92). Under the condition of invariance, the second term in the expression for the mean squared error disappears, and the objective function to be minimized is of the following form:

$$\text{MSE}(\mathbf{1}^T \mathbf{M} \mathbf{l}) = 2 \text{tr}(\mathbf{C}_\theta \mathbf{M})^2. \quad (2.127)$$

Minimization of eqn. (2.127) results in the set of linear equations, given by Caspary [1987] and Searle et al. [1992] as:

$$\mathbf{S}_{\text{UI}\tilde{\theta}} \theta = \mathbf{q}_{\text{UI}\tilde{\theta}}, \quad (2.128)$$

where  $\mathbf{S}_{\text{UI}\tilde{\theta}}$  is given by eqn. (2.106) and  $\mathbf{q}_{\text{UI}\tilde{\theta}}$  is given by eqn. (2.107). The BIQUE of  $\gamma = \mathbf{p}^T \theta$  is  $\hat{\gamma} = \mathbf{p}^T \hat{\theta}$ , where  $\hat{\theta}$  is a solution to eqn. (2.128). Numerically, BIQUE is identical to MIVQUE (under the assumption that the vector of observations is normal), and to MINQE(U,I).

The BIQUE estimator, which is computed with a pre-assigned vector of approximate variance-covariance components, is only locally best. In order to obtain globally best estimate, an iterative procedure must be employed. Estimates of the variance-covariance

components from the initial step are used as the new approximate values for the next iteration step. The iteration process is continued until convergence is achieved. The Iterated BIQUE is numerically equivalent to IMINQE(U,I), and to the REML estimation.

### ***2.4.3 Minimum Variance Quadratic Unbiased Estimation (MIVQUE)***

The Minimum Variance Quadratic Unbiased Estimation algorithm can be found in [Rao, 1971b]. Given the general mixed model, eqn. (2.1), the principle of MIVQUE [Swallow and Searle, 1978] is to find an unbiased and invariant quadratic estimator,  $\hat{\gamma} = \mathbf{I}^T \mathbf{M} \mathbf{I}$ , (where  $\mathbf{M}$  is a symmetric matrix to be determined) of a linear combination of variance-covariance components,  $\gamma = \mathbf{p}^T \theta$ , (where  $\mathbf{p}$  is a known vector) that has the minimum variance property:

$$\text{var}(\mathbf{I}^T \mathbf{M} \mathbf{I}) = 2\text{tr}(\mathbf{C}_\theta \mathbf{M})^2 + \left( \begin{array}{l} \text{a term depending on } \mathbf{M} \\ \text{and the kurtosis parameters} \end{array} \right) = \text{minimum}. \quad (2.129)$$

In the general case, i.e., with the vector of observations having unspecified distribution, the variance of the quadratic form  $\hat{\gamma} = \mathbf{I}^T \mathbf{M} \mathbf{I}$  is a function of the moments of the vector of observations - up to the fourth order.

If the condition of normality is additionally imposed on the vector of observations, then the kurtosis parameters are identical to zero [Swallow and Searle, 1978], and the variance function to be minimized takes the following form:

$$\text{var}(\mathbf{I}^T \mathbf{M} \mathbf{I}) = 2\text{tr}(\mathbf{C}_\theta \mathbf{M})^2. \quad (2.130)$$

Under the assumption that the vector of observations is normally distributed the variance of this estimator is identical to the mean squared error given by eqn. (2.127), and the Euclidean norm given by eqn. (2.96). As a result, MIVQUE is numerically equivalent to BIQUE, and to MINQE(U,I) [Swallow and Monahan, 1984].

#### ***2.4.4 Almost Unbiased Estimation (AUE)***

The concept of Almost Unbiased Estimation (AUE) was developed by Horn et al. [1975] and Horn and Horn [1975]. The AUE [Rao and Kleffe, 1988] relies on the principle of minimization of the following criterion:

$$f(\mathbf{M}) = 2\text{tr}(\mathbf{C}_\theta \mathbf{M})^2, \quad (2.131)$$

subject to the condition of invariance, eqn. (2.99), and "almost" unbiasedness:

$$\begin{cases} \mathbf{A}^T \mathbf{M} \mathbf{A} = \mathbf{0} \\ \text{tr}(\mathbf{M} \mathbf{C}_{\tilde{\theta}}) = \mathbf{p} \tilde{\theta}, \end{cases} \quad (2.132)$$

where  $\tilde{\theta}$  is the initial vector of approximate variance-covariance components.

If the initial, approximate values of the variance-covariance components are proportional to the true values, then AUE is unbiased [Lucas et al., 1985]. In reality, this condition is never satisfied, and the estimators are always biased. The bias, however, always remains relatively small - hence the name: Almost Unbiased Estimator.

The AUE of the k-th variance-covariance component is given by Förstner [1979a and 1979b] and by Rao and Kleffe [1988] as:

$$\hat{\theta}_k = \frac{\tilde{\theta}_k}{\text{tr}(\mathbf{R}_{\tilde{\theta}} \mathbf{T}_k)} \mathbf{I}^T \mathbf{R}_{\tilde{\theta}} \mathbf{T}_k \mathbf{R}_{\tilde{\theta}} \mathbf{I}, \quad (2.133)$$

where  $\tilde{\theta}_k$  is the k-th component of the vector of initial, approximate values of variance-covariance components.

The AUE estimates can be used as approximate starting values for the next iteration. The limiting estimator, if it exists, is called the Iterated Almost Unbiased Estimator (IAUE) (cf. Subsection 2.4.1). The IAUE estimate of the k-th variance-covariance component resulting from the (i+1)-th iterative step is expressed (see, e.g., [Rao and Kleffe, 1988]) as:

$$\hat{\theta}_{(i+1),k} = \frac{\hat{\theta}_{(i),k}}{\text{tr}(\mathbf{R}_{\hat{\theta}_{(i)}} \mathbf{T}_k)} \mathbf{I}^T \mathbf{R}_{\hat{\theta}_{(i)}} \mathbf{T}_k \mathbf{R}_{\hat{\theta}_{(i)}} \mathbf{I}. \quad (2.134)$$

An interesting property of IAUE is that the variance-covariance components do not change sign through the entire iteration process [Horn and Horn, 1975]. This property is especially attractive in the context of estimation of variance components, which are by definition non-negative, and would remain non-negative if the iterative process started with non-negative values. The IAUE can be regarded as an alternative computational method for solving the REML scoring equations, eqn. (2.63), or, equivalently, for

obtaining the  $IMINQE(U,I)$  - if the solution exists. The IAUE is slightly more computationally efficient than the Fisher scoring method when applied to REML estimation method, eqn. (2.63), or, equivalently, than the  $IMINQE(U,I)$  iterative procedure given by eqn. (2.125), as it does not require inversion of an  $r$  by  $r$  matrix at each iteration step. It should be pointed out that if  $r$  (the number of variance-covariance components) stays relatively small, then the gain in computational efficiency is insignificant. Moreover, the IAUE does not provide the asymptotic covariance matrix of the estimated variance-covariance components at each iteration step.

## **2.5 Comparison of the Reviewed Methods**

Despite the fact that there exist a substantial number of various statistical methods for estimation of variance-covariance components for unbalanced data, most of these methods are numerically equivalent. In other words, even though each method is derived from different basic principles, the computational algorithms, and hence the numerical results, are, under certain restrictions, the same.

As far as the computational procedures are concerned, the broadest classification of the discussed methods is classification into non-iterative and iterative methods.

The non-iterative methods include: (i) general ANOVA methods with the Helmert-type estimation method as an example, (ii) MINQE methods, (iii) BQE methods, (iv) MIVQUE method, and (v) AUE method. The following relationships exist between these methods:

1. MINQE(U,I) is numerically equivalent to BIQUE, to MIVQUE under the restriction of normality of observations to AUE (under the assumption that the initial, approximate values of the variance-covariance components are proportional to their true values), to the Helmert-type estimation (for variance components only), and to the first step REML estimation.
2. MINQE(I) is numerically equivalent to the first step ML estimation.

The iterative methods include: (i) the iterated Helmert-type estimation method, (ii) the ML method, (iii) the REML method, (iv) GML methods, (v) Iterated MINQE methods, (vi) Iterated BQE methods, and (vii) the IAUE method. The following relationships exist between these methods:

1. REML estimation is numerically equivalent to IMINQE(U,I), to iterated BIQUE, IAUE (which may be regarded as one of the algorithms for solution of REML), and to the iterated Helmert-type estimation (for variance components only).
2. ML estimation is numerically equivalent to IMINQE(I).

From all the presented methods, only the Bayesian methods make a provision for introduction of prior information about the unknown variance-covariance components. While the computational approach of the GML method is similar to that of the maximum likelihood methods (ML and REML), the other two Bayesian methods, i.e., the Bayes estimation and the interval estimation (see Subsection 2.3.2 and Subsection 2.3.3), employ completely different computational procedures. It should be noted that both of these



methods (Bayes estimation and interval estimation) are computationally involved, and therefore impractical for all but trivial cases (cf. Subsection 2.3.4).

As far as the general ANOVA methods are concerned, because of their non-uniqueness and lack of optimal properties, they should be regarded as inferior to the other presented methods for estimating variance-covariance components for unbalanced data.

## CHAPTER 3

### **INTRODUCTION OF PRIOR INFORMATION: GML ESTIMATION OF VARIANCE COMPONENTS AND THE DISPERSION-MEAN MODEL WITH WEIGHTED CONSTRAINTS**

As shown in Section 2.3, application of the Bayesian methodology to the estimation of variance-covariance components results in the posterior probability density function,  $p(\theta|I)$ , of the vector of variance-covariance components  $\theta$ . This posterior probability density function constitutes the foundation for subsequent estimation of  $\theta$ . Neither the Bayes estimation (see Subsection 2.3.2) nor the interval estimation (see Subsection 2.3.3) methods can be solved analytically. In general, they require extensive application of numerical integration techniques which, because these techniques are computationally intensive, makes them practicable only for variance-covariance models that involve a limited number of components. The Generalized Maximum Likelihood (GML) methodology, on the other hand, is much simpler computationally, as it does not require numerical integration to get the point estimate of the vector of variance-covariance components.

The principle of the Generalized Maximum Likelihood estimation is applied in this chapter to the estimation of the variance components in the following variance components model:

$$\mathbf{l} = \mathbf{A}\mathbf{x} + \mathbf{v}, \quad \mathbf{v} \sim (\mathbf{0}, \mathbf{C}_\theta = \theta_1 \mathbf{T}_1 + \theta_2 \mathbf{T}_2 + \dots + \theta_r \mathbf{T}_r), \quad (3.1)$$

where:  $\mathbf{l}$  is a vector of  $m$  observations,

$\mathbf{A}$  is a given first order design matrix,

$\mathbf{x}$  is a vector of  $u$  unknown location parameters,

$\mathbf{v}$  is a vector of  $m$ , normally distributed, unobservable random errors,

$\mathbf{T}_1, \dots, \mathbf{T}_r$  are given, linearly independent, positive semi-definite diagonal matrices, and

$\theta = (\theta_1, \dots, \theta_r)$  is a vector of  $r$  unknown positive variance components.

In Sections 3.1 and 3.2, under the assumption that the constraints on the parameter space, implied by the requirement of positiveness of the variance components, are ignored, the Fisher scoring method is used to find a solution to the following non-linear unconstrained GML optimization problem:

$$l_P(\hat{\theta}) = \sup_{\theta} l_P, \quad (3.2)$$

where  $l_P$  is the natural logarithm of the posterior probability density function. Two types of priors are used in this approach. First, the inverted gamma probability density function, eqn. (2.74), is used as an informative prior. Then, the Jeffrey's prior, eqn. (2.71), which is proportional to the square root of the determinant of the information matrix associated with the marginal likelihood function, eqn. (2.44), is used as a noninformative prior: this

expressing the lack of prior knowledge about the variance components. As the application of the Fisher scoring method requires that the expressions for both the information matrix and the vector of scores of the posterior probability density function are known, these two expressions - for both types of priors - are derived in this chapter.

In Section 3.3, the principle of mixed estimation (see, e.g., [Bossler, 1972], [Belsley et al., 1980], [Vanicek and Krakiwsky, 1986] or [Wells, 1990]) - which allows, in the least squares adjustment method, the introduction of prior information about the vector of location parameters,  $\mathbf{x}$ , in the form of weighted constraints - is extended to the dispersion-mean model. The dispersion-mean model reformulates the general mixed model, eqn. (2.1), in such a way that the resulting model is linear in terms of the vector of variance-covariance components [Searle et al., 1992]. A link between the mixed estimation of the dispersion-mean model, i.e., estimation of the dispersion-mean model with weighted constraints, and the GML methodology is subsequently investigated. Finally, the form of a hypothetical prior probability density function,  $p(\theta)$ , that would result in the estimating equations of the dispersion-mean model with weighted constraints, if the Fisher scoring method were applied to the GML optimization problem with such prior, is derived.

With the exception of the dispersion-mean model with weighted constraints the covariance components are not investigated in this chapter.

To simplify the notation, in this chapter in all subsequent formulae,  $\mathbf{R}_\theta$  will be replaced by  $\mathbf{R}$  and  $\mathbf{P}_\theta$  will be replaced by  $\mathbf{P}$ .

### 3.1 GML Estimation with Inverted Gamma Prior

The prior distribution function of the vector of variance components is represented by the inverted gamma distribution (cf. Section 2.3.1) in this chapter.

Assuming that the variance components are independent, the inverted gamma probability density function has the following form:

$$p(\theta) \propto \prod_{i=1}^r \left[ \left( \frac{1}{\theta_i} \right)^{p_i+1} \exp\left(-\frac{b_i}{\theta_i}\right) \right] \quad (3.3)$$

with:  $b_i > 0$ ,  $p_i > 0$ , and  $0 < \theta_i < \infty$ , for  $i = 1, 2, \dots, r$ ,

$$p_i = \frac{\theta_{0_i}^2}{V_{\theta_{0_i}}} + 2, \text{ and}$$

$$b_i = (p_i - 1)\theta_{0_i},$$

where  $\theta_{0_i}$  is the prior value (mean of the prior distribution) of the  $i$ -th variance

component, and

$V_{\theta_{0_i}}$  is the variance of the prior distribution for the  $i$ -th variance component.

The posterior probability density function results from application of the Bayes' Theorem eqn. (2.65), which with the marginal likelihood function  $p(\mathbf{l}|\theta) = L_R$ , eqn. (2.44), gives:

$$\begin{aligned}
p(\boldsymbol{\theta}|\mathbf{l}) &\propto p(\boldsymbol{\theta})p(\mathbf{l}|\boldsymbol{\theta}) = p(\boldsymbol{\theta})L_{\mathbf{R}} \\
&\propto \prod_{i=1}^r \left[ \left( \frac{1}{\theta_i} \right)^{p_i+1} \exp\left(-\frac{b_i}{\theta_i}\right) \right] \left( |\mathbf{C}_{\boldsymbol{\theta}}| |\mathbf{A}^T \mathbf{C}_{\boldsymbol{\theta}}^{-1} \mathbf{A}| \right)^{-\frac{1}{2}} \exp\left(-\frac{\mathbf{l}^T \mathbf{R} \mathbf{l}}{2}\right).
\end{aligned} \tag{3.4}$$

Taking the natural logarithms of both sides we get:

$$\begin{aligned}
l_{\mathbf{P}} &= \ln[p(\boldsymbol{\theta}|\mathbf{l})] = c + \ln[p(\boldsymbol{\theta})] + \ln(L_{\mathbf{R}}) = c + \ln[p(\boldsymbol{\theta})] + l_{\mathbf{R}} \\
&= c - \sum_{i=1}^r [(p_i + 1)\ln(\theta_i)] - \sum_{i=1}^r \left( \frac{b_i}{\theta_i} \right) - \frac{1}{2} \ln|\mathbf{C}_{\boldsymbol{\theta}}| - \frac{1}{2} \ln|\mathbf{A}^T \mathbf{C}_{\boldsymbol{\theta}}^{-1} \mathbf{A}| - \frac{\mathbf{l}^T \mathbf{R} \mathbf{l}}{2},
\end{aligned} \tag{3.5}$$

where  $c$  is some constant.

### 3.1.1 Derivation of the Vector of Scores

The expression for the vector of scores, which is defined as a vector of  $r$  partial first derivatives of  $l_{\mathbf{P}}$  with respect to the variance components  $\theta_k$ :

$$\mathbf{s}_{\mathbf{P}} = \left\{ c \frac{\partial l_{\mathbf{P}}}{\partial \theta_k} \right\}_{k=1}^r = \frac{\partial l_{\mathbf{P}}}{\partial \boldsymbol{\theta}}, \tag{3.6}$$

is derived from eqn. (3.5).

As seen in eqn. (3.5), the vector of scores can be decomposed into the sum of the vectors of first derivatives of the natural logarithm of the prior,  $\ln[p(\boldsymbol{\theta})]$  and the marginal log-likelihood function  $l_{\mathbf{R}}$ , eqn. (2.45), as:

$$\mathbf{s}_{\mathbf{P}} = \frac{\partial l_{\mathbf{P}}}{\partial \boldsymbol{\theta}} = \frac{\partial \ln[p(\boldsymbol{\theta})]}{\partial \boldsymbol{\theta}} + \frac{\partial l_{\mathbf{R}}}{\partial \boldsymbol{\theta}} = \mathbf{s}_{\mathbf{P}0} + \mathbf{s}_{\mathbf{R}}, \tag{3.7}$$

where the first derivative of the marginal log-likelihood function is given by Koch [1990] as:

$$\mathbf{s}_R = \frac{\partial \ln \mathbf{L}_R}{\partial \boldsymbol{\theta}} = \left\{ c \left( -\frac{1}{2} \text{tr}(\mathbf{R}\mathbf{T}_k) + \frac{1}{2} \mathbf{1}^T \mathbf{R}\mathbf{T}_k \mathbf{R}\mathbf{1} \right) \right\}_{k=1}^r. \quad (3.8)$$

To get an expression for the vector of scores, the first derivative of the natural logarithm of the prior probability density function,  $\ln[p(\boldsymbol{\theta})]$ , is derived as follows.

The partial first derivative of the natural logarithm of  $p(\boldsymbol{\theta})$  with respect to the  $k$ -th variance component is

$$\frac{\partial \ln[p(\boldsymbol{\theta})]}{\partial \theta_k} = \frac{b_k}{\theta_k^2} - \frac{p_k + 1}{\theta_k}. \quad (3.9)$$

As a result, the first derivative of the natural logarithm of the prior probability density function,  $\ln[p(\boldsymbol{\theta})]$ , with respect to the vector of variance components, is:

$$\mathbf{s}_{P0} = \frac{\partial \ln[p(\boldsymbol{\theta})]}{\partial \boldsymbol{\theta}} = \left\{ c \left( \frac{b_k}{\theta_k^2} - \frac{p_k + 1}{\theta_k} \right) \right\}_{k=1}^r. \quad (3.10)$$

Taking into account the results of eqn. (3.7), eqn. (3.8), and eqn. (3.9) the vector of scores is finally given as:

$$\begin{aligned}
\mathbf{s}_P &= \frac{\partial \mathbf{l}_P}{\partial \boldsymbol{\theta}} = \mathbf{s}_{P0} + \mathbf{s}_R = \left\{ c \left( \frac{\mathbf{b}_k}{\theta_k^2} - \frac{p_k + 1}{\theta_k} \right) \right\}_{k=1}^r + \left\{ c - \frac{1}{2} \text{tr}(\mathbf{R}\mathbf{T}_k) + \frac{1}{2} \mathbf{1}^T \mathbf{R}\mathbf{T}_k \mathbf{R}\mathbf{1} \right\}_{k=1}^r \\
&= \mathbf{s}_{P0} - \left\{ c \left( \frac{1}{2} \text{tr}(\mathbf{R}\mathbf{T}_k) \right) \right\}_{k=1}^r + \mathbf{q}_R,
\end{aligned} \tag{3.11}$$

where  $\mathbf{s}_R$  is the vector of marginal likelihood scores (see eqn. (2.56)) and  $\mathbf{q}_R$  is the vector of quadratic forms of observations in the REML scoring equations (see eqn. (2.63)).

### 3.1.2 Derivation of the Information Matrix

The expression for the vector of scores,  $\mathbf{s}_P$ , eqn. (3.16), is utilized in this chapter to derive the Hessian matrix,  $\mathbf{H}_P$ , which is defined as an  $r$  by  $r$  matrix of second partial derivatives of the natural logarithm of the posterior probability density function,  $\mathbf{l}_P$ , with respect to the  $k$ -th and  $m$ -th variance components,  $\theta_k$  and  $\theta_m$ :

$$\mathbf{H}_P = \left\{ \frac{\partial^2 \mathbf{l}_P}{\partial \theta_k \partial \theta_m} \right\}_{k,m=1}^r = \frac{\partial^2 \mathbf{l}_P}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}} = \frac{\partial \mathbf{s}_P}{\partial \boldsymbol{\theta}}. \tag{3.12}$$

Subsequently, the information matrix,  $\mathbf{J}_P$ , which is defined as a negative of the expected value (where the expectation is taken with respect to the vector of observations) of the Hessian matrix  $\mathbf{H}_P$ ,

$$\mathbf{J}_P = -\mathbf{E}(\mathbf{H}_P), \tag{3.13}$$

is derived.



Accounting for the relationships given in eqn. (3.5), the Hessian matrix,  $\mathbf{H}_P$ , can be decomposed into the sum of the Hessian matrix of the natural logarithm of the prior,  $\ln[p(\theta)]$ , and the Hessian matrix of the marginal log-likelihood function,  $l_R$ , as:

$$\mathbf{H}_P = \frac{\partial^2 \ln[p(\theta)]}{\partial \theta \partial \theta} + \frac{\partial^2 l_R}{\partial \theta \partial \theta} = \mathbf{H}_{P0} + \mathbf{H}_R, \quad (3.14)$$

where the Hessian matrix of the marginal log-likelihood function,  $\mathbf{H}_R$  (see eqn. (2.57)), is given by Koch [1990] and Searle et al. [1992] as:

$$\mathbf{H}_R = \frac{\partial^2 l_R}{\partial \theta \partial \theta} = \left\{ \frac{1}{m} \frac{1}{2} \text{tr}(\mathbf{R}\mathbf{T}_k \mathbf{R}\mathbf{T}_m) - \mathbf{I}^T \mathbf{R}\mathbf{T}_k \mathbf{R}\mathbf{T}_m \mathbf{R}\mathbf{I} \right\}_{k,m=1}^r. \quad (3.15)$$

The Hessian matrix of the natural logarithm of the prior probability density function,  $\mathbf{H}_{P0}$ , is derived as follows.

From the results of eqn. (3.10), with  $k \neq m$ , we find that the second partial derivative of the function  $\ln[p(\theta)]$  with respect to  $k$ -th and  $m$ -th variance components is equal to zero:

$$\frac{\partial^2 \ln[p(\theta)]}{\partial \theta_k \partial \theta_m} = \frac{\partial \left( \frac{b_k}{\theta_k^2} - \frac{p_k + 1}{\theta_k} \right)}{\partial \theta_m} = 0. \quad (3.16)$$

With  $k = m$ , the second partial derivative of the function  $\ln[p(\theta)]$  with respect to  $k$ -th and  $m$ -th variance components is found to be:

$$\frac{\partial^2 \ln[p(\theta)]}{\partial \theta_k \partial \theta_m} = \frac{\partial^2 \ln[p(\theta)]}{\partial \theta_k \partial \theta_k} = \frac{\partial \left( \frac{b_k}{\theta_k^2} - \frac{p_k + 1}{\theta_k} \right)}{\partial \theta_k} = \frac{p_k + 1}{\theta_k^2} - \frac{2b_k}{\theta_k^3} \quad (3.17)$$

As a result of eqn. (3.16) and eqn. (3.17), the Hessian matrix of the natural logarithm of the inverted gamma probability density function is given as:

$$\mathbf{H}_{P0} = \frac{\partial^2 \ln[p(\theta)]}{\partial \theta \partial \theta} = \left\{ d \left( \frac{p_k + 1}{\theta_k^2} - \frac{2b_k}{\theta_k^3} \right) \right\}_{k=1}^r \quad (3.18)$$

Finally, following from equations (3.15) and (3.18), we find the Hessian matrix of the natural logarithm of the posterior probability density function,  $l_p$ , with the inverted gamma distribution function chosen for a prior, to be:

$$\begin{aligned} \mathbf{H}_P &= \mathbf{H}_{P0} + \mathbf{H}_R \\ &= \left\{ d \left( \frac{p_k + 1}{\theta_k^2} - \frac{2b_k}{\theta_k^3} \right) \right\}_{k=1}^r + \left\{ \frac{1}{2} \text{tr}(\mathbf{R}\mathbf{T}_k\mathbf{R}\mathbf{T}_m) - \mathbf{1}^T \mathbf{R}\mathbf{T}_k\mathbf{R}\mathbf{T}_m \mathbf{R}\mathbf{1} \right\}_{k,m=1}^r \end{aligned} \quad (3.19)$$

The information matrix,  $\mathbf{J}_R$ , is subsequently derived from the Hessian matrix, eqn. (3.19), as:

$$\mathbf{J}_P = -\mathbf{E}(\mathbf{H}_P) = -\mathbf{E}(\mathbf{H}_{P0}) - \mathbf{E}(\mathbf{H}_R) = \mathbf{J}_{P0} + \mathbf{J}_R, \quad (3.20)$$

where  $\mathbf{J}_R$  is the marginal likelihood information matrix (see eqn. (2.61)), and is given by

Searle et al. [1992] as:

$$\mathbf{J}_R = -\mathbf{E} \left( \left\{ \frac{1}{2} \text{tr}(\mathbf{RT}_k \mathbf{RT}_m) - \mathbf{I}^T \mathbf{RT}_k \mathbf{RT}_m \mathbf{I} \right\}_{k,m=1}^r \right) = \left\{ \frac{1}{2} \text{tr}(\mathbf{RT}_k \mathbf{RT}_m) \right\}_{k,m=1}^r. \quad (3.21)$$

Since the expectations in eqn. (3.20) are taken with respect to the vector of observations,  $\mathbf{I}$ , the expected value of the Hessian matrix  $\mathbf{H}_{P0}$  is:

$$\mathbf{E}(\mathbf{H}_{P0}) = \mathbf{H}_{P0}. \quad (3.22)$$

Therefore, the information matrix of the posterior probability density function,  $p(\theta|\mathbf{I})$ , with the inverted gamma distribution function chosen for a prior, is finally found to be:

$$\mathbf{J}_P = \mathbf{J}_{P0} + \mathbf{J}_R = \left\{ d \left( \frac{2b_k}{\theta_k^3} - \frac{p_k + 1}{\theta_k^2} \right) \right\}_{k=1}^r + \left\{ \frac{1}{2} \text{tr}(\mathbf{RT}_k \mathbf{RT}_m) \right\}_{k,m=1}^r. \quad (3.23)$$

### 3.1.3 Application of the Fisher Scoring Method

In this section the Fisher scoring method is used to obtain the estimating equations for the vector of variance components. Application of the Fisher scoring method to the natural logarithm of the posterior probability density function,  $l_p$ , eqn. (3.5), yields the following expression:

$$\hat{\theta}_{(m+1)} = \hat{\theta}_{(m)} + \left[ \mathbf{J}_P(\hat{\theta}_{(m)}) \right]^{-1} \mathbf{s}_P(\hat{\theta}_{(m)}), \quad (3.24)$$

where:  $\hat{\theta}_{(m)}$  is the GML estimate of the vector of variance components from  $m$ -th

iterative step,

$\mathbf{J}_P$  is the information matrix, eqn. (3.23), and

$\mathbf{s}_P$  is the vector of scores, eqn. (3.11).

As shown in eqn. (3.11), the vector of scores can be decomposed into two parts:

$$\mathbf{s}_P = \mathbf{s}_{P0} + \mathbf{s}_R, \quad (3.25)$$

where  $\mathbf{s}_{P0}$  is the first derivative of the natural logarithm of the prior probability density function, eqn. (3.10):

$$\mathbf{s}_{P0} = \frac{\partial \ln[p(\theta)]}{\partial \theta} = \left\{ c \left( \frac{b_k}{\theta_k^2} - \frac{p_k + 1}{\theta_k} \right) \right\}_{k=1}^r, \quad (3.26)$$

$\mathbf{s}_R$  is the vector of marginal likelihood scores:

$$\mathbf{s}_R = - \left\{ c \frac{1}{2} \text{tr}(\mathbf{RT}_k) \right\}_{k=1}^r + \mathbf{q}_R, \text{ and} \quad (3.27)$$

$\mathbf{q}_R$  is the quadratic form of observations in the REML scoring equations (see eqn. (2.63)),

$$\mathbf{q}_R = \left\{ c \frac{1}{2} \mathbf{1}^T \mathbf{RT}_k \mathbf{R} \mathbf{1} \right\}_{k=1}^r. \quad (3.28)$$

Introducing the results of eqn. (3.23) and eqn. (3.25) into the Fisher scoring equations, eqn. (3.24), and making use of the following relationship (see, e.g., [Searle et al., 1992]):

$$\left\{ c \frac{1}{2} \text{tr}(\mathbf{RT}_k) \right\}_{k=1}^r = \left\{ c \frac{1}{2} \text{tr}(\mathbf{RT}_k \mathbf{RT}_m) \right\}_{k,m=1}^r \quad \theta = \mathbf{J}_R \theta, \quad (3.29)$$

we get

$$\hat{\theta}_{(m+1)} = \hat{\theta}_{(m)} + \left[ \mathbf{J}_{P0}(\hat{\theta}_{(m)}) + \mathbf{J}_R(\hat{\theta}_{(m)}) \right]^{-1} \left[ \mathbf{s}_{P0}(\hat{\theta}_{(m)}) + \mathbf{q}_R(\hat{\theta}_{(m)}) - \mathbf{J}_R(\hat{\theta}_{(m)}) \hat{\theta}_{(m)} \right]. \quad (3.30)$$

Subtracting and adding the following product:

$$\mathbf{J}_{P0}(\hat{\boldsymbol{\theta}}_{(m)})\hat{\boldsymbol{\theta}}_{(m)} = \left\{ \mathbf{d} \left( \frac{2\mathbf{b}_k}{\hat{\boldsymbol{\theta}}_{k(m)}^3} - \frac{\mathbf{p}_k + 1}{\hat{\boldsymbol{\theta}}_{k(m)}^2} \right) \right\}_{k=1}^r \hat{\boldsymbol{\theta}}_{(m)} = \left\{ \mathbf{c} \left( \frac{2\mathbf{b}_k}{\hat{\boldsymbol{\theta}}_{k(m)}^2} - \frac{\mathbf{p}_k + 1}{\hat{\boldsymbol{\theta}}_{k(m)}} \right) \right\}_{k=1}^r, \quad (3.31)$$

(where  $\hat{\boldsymbol{\theta}}_{k(m)}$  is the is the GML estimate of the k-th variance component from m-th iterative step) from the vector of scores,  $\mathbf{s}_P$  (eqn. 3.25), we get:

$$\begin{aligned} \hat{\boldsymbol{\theta}}_{(m+1)} &= \hat{\boldsymbol{\theta}}_{(m)} + \left[ \mathbf{J}_{P0}(\hat{\boldsymbol{\theta}}_{(m)}) + \mathbf{J}_R(\hat{\boldsymbol{\theta}}_{(m)}) \right]^{-1} \\ &\times \left[ \mathbf{s}_{P0}(\hat{\boldsymbol{\theta}}_{(m)}) + \mathbf{q}_R(\hat{\boldsymbol{\theta}}_{(m)}) - \mathbf{J}_R(\hat{\boldsymbol{\theta}}_{(m)})\hat{\boldsymbol{\theta}}_{(m)} - \mathbf{J}_{P0}(\hat{\boldsymbol{\theta}}_{(m)})\hat{\boldsymbol{\theta}}_{(m)} + \mathbf{J}_{P0}(\hat{\boldsymbol{\theta}}_{(m)})\hat{\boldsymbol{\theta}}_{(m)} \right] \\ &= \left[ \mathbf{J}_{P0}(\hat{\boldsymbol{\theta}}_{(m)}) + \mathbf{J}_R(\hat{\boldsymbol{\theta}}_{(m)}) \right]^{-1} \left[ \mathbf{s}_{P0}(\hat{\boldsymbol{\theta}}_{(m)}) + \mathbf{q}_R(\hat{\boldsymbol{\theta}}_{(m)}) + \mathbf{J}_{P0}(\hat{\boldsymbol{\theta}}_{(m)})\hat{\boldsymbol{\theta}}_{(m)} \right] \\ &= \left[ \mathbf{J}_P(\hat{\boldsymbol{\theta}}_{(m)}) \right]^{-1} \mathbf{q}_P(\hat{\boldsymbol{\theta}}_{(m)}), \end{aligned} \quad (3.32)$$

with

$$\begin{aligned} \mathbf{q}_P &= \mathbf{s}_{P0} + \mathbf{q}_R + \mathbf{J}_{P0}\boldsymbol{\theta} = \\ &\left\{ \mathbf{c} \left( \frac{\mathbf{b}_k}{\boldsymbol{\theta}_k^2} - \frac{\mathbf{p}_k + 1}{\boldsymbol{\theta}_k} \right) \right\}_{k=1}^r + \left\{ \mathbf{c} \frac{1}{2} \mathbf{I}^T \mathbf{R} \mathbf{T}_k \mathbf{R} \mathbf{I} \right\}_{k=1}^r + \left\{ \mathbf{c} \left( \frac{2\mathbf{b}_k}{\boldsymbol{\theta}_k^2} - \frac{\mathbf{p}_k + 1}{\boldsymbol{\theta}_k} \right) \right\}_{k=1}^r \\ &= \left\{ \mathbf{c} \left( \frac{3\mathbf{b}_k}{\boldsymbol{\theta}_k^2} - \frac{2\mathbf{p}_k + 2}{\boldsymbol{\theta}_k} \right) \right\}_{k=1}^r + \left\{ \mathbf{c} \frac{1}{2} \mathbf{I}^T \mathbf{R} \mathbf{T}_k \mathbf{R} \mathbf{I} \right\}_{k=1}^r = \mathbf{q}_{P0} + \mathbf{q}_R. \end{aligned} \quad (3.33)$$

Taking into account the results of eqn. (3.32) and eqn. (3.33) we get the final form of the Fisher scoring equations as:

$$\begin{aligned}
\hat{\theta}_{(m+1)} &= [\mathbf{J}_P(\hat{\theta}_{(m)})]^{-1} \mathbf{q}_P(\hat{\theta}_{(m)}) = [\mathbf{J}_{P0}(\hat{\theta}_{(m)}) + \mathbf{J}_R(\hat{\theta}_{(m)})]^{-1} [\mathbf{q}_{P0}(\hat{\theta}_{(m)}) + \mathbf{q}_R(\hat{\theta}_{(m)})] \\
&= \left[ \left\{ d \left( \frac{2b_k}{\hat{\theta}_{k(m)}^3} - \frac{p_k + 1}{\hat{\theta}_{k(m)}^2} \right) \right\}_{k=1}^r + \left\{ m \frac{1}{2} \text{tr}(\mathbf{R}_{\hat{\theta}_{(m)}} \mathbf{T}_k \mathbf{R}_{\hat{\theta}_{(m)}} \mathbf{T}_m) \right\}_{k,m=1}^r \right]^{-1} \\
&\times \left[ \left\{ c \left( \frac{3b_k}{\hat{\theta}_{k(m)}^2} - \frac{2p_k + 2}{\hat{\theta}_{k(m)}} \right) \right\}_{k=1}^r + \left\{ c \frac{1}{2} \mathbf{I}^T \mathbf{R}_{\hat{\theta}_{(m)}} \mathbf{T}_k \mathbf{R}_{\hat{\theta}_{(m)}} \mathbf{I} \right\}_{k=1}^r \right].
\end{aligned} \tag{3.34}$$

In the limit, eqn. (3.34) yields the GML estimator (with the inverted gamma prior),  $\hat{\theta}$ , of the vector of variance components. The asymptotic covariance matrix of this GML estimator is given by the inverse of the information matrix (see Appendix B) as:

$$\text{var}(\hat{\theta}) \approx [\mathbf{J}_P(\hat{\theta})]^{-1}. \tag{3.35}$$

### 3.2 GML Estimation with Noninformative Prior

The prior distribution function for the vector of variance components is represented in this chapter by the Jeffrey's noninformative prior, the probability density function of which is given by eqn. (2.71). For the sake of completeness, the expressions for the probability density function of the noninformative prior and the posterior probability density function, which are given in Subsection 2.3.1, are repeated in this section.

Under the assumption that the likelihood function of the vector of observations is identical to the marginal likelihood function, eqn. (2.44), the noninformative prior probability density function for the vector of variance components has the following form:

$$p(\theta) \propto |\mathbf{J}_R|^{-\frac{1}{2}}, \quad (3.36)$$

where  $\mathbf{J}_R$  is the information matrix (see eqn. (2.61)) associated with the marginal likelihood function.

The posterior probability density function results from application of the Bayes' Theorem, eqn. (2.65), which with the marginal likelihood function, eqn. (2.44), and the noninformative prior probability density function, eqn. (3.36), gives:

$$p(\theta|\mathbf{l}) \propto p(\theta)p(\mathbf{l}|\theta) \propto p(\theta)L_R \propto |\mathbf{J}_R|^{-\frac{1}{2}} \left( |\mathbf{C}_\theta| |\mathbf{A}^T \mathbf{C}_\theta^{-1} \mathbf{A}| \right)^{-\frac{1}{2}} \exp\left(-\frac{\mathbf{l}^T \mathbf{R} \mathbf{l}}{2}\right). \quad (3.37)$$

By taking natural logarithms of both sides of eqn. (3.37) we get:

$$\begin{aligned} l_N &= \ln[p(\theta|\mathbf{l})] = c + \ln[p(\theta)] + \ln(L_R) \\ &= c + \frac{1}{2} \ln|\mathbf{J}_R| - \frac{1}{2} \ln|\mathbf{C}_\theta| - \frac{1}{2} \ln|\mathbf{A}^T \mathbf{C}_\theta^{-1} \mathbf{A}| - \frac{\mathbf{l}^T \mathbf{R} \mathbf{l}}{2}, \end{aligned} \quad (3.38)$$

where  $c$  is some constant.

### 3.2.1 Derivation of the Vector of Scores

The vector of scores, being a vector of  $r$  partial first derivatives of  $l_N$  with respect to the variance components  $\theta_k$ :

$$\mathbf{s}_N = \left\{ \frac{\partial l_N}{\partial \theta_k} \right\}_{k=1}^r = \frac{\partial l_N}{\partial \theta}, \quad (3.39)$$

is derived in this section.

As seen in eqn. (3.37), the vector of scores can be decomposed into the sum of the vector of first derivatives of the natural logarithm of the noninformative prior,  $\ln[p(\theta)]$ , and the marginal log-likelihood function  $l_R$ , eqn. (2.45):

$$\mathbf{s}_N = \frac{\partial l_N}{\partial \theta} = \frac{\partial \ln[p(\theta)]}{\partial \theta} + \frac{\partial l_R}{\partial \theta} = \mathbf{s}_{N0} + \mathbf{s}_R, \quad (3.40)$$

where the first derivative of the marginal log-likelihood function is given by Koch [1990] as:

$$\mathbf{s}_R = \frac{\partial l_R}{\partial \theta} = \left\{ c \left( -\frac{1}{2} \text{tr}(\mathbf{R}\mathbf{T}_k) + \frac{1}{2} \mathbf{1}^T \mathbf{R}\mathbf{T}_k \mathbf{R}\mathbf{1} \right) \right\}_{k=1}^r. \quad (3.41)$$

The first derivative of the natural logarithm of the noninformative prior

$$\mathbf{s}_{N0} = \frac{\partial \ln[p(\theta)]}{\partial \theta} = \frac{1}{2} \frac{\partial \ln|\mathbf{J}_R|}{\partial \theta}, \quad (3.42)$$

where:

$$\mathbf{J}_R = \left\{ \frac{1}{2} \text{tr}(\mathbf{R}\mathbf{T}_i \mathbf{R}\mathbf{T}_j) \right\}_{i,j=1}^r,$$

(3.43)

$$\mathbf{R} = \mathbf{C}_\theta^{-1} \left[ \mathbf{I} - \mathbf{A} (\mathbf{A}^T \mathbf{C}_\theta^{-1} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{C}_\theta^{-1} \right], \text{ and} \quad (3.44)$$

$$\mathbf{C}_\theta = \sum_{i=1}^r \mathbf{T}_i \theta_i, \quad (3.45)$$

is derived as follows.



Searle et al. [1992] give the following rule for taking a derivative of the natural logarithm of the determinant of a regular symmetric matrix  $\mathbf{A}$ , being a function of  $a_i$ , with respect to  $a_i$ :

$$\frac{\partial \ln|\mathbf{A}|}{\partial a_i} = \text{tr}\left(\mathbf{A}^{-1} \frac{\partial \mathbf{A}}{\partial a_i}\right). \quad (3.46)$$

Using this rule we get the first partial derivative of the natural logarithm of the determinant of  $\mathbf{J}_R$  as:

$$\frac{\partial \ln|\mathbf{J}_R|}{\partial \theta_k} = \text{tr}\left(\mathbf{J}_R^{-1} \frac{\partial \mathbf{J}_R}{\partial \theta_k}\right). \quad (3.47)$$

The next step is to find the first partial derivative of the information matrix  $\mathbf{J}_R$

$$\frac{\partial \mathbf{J}_R}{\partial \theta_k} = \left\{ \begin{matrix} \partial j_{ij} \\ m \partial \theta_k \end{matrix} \right\}_{i,j=1}^r, \quad (3.48)$$

where

$$\frac{\partial j_{ij}}{\partial \theta_k} = \frac{d\left[\frac{1}{2} \text{tr}(\mathbf{R}\mathbf{T}_i\mathbf{R}\mathbf{T}_j)\right]}{\partial \theta_k}. \quad (3.49)$$

Using the chain rule for the matrix functions [Magnus and Neudecker, 1988] we get the derivative of (i,j)-th component of the information matrix,  $\mathbf{J}_R$ , with respect to the k-th variance component,  $\theta_k$ , as:

$$\frac{\partial j_{ij}}{\partial \theta_k} = \left[\mathbf{D}j_{ij}(\theta_k)\right]_{1 \times 1} = \frac{\partial j_{ij}}{\partial \text{vec}\mathbf{R}} \frac{\partial \text{vec}\mathbf{R}}{\partial \theta_k} = \left[\mathbf{D}j_{ij}(\mathbf{R})\right]_{1 \times m^2} \left[\mathbf{D}\mathbf{R}(\theta_k)\right]_{m^2 \times 1}. \quad (3.50)$$

Derivation of an expression for the first derivative, eqn. (3.50), of (i,j)-th component of the information matrix,  $\mathbf{J}_{\mathbf{R}}$ , with respect to the k-th variance component,  $\theta_k$ , will be divided into two steps. In the first step the derivative matrix  $\frac{\partial j_{ij}}{\partial \text{vec}\mathbf{R}} = [\mathbf{D}j_{ij}(\mathbf{R})]_{1 \times m^2}$  will be derived. In the second step an expression for the derivative matrix  $\frac{\partial \text{vec}\mathbf{R}}{\partial \theta_k} = [\mathbf{DR}(\theta_k)]_{m^2 \times 1}$  will be found.

**(1) The derivative matrix**  $\frac{\partial j_{ij}}{\partial \text{vec}\mathbf{R}} = [\mathbf{D}j_{ij}(\mathbf{R})]_{1 \times m^2}$

The differential and the derivative matrix of the scalar function  $j_{ij}$  of the matrix  $\mathbf{R}$  can be computed by means of the “identification theorem for matrix functions” [Magnus and Neudecker, 1988]. In the case of a scalar function of a matrix, this theorem implies that, in order to find the derivative matrix  $\mathbf{D}j_{ij}(\mathbf{R})$ , the differential  $dj_{ij}$  has to be converted into the following form:

$$dj_{ij} = \text{tr}(\mathbf{X}^T d\mathbf{R}) = (\text{vec}\mathbf{X})^T d(\text{vec}\mathbf{R}) = [\mathbf{D}j_{ij}(\mathbf{R})]_{1 \times m^2} d(\text{vec}\mathbf{R})_{m^2 \times 1}, \quad (3.51)$$

where  $[\mathbf{D}j_{ij}(\mathbf{R})]_{1 \times m^2}$  is the Jacobian (derivative) matrix, and  $\mathbf{X}$  is some matrix such that

$$[\mathbf{D}j_{ij}(\mathbf{R})]_{1 \times m^2} = (\text{vec}\mathbf{X})^T = \frac{\partial j_{ij}}{\partial \text{vec}\mathbf{R}}_{m^2 \times 1}. \quad (3.52)$$

Starting from:

$$\mathbf{j}_{ij} = \frac{1}{2} \text{tr}(\mathbf{RT}_i \mathbf{RT}_j), \quad (3.53)$$

and applying the identity (see, e.g., [Searle, 1982]):

$$\text{tr}(\mathbf{A}^T \mathbf{B}) = (\text{vec} \mathbf{A})^T \text{vec} \mathbf{B}, \quad (3.54)$$

we get the following expression for the differential  $d\mathbf{j}_{ij}$ :

$$\begin{aligned} d\mathbf{j}_{ij} &= d\left[\frac{1}{2} \text{tr}(\mathbf{RT}_i \mathbf{RT}_j)\right] = \frac{1}{2} \text{tr}\left[d(\mathbf{RT}_i \mathbf{RT}_j)\right] = \frac{1}{2} \text{tr}\left[d(\mathbf{RT}_i) \mathbf{RT}_j + \mathbf{RT}_i d(\mathbf{RT}_j)\right] \\ &= \frac{1}{2} \text{tr}(d\mathbf{RT}_i \mathbf{RT}_j + \mathbf{RT}_i d\mathbf{RT}_j) = \frac{1}{2} \text{tr}(\mathbf{T}_i \mathbf{RT}_j d\mathbf{R}) + \frac{1}{2} \text{tr}(\mathbf{T}_j \mathbf{RT}_i d\mathbf{R}) \\ &= \frac{1}{2} \text{tr}\left[(\mathbf{T}_i \mathbf{RT}_j + \mathbf{T}_j \mathbf{RT}_i) d\mathbf{R}\right] = \frac{1}{2} \left[\text{vec}(\mathbf{T}_i \mathbf{RT}_j + \mathbf{T}_j \mathbf{RT}_i)\right]^T d(\text{vec} \mathbf{R}). \end{aligned} \quad (3.55)$$

Applying “the identification theorem for matrix functions”, eqn. (3.51), to the expression for the differential  $d\mathbf{j}_{ij}$ , eqn. (3.55), we finally find the derivative matrix  $\mathbf{D}\mathbf{j}_{ij}$  to be:

$$\left[\mathbf{D}\mathbf{j}_{ij}(\mathbf{R})\right]_{1 \times m^2} = \frac{\partial \mathbf{j}_{ij}}{\partial \text{vec} \mathbf{R}} = \frac{1}{2} \left[\text{vec}(\mathbf{T}_i \mathbf{RT}_j + \mathbf{T}_j \mathbf{RT}_i)\right]^T_{m^2 \times 1}. \quad (3.56)$$

**(2) The derivative matrix**  $\frac{\partial \text{vec} \mathbf{R}}{\partial \theta_k} = \left[\mathbf{DR}(\theta_k)\right]_{m^2 \times 1}$

The matrix function  $\mathbf{R}$  of the scalar  $\theta_k$ ,  $\mathbf{R}(\theta_k) = \mathbf{R}\{\mathbf{P}[\mathbf{C}_\theta(\theta_k)]\}$ , will be treated as a superposition of the following three functions:

$$\mathbf{R}(\mathbf{P}) = \mathbf{P} \left[ \mathbf{I} - \mathbf{A} (\mathbf{A}^T \mathbf{P} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{P} \right], \quad (3.57)$$

$$\mathbf{P}(\mathbf{C}_\theta) = \mathbf{C}_\theta^{-1}, \text{ and} \quad (3.58)$$

$$\mathbf{C}_\theta(\theta_k) = \sum_{i=1}^I \mathbf{T}_i \theta_i. \quad (3.59)$$

The derivative matrix,  $\mathbf{DR}(\theta_k)$ , will be computed by means of the chain rule, as the product of the following three derivative matrices:

$$\begin{aligned} [\mathbf{DR}(\theta_k)]_{m^2 \times 1} &= \frac{\partial \text{vec} \mathbf{R}}{\partial \theta_k} = \frac{\partial \text{vec} \mathbf{R}}{\partial \text{vec} \mathbf{P}} \frac{\partial \text{vec} \mathbf{P}}{\partial \text{vec} \mathbf{C}_\theta} \frac{\partial \text{vec} \mathbf{C}_\theta}{\partial \theta_k} \\ &= [\mathbf{DR}(\mathbf{P})]_{m^2 \times m^2} [\mathbf{DP}(\mathbf{C}_\theta)]_{m^2 \times m^2} [\mathbf{DC}_\theta(\theta_k)]_{m^2 \times 1}. \end{aligned} \quad (3.60)$$

The three derivative matrices,  $\mathbf{DR}$ ,  $\mathbf{DP}$ , and  $\mathbf{DC}_\theta$ , are derived as follows.

(2a) The derivative matrix of  $\mathbf{R}$  with respect to  $\mathbf{P}$

For a matrix function of a matrix, the “identification theorem for matrix functions” implies that, in order to find the derivative matrix  $\mathbf{DR}(\mathbf{P})$ , we have to arrive at the following form of the differential  $d(\text{vec} \mathbf{R})$ :

$$d(\text{vec} \mathbf{R}) = \mathbf{X} d(\text{vec} \mathbf{P}) = [\mathbf{DR}(\mathbf{P})]_{m^2 \times m^2} d(\text{vec} \mathbf{P})_{m^2 \times 1}, \quad (3.61)$$

where  $\mathbf{DR}(\mathbf{P})$  is the Jacobian (derivative) matrix, and  $\mathbf{X}$  is some matrix such that

$$[\mathbf{DR}(\mathbf{P})]_{m^2 \times m^2} = \mathbf{X} = \frac{\partial \text{vec} \mathbf{R}_{m^2 \times 1}}{\partial \text{vec} \mathbf{P}_{m^2 \times 1}}. \quad (3.62)$$

Starting from eqn. (3.57):

$$\mathbf{R}(\mathbf{P}) = \mathbf{P} \left[ \mathbf{I} - \mathbf{A}(\mathbf{A}^T \mathbf{P} \mathbf{A})^{-1} \mathbf{A}^T \right] = \mathbf{P} - \mathbf{P} \mathbf{A}(\mathbf{A}^T \mathbf{P} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{P}, \quad (3.63)$$

we get the following expression for the differential  $d\mathbf{R}$ :

$$\begin{aligned} d\mathbf{R}(\mathbf{P}) &= d\mathbf{P} - d \left[ \mathbf{P} \mathbf{A}(\mathbf{A}^T \mathbf{P} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{P} \right] = d\mathbf{P} - d(\mathbf{P} \mathbf{A})(\mathbf{A}^T \mathbf{P} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{P} - \mathbf{P} \mathbf{A} d \left[ (\mathbf{A}^T \mathbf{P} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{P} \right] \\ &= d\mathbf{P} - d\mathbf{P} \mathbf{A}(\mathbf{A}^T \mathbf{P} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{P} - \mathbf{P} \mathbf{A} \left\{ d \left[ (\mathbf{A}^T \mathbf{P} \mathbf{A})^{-1} \right] \mathbf{A}^T \mathbf{P} + (\mathbf{A}^T \mathbf{P} \mathbf{A})^{-1} d(\mathbf{A}^T \mathbf{P}) \right\} \\ &= d\mathbf{P} - d\mathbf{P} \mathbf{A}(\mathbf{A}^T \mathbf{P} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{P} - \mathbf{P} \mathbf{A} \left[ -(\mathbf{A}^T \mathbf{P} \mathbf{A})^{-1} \mathbf{A}^T d\mathbf{P} \mathbf{A}(\mathbf{A}^T \mathbf{P} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{P} + (\mathbf{A}^T \mathbf{P} \mathbf{A})^{-1} \mathbf{A}^T d\mathbf{P} \right] \\ &= d\mathbf{P} - d\mathbf{P} \mathbf{A}(\mathbf{A}^T \mathbf{P} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{P} + \mathbf{P} \mathbf{A}(\mathbf{A}^T \mathbf{P} \mathbf{A})^{-1} \mathbf{A}^T d\mathbf{P} \mathbf{A}(\mathbf{A}^T \mathbf{P} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{P} - \mathbf{P} \mathbf{A}(\mathbf{A}^T \mathbf{P} \mathbf{A})^{-1} \mathbf{A}^T d\mathbf{P} \\ &= \mathbf{I} d\mathbf{P} \mathbf{I} - \mathbf{I} d\mathbf{P} \mathbf{B} + \mathbf{B}^T d\mathbf{P} \mathbf{B} - \mathbf{B}^T d\mathbf{P} \mathbf{I}, \end{aligned} \quad (3.64)$$

where:

$$\mathbf{B} = \mathbf{A}(\mathbf{A}^T \mathbf{P} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{P}. \quad (3.65)$$

Applying the following matrix identity (see, e.g., [Searle, 1982]):

$$\text{vec}(\mathbf{ABC}) = (\mathbf{C}^T \otimes \mathbf{A}) \text{vec} \mathbf{B} \quad (3.66)$$

to eqn. (3.64) we get:

$$\begin{aligned} d[\text{vec} \mathbf{R}(\mathbf{P})]_{m^2 \times 1} &= (\mathbf{I}_{m \times m} \otimes \mathbf{I}_{m \times m}) d(\text{vec} \mathbf{P})_{m^2 \times 1} - (\mathbf{B}_{m \times m}^T \otimes \mathbf{I}_{m \times m}) d(\text{vec} \mathbf{P})_{m^2 \times 1} \\ &+ (\mathbf{B}_{m \times m}^T \otimes \mathbf{B}_{m \times m}^T) d(\text{vec} \mathbf{P})_{m^2 \times 1} - (\mathbf{I}_{m \times m} \otimes \mathbf{B}_{m \times m}^T) d(\text{vec} \mathbf{P})_{m^2 \times 1} \\ &= \left[ (\mathbf{I}_{m \times m} \otimes \mathbf{I}_{m \times m}) - (\mathbf{B}_{m \times m}^T \otimes \mathbf{I}_{m \times m}) + (\mathbf{B}_{m \times m}^T \otimes \mathbf{B}_{m \times m}^T) - (\mathbf{I}_{m \times m} \otimes \mathbf{B}_{m \times m}^T) \right] d(\text{vec} \mathbf{P})_{m^2 \times 1}. \end{aligned} \quad (3.67)$$

Applying the “identification theorem for matrix functions”, eqn. (3.61), to the expression for the differential  $d(\text{vec}\mathbf{R})$ , eqn. (3.67), we finally find the derivative matrix  $\mathbf{DR}$  to be:

$$\begin{aligned} [\mathbf{DR}(\mathbf{P})]_{m^2 \times m^2} &= \frac{\partial \text{vec}\mathbf{R}}{\partial \text{vec}\mathbf{P}} = \\ & \left[ \left( \mathbf{I}_{m \times m} \otimes \mathbf{I}_{m \times m} \right) - \left( \mathbf{B}_{m \times m}^T \otimes \mathbf{I}_{m \times m} \right) + \left( \mathbf{B}_{m \times m}^T \otimes \mathbf{B}_{m \times m}^T \right) - \left( \mathbf{I}_{m \times m} \otimes \mathbf{B}_{m \times m}^T \right) \right]_{m^2 \times m^2}. \end{aligned} \quad (3.68)$$

(2b) The derivative matrix of  $\mathbf{P}$  with respect to  $\mathbf{C}_\theta$

It follows from the “identification theorem for matrix functions”, eqn. (3.61), that in order to find the derivative matrix  $\mathbf{DP}$ , the differential  $d(\text{vec}\mathbf{P})$  has to be converted into the following form:

$$d(\text{vec}\mathbf{P}) = \mathbf{X} d(\text{vec}\mathbf{C}_\theta) = [\mathbf{DP}(\mathbf{C}_\theta)]_{m^2 \times m^2} d(\text{vec}\mathbf{C}_\theta), \quad (3.69)$$

where  $\mathbf{DP}(\mathbf{C}_\theta)$  is the Jacobian (derivative) matrix and  $\mathbf{X}$  is some matrix such that

$$[\mathbf{DP}(\mathbf{C}_\theta)]_{m^2 \times m^2} = \mathbf{X} = \frac{\partial \text{vec}\mathbf{P}_{m^2 \times 1}}{\partial \text{vec}\mathbf{C}_\theta_{m^2 \times 1}}. \quad (3.70)$$

Starting with eqn. (3.58):

$$\mathbf{P}(\mathbf{C}_\theta) = \mathbf{C}_\theta^{-1},$$

we get the following expression for the differential  $d\mathbf{P}$ :

$$d\mathbf{P}(\mathbf{C}_\theta) = -\mathbf{C}_\theta^{-1} d\mathbf{C}_\theta \mathbf{C}_\theta^{-1}. \quad (3.71)$$

Application of the matrix identity given in eqn. (3.66) to eqn. (3.71) yields:

$$d[\text{vec}\mathbf{P}(\mathbf{C}_\theta)]_{m^2 \times 1} = -(\mathbf{C}_\theta^{-1} \otimes \mathbf{C}_\theta^{-1})_{m^2 \times m^2} d(\text{vec}\mathbf{C}_\theta)_{m^2 \times 1} . \quad (3.72)$$

Applying the results of eqn. (3.69) to the expression for the differential  $d(\text{vec}\mathbf{P})$ , eqn.

(3.72), we finally find the derivative matrix  $\mathbf{DP}$  to be:

$$[\mathbf{DP}(\mathbf{C}_\theta)]_{m^2 \times m^2} = \frac{\partial \text{vec}\mathbf{P}}{\partial \text{vec}\mathbf{C}_\theta} = -(\mathbf{C}_\theta^{-1} \otimes \mathbf{C}_\theta^{-1})_{m^2 \times m^2} . \quad (3.73)$$

(2c) The derivative matrix of  $\mathbf{C}_\theta$  with respect to  $\theta_k$

In the case of a matrix function of a scalar, the “identification theorem for matrix functions” implies that, in order to find the derivative matrix  $\mathbf{DC}_\theta(\theta_k)$ , the differential  $d(\text{vec}\mathbf{C}_\theta)$  has to be converted into the following form:

$$d(\text{vec}\mathbf{C}_\theta) = \text{vec}\mathbf{X} d\theta_k = [\mathbf{DC}_\theta(\theta_k)]_{m^2 \times 1} d\theta_k , \quad (3.74)$$

where  $\mathbf{DC}_\theta(\theta_k)$  is the Jacobian (derivative) matrix, and  $\mathbf{X}$  is some matrix such that:

$$[\mathbf{DC}_\theta(\theta_k)]_{m^2 \times 1} = \text{vec}\mathbf{X} = \frac{\partial \text{vec}\mathbf{C}_\theta_{m^2 \times 1}}{\partial \theta_{k1 \times 1}} . \quad (3.75)$$

Differentiating both sides of eqn. (3.59), we get the following expression for the differential  $d\mathbf{C}_\theta$ :

$$d\mathbf{C}_\theta(\theta_k) = \mathbf{T}_k d\theta_k . \quad (3.76)$$

This becomes on vectorizing both sides of eqn. (3.76)

$$d[\text{vec}\mathbf{C}_\theta(\theta_k)]_{m^2 \times 1} = \text{vec}\mathbf{T}_k d\theta_k. \quad (3.77)$$

Applying the “identification theorem for matrix functions”, eqn. (3.74), to the expression for the differential  $d(\text{vec}\mathbf{C}_\theta)$ , eqn. (3.77), we subsequently get:

$$[\mathbf{DC}_\theta(\theta_k)]_{m^2 \times 1} = \frac{\partial \text{vec}\mathbf{C}_\theta}{\partial \theta_k} = \text{vec}\mathbf{T}_k. \quad (3.78)$$

Finally, accounting for the results derived in points (2a), (2b), and (2c), the derivative matrix of  $\mathbf{R}$  with respect to the  $k$ -th variance component,  $\theta_k$ , is:

$$\begin{aligned} [\mathbf{DR}(\theta_k)]_{m^2 \times 1} &= [\mathbf{DR}(\mathbf{P})]_{m^2 \times m^2} [\mathbf{DP}(\mathbf{C}_\theta)]_{m^2 \times m^2} [\mathbf{DC}_\theta(\theta_k)]_{m^2 \times 1} \\ &= \left[ (\mathbf{I} \otimes \mathbf{I}) - (\mathbf{B}^T \otimes \mathbf{I}) + (\mathbf{B}^T \otimes \mathbf{B}^T) - (\mathbf{I} \otimes \mathbf{B}^T) \right]_{m^2 \times m^2} \\ &\quad \times \left[ -(\mathbf{C}_\theta^{-1} \otimes \mathbf{C}_\theta^{-1}) \right]_{m^2 \times m^2} (\text{vec}\mathbf{T}_k)_{m^2 \times 1}. \end{aligned} \quad (3.79)$$

To simplify the form of eqn. (3.79), the following two matrix identities, which are given by Searle [1982] as:

$$(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = \mathbf{AC} \otimes \mathbf{BD}, \quad (3.80)$$

and

$$\text{vec}(\mathbf{ABC}) = (\mathbf{C}^T \otimes \mathbf{A})\text{vec}\mathbf{B}, \quad (3.81)$$

will be used. The development is then as follows:



$$\begin{aligned}
[\mathbf{DR}(\theta_k)]_{m^2 \times 1} &= [(\mathbf{I} \otimes \mathbf{I}) - (\mathbf{B}^T \otimes \mathbf{I}) + (\mathbf{B}^T \otimes \mathbf{B}^T) - (\mathbf{I} \otimes \mathbf{B}^T)]_{m^2 \times m^2} \times [-(\mathbf{C}_\theta^{-1} \otimes \mathbf{C}_\theta^{-1})]_{m^2 \times m^2} \\
&\times (\text{vec} \mathbf{T}_k)_{m^2 \times 1} \\
&= -[(\mathbf{C}_\theta^{-1} \otimes \mathbf{C}_\theta^{-1}) - (\mathbf{B}^T \mathbf{C}_\theta^{-1} \otimes \mathbf{C}_\theta^{-1}) + (\mathbf{B}^T \mathbf{C}_\theta^{-1} \otimes \mathbf{B}^T \mathbf{C}_\theta^{-1}) - (\mathbf{C}_\theta^{-1} \otimes \mathbf{B}^T \mathbf{C}_\theta^{-1})]_{m^2 \times m^2} \\
&\times (\text{vec} \mathbf{T}_k)_{m^2 \times 1} \\
&= -(\mathbf{C}_\theta^{-1} \otimes \mathbf{C}_\theta^{-1}) \text{vec} \mathbf{T}_k + (\mathbf{B}^T \mathbf{C}_\theta^{-1} \otimes \mathbf{C}_\theta^{-1}) \text{vec} \mathbf{T}_k - (\mathbf{B}^T \mathbf{C}_\theta^{-1} \otimes \mathbf{B}^T \mathbf{C}_\theta^{-1}) \text{vec} \mathbf{T}_k \\
&+ (\mathbf{C}_\theta^{-1} \otimes \mathbf{B}^T \mathbf{C}_\theta^{-1}) \text{vec} \mathbf{T}_k \\
&= -\text{vec}(\mathbf{C}_\theta^{-1} \mathbf{T}_k \mathbf{C}_\theta^{-1}) + \text{vec}(\mathbf{C}_\theta^{-1} \mathbf{T}_k \mathbf{C}_\theta^{-1} \mathbf{B}) - \text{vec}(\mathbf{B}^T \mathbf{C}_\theta^{-1} \mathbf{T}_k \mathbf{C}_\theta^{-1} \mathbf{B}) + \text{vec}(\mathbf{B}^T \mathbf{C}_\theta^{-1} \mathbf{T}_k \mathbf{C}_\theta^{-1}).
\end{aligned} \tag{3.82}$$

Finally, with  $\mathbf{B}$  given by eqn. (3.65) and  $\mathbf{P}$  by eqn (3.58) we find the expression for the derivative matrix of  $\mathbf{R}$  with respect to the  $k$ -th variance component,  $\theta_k$ , to be:

$$\begin{aligned}
[\mathbf{DR}(\theta_k)]_{m^2 \times 1} &= -\text{vec}(\mathbf{P} \mathbf{T}_k \mathbf{P}) + \text{vec} \left[ \mathbf{P} \mathbf{T}_k \mathbf{P} \mathbf{A} (\mathbf{A}^T \mathbf{P} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{P} \right] \\
&- \text{vec} \left[ \mathbf{P} \mathbf{A} (\mathbf{A}^T \mathbf{P} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{P} \mathbf{T}_k \mathbf{P} \mathbf{A} (\mathbf{A}^T \mathbf{P} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{P} \right] + \text{vec} \left[ \mathbf{P} \mathbf{A} (\mathbf{A}^T \mathbf{P} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{P} \mathbf{T}_k \mathbf{P} \right] \\
&= -\text{vec} \left\{ \mathbf{P} \mathbf{T}_k \mathbf{P} \left[ \mathbf{I} - \mathbf{A} (\mathbf{A}^T \mathbf{P} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{P} \right] \right\} + \text{vec} \left\{ \mathbf{P} \mathbf{A} (\mathbf{A}^T \mathbf{P} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{P} \mathbf{T}_k \mathbf{P} \left[ \mathbf{I} - \mathbf{A} (\mathbf{A}^T \mathbf{P} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{P} \right] \right\} \\
&= \text{vec} \left\{ \left[ \mathbf{P} \mathbf{A} (\mathbf{A}^T \mathbf{P} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{P} \mathbf{T}_k \mathbf{P} - \mathbf{P} \mathbf{T}_k \mathbf{P} \right] \left[ \mathbf{I} - \mathbf{A} (\mathbf{A}^T \mathbf{P} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{P} \right] \right\} \\
&= \text{vec} \left\{ \left[ \mathbf{P} \mathbf{A} (\mathbf{A}^T \mathbf{P} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{P} - \mathbf{P} \right] \mathbf{T}_k \left[ \mathbf{P} - \mathbf{P} \mathbf{A} (\mathbf{A}^T \mathbf{P} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{P} \right] \right\} \\
&= -\text{vec}(\mathbf{R} \mathbf{T}_k \mathbf{R}).
\end{aligned} \tag{3.83}$$

Coming back to the original objective, i.e., derivation of the first partial derivative of the  $(i,j)$ -th element of the information matrix,  $\mathbf{J}_R$ , with respect to the  $k$ -th variance component,  $\theta_k$ , by making use of the results given in eqn. (3.56) and eqn. (3.83), we get:

$$\begin{aligned}
\frac{\partial j_{ij}}{\partial \theta_k} &= [\mathbf{D}j_{ij}(\theta_k)]_{1 \times 1} = [\mathbf{D}j_{ij}(\mathbf{R})]_{1 \times m^2} [\mathbf{D}\mathbf{R}(\theta_k)]_{m^2 \times 1} \\
&= \frac{1}{2} [\text{vec}(\mathbf{T}_i \mathbf{R} \mathbf{T}_j + \mathbf{T}_j \mathbf{R} \mathbf{T}_i)]^T [-\text{vec}(\mathbf{R} \mathbf{T}_k \mathbf{R})] \\
&= -\frac{1}{2} [\text{vec}(\mathbf{T}_j \mathbf{R} \mathbf{T}_i)]^T \text{vec}(\mathbf{R} \mathbf{T}_k \mathbf{R}) - \frac{1}{2} [\text{vec}(\mathbf{T}_i \mathbf{R} \mathbf{T}_j)]^T \text{vec}(\mathbf{R} \mathbf{T}_k \mathbf{R}).
\end{aligned} \tag{3.84}$$

Applying the following matrix identity (see, e.g., [Searle, 1982]):

$$\text{tr}(\mathbf{A}^T \mathbf{B}) = (\text{vec} \mathbf{A})^T \text{vec} \mathbf{B}, \tag{3.85}$$

to eqn. (3.84), we get

$$\begin{aligned}
\frac{\partial j_{ij}}{\partial \theta_k} &= [\mathbf{D}j_{ij}(\theta_k)]_{1 \times 1} = -\frac{1}{2} [\text{vec}(\mathbf{T}_j \mathbf{R} \mathbf{T}_i)]^T \text{vec}(\mathbf{R} \mathbf{T}_k \mathbf{R}) - \frac{1}{2} [\text{vec}(\mathbf{T}_i \mathbf{R} \mathbf{T}_j)]^T \text{vec}(\mathbf{R} \mathbf{T}_k \mathbf{R}) \\
&= -\frac{1}{2} \text{tr}(\mathbf{T}_i \mathbf{R} \mathbf{T}_j \mathbf{R} \mathbf{T}_k \mathbf{R}) - \frac{1}{2} \text{tr}(\mathbf{T}_j \mathbf{R} \mathbf{T}_i \mathbf{R} \mathbf{T}_k \mathbf{R}).
\end{aligned} \tag{3.86}$$

Applying further the following matrix identity (see, e.g., [Searle, 1982]):

$$\text{tr}(\mathbf{A}\mathbf{B}) = \text{tr}(\mathbf{B}^T \mathbf{A}^T), \tag{3.87}$$

to eqn. (3.86), we get

$$\begin{aligned}
\frac{\partial j_{ij}}{\partial \theta_k} &= [\mathbf{D}j_{ij}(\theta_k)]_{1 \times 1} = -\frac{1}{2} \text{tr}(\mathbf{T}_i \mathbf{R} \mathbf{T}_j \mathbf{R} \mathbf{T}_k \mathbf{R}) - \frac{1}{2} \text{tr}(\mathbf{T}_j \mathbf{R} \mathbf{T}_i \mathbf{R} \mathbf{T}_k \mathbf{R}) \\
&= -\frac{1}{2} \text{tr}(\mathbf{T}_i \mathbf{R} \mathbf{T}_j \mathbf{R} \mathbf{T}_k \mathbf{R}) - \frac{1}{2} \text{tr}[(\mathbf{R} \mathbf{T}_k \mathbf{R})^T (\mathbf{T}_j \mathbf{R} \mathbf{T}_i)^T] \\
&= -\frac{1}{2} \text{tr}(\mathbf{T}_i \mathbf{R} \mathbf{T}_j \mathbf{R} \mathbf{T}_k \mathbf{R}) - \frac{1}{2} \text{tr}(\mathbf{R} \mathbf{T}_k \mathbf{R} \mathbf{T}_i \mathbf{R} \mathbf{T}_j) \\
&= -\text{tr}(\mathbf{R} \mathbf{T}_i \mathbf{R} \mathbf{T}_j \mathbf{R} \mathbf{T}_k).
\end{aligned} \tag{3.88}$$

Finally, writing the results of eqn. (3.88) in matrix form, we find the first derivative of the information matrix,  $\mathbf{J}_R$ , with respect to the k-th variance component,  $\theta_k$ , to be:

$$\frac{\partial \mathbf{J}_R}{\partial \theta_k} = \left\{ \frac{\partial j_{ij}}{\partial \theta_k} \right\}_{i,j=1}^r = - \left\{ \frac{1}{m} \text{tr}(\mathbf{R}\mathbf{T}_i\mathbf{R}\mathbf{T}_j\mathbf{R}\mathbf{T}_k) \right\}_{i,j=1}^r = - \left\{ \frac{1}{m} f_{ij} \right\}_{i,j=1}^r. \quad (3.89)$$

The results of eqn. (3.89), together with eqn. (3.43), yield the final form of the first partial derivative of the natural logarithm of the noninformative prior probability density function with respect to the k-th variance component,  $\theta_k$ , given as:

$$\frac{1}{2} \frac{\partial (\ln |\mathbf{J}_R|)}{\partial \theta_k} = \frac{1}{2} \text{tr} \left( \mathbf{J}_R^{-1} \frac{\partial \mathbf{J}_R}{\partial \theta_k} \right) = - \frac{1}{2} \text{tr} \left[ \left( \left\{ \frac{1}{m} \frac{1}{2} \text{tr}(\mathbf{R}\mathbf{T}_i\mathbf{R}\mathbf{T}_j) \right\}_{i,j=1}^r \right)^{-1} \left\{ \frac{1}{m} \text{tr}(\mathbf{R}\mathbf{T}_i\mathbf{R}\mathbf{T}_j\mathbf{R}\mathbf{T}_k) \right\}_{i,j=1}^r \right]. \quad (3.90)$$

The derivative of the natural logarithm of the noninformative prior, eqn. (3.36), with respect to the vector of variance components follows with

$$\mathbf{s}_{\text{NO}} = \frac{1}{2} \frac{\partial (\ln |\mathbf{J}_R|)}{\partial \boldsymbol{\theta}} = \left\{ \frac{1}{2} \text{tr} \left( \mathbf{J}_R^{-1} \frac{\partial \mathbf{J}_R}{\partial \theta_k} \right) \right\}_{k=1}^r. \quad (3.91)$$

Taking into account the results given in eqn. (3.41), eqn. (3.90) and eqn. (3.91), we get the final expression for the vector of scores as:

$$\begin{aligned}
\mathbf{s}_N &= \frac{\partial l_N}{\partial \boldsymbol{\theta}} = \mathbf{s}_{N0} + \mathbf{s}_R = \left\{ \mathbf{c} \frac{1}{2} \operatorname{tr} \left( \mathbf{J}_R^{-1} \frac{\partial \mathbf{J}_R}{\partial \boldsymbol{\theta}_k} \right) \right\}_{k=1}^r + \left\{ \mathbf{c} - \frac{1}{2} \operatorname{tr}(\mathbf{R}\mathbf{T}_k) + \frac{1}{2} \mathbf{1}^T \mathbf{R}\mathbf{T}_k \mathbf{R}\mathbf{1} \right\}_{k=1}^r \\
&= \left\{ \mathbf{c} - \frac{1}{2} \operatorname{tr} \left[ \left( \left\{ \mathbf{m} \frac{1}{2} \operatorname{tr}(\mathbf{R}\mathbf{T}_i \mathbf{R}\mathbf{T}_j) \right\}_{i,j=1}^r \right)^{-1} \left\{ \mathbf{m} \operatorname{tr}(\mathbf{R}\mathbf{T}_i \mathbf{R}\mathbf{T}_j \mathbf{R}\mathbf{T}_k) \right\}_{i,j=1}^r \right] - \frac{1}{2} \operatorname{tr}(\mathbf{R}\mathbf{T}_k) + \frac{1}{2} \mathbf{1}^T \mathbf{R}\mathbf{T}_k \mathbf{R}\mathbf{1} \right\}_{k=1}^r \\
&= \mathbf{s}_{N0} - \left\{ \mathbf{c} \frac{1}{2} \operatorname{tr}(\mathbf{R}\mathbf{T}_k) \right\}_{k=1}^r + \mathbf{q}_R.
\end{aligned} \tag{3.92}$$

### 3.2.2 Derivation of the Information Matrix

The expression for the vector of scores, eqn. (3.92), will be utilized in this section for derivation of the Hessian matrix,  $\mathbf{H}_N$ , which is defined as an  $r$  by  $r$  matrix of second partial derivatives of the natural logarithm of the posterior probability density function,  $l_N$ , with respect to the  $k$ -th and  $m$ -th variance-covariance components,  $\theta_k$  and  $\theta_m$ :

$$\mathbf{H}_N = \left\{ \mathbf{m} \frac{\partial^2 l_N}{\partial \theta_k \partial \theta_m} \right\}_{k,m=1}^r = \frac{\partial^2 l_N}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}} = \frac{\partial \mathbf{s}_N}{\partial \boldsymbol{\theta}}. \tag{3.93}$$

Subsequently, the information matrix,  $\mathbf{J}_N$ , which is defined as a negative of the expected value (where the expectation is taken with respect to the vector of observations) of the Hessian matrix  $\mathbf{H}_N$ ,

$$\mathbf{J}_N = -\mathbf{E}(\mathbf{H}_N), \tag{3.94}$$

will be derived.

Considering the relationship given in eqn. (3.37), the Hessian matrix  $\mathbf{H}_N$  can be decomposed into the sum of the Hessian matrix of the natural logarithm of the noninformative prior probability density function,  $\ln[p(\theta)]$ , and the Hessian matrix of the marginal log-likelihood function  $l_R$ :

$$\mathbf{H}_N = \frac{\partial^2 \ln[p(\theta)]}{\partial \theta \partial \theta} + \frac{\partial^2 l_R}{\partial \theta \partial \theta} = \frac{1}{2} \frac{\partial^2 \ln|\mathbf{J}_R|}{\partial \theta \partial \theta} + \frac{\partial^2 l_R}{\partial \theta \partial \theta} = \mathbf{H}_{N0} + \mathbf{H}_R, \quad (3.95)$$

where  $\mathbf{H}_R$  is the Hessian matrix of the marginal log-likelihood function (see eqn. (2.57)).

The expression for the Hessian matrix of the natural logarithm of the noninformative prior probability density function,

$$\mathbf{H}_{N0} = \frac{\partial^2 \ln[p(\theta)]}{\partial \theta \partial \theta} = \frac{1}{2} \frac{\partial^2 \ln|\mathbf{J}_R|}{\partial \theta \partial \theta} = \frac{\partial \mathbf{s}_{N0}}{\partial \theta}, \quad (3.96)$$

is derived as follows. Using the previously derived expression for the first partial derivative of the natural logarithm of the noninformative prior probability density function, eqn. (3.90), and applying the following matrix identity (see, e.g., [Searle, 1982]):

$$\text{tr}(\mathbf{A}^T \mathbf{B}) = (\text{vec} \mathbf{A})^T \text{vec} \mathbf{B}, \quad (3.97)$$

we determine the second partial derivative as:

$$\begin{aligned}
\frac{1}{2} \frac{\partial^2 (\ln |\mathbf{J}_R|)}{\partial \theta_k \partial \theta_m} &= \frac{1}{2} \frac{\partial}{\partial \theta_m} \left[ \frac{\partial (\ln |\mathbf{J}_R|)}{\partial \theta_k} \right] = \frac{1}{2} \frac{\partial \text{tr} \left( \mathbf{J}_R^{-1} \frac{\partial \mathbf{J}_R}{\partial \theta_k} \right)}{\partial \theta_m} = \frac{1}{2} \frac{\partial \left[ \text{vec}(\mathbf{J}_R^{-1}) \right]^T \text{vec} \left( \frac{\partial \mathbf{J}_R}{\partial \theta_k} \right)}{\partial \theta_m} \\
&= \frac{1}{2} \frac{\partial \left[ \text{vec}(\mathbf{J}_R^{-1}) \right]^T}{\partial \theta_m} \text{vec} \left( \frac{\partial \mathbf{J}_R}{\partial \theta_k} \right) + \frac{1}{2} \left[ \text{vec}(\mathbf{J}_R^{-1}) \right]^T \frac{\partial \text{vec} \left( \frac{\partial \mathbf{J}_R}{\partial \theta_k} \right)}{\partial \theta_m} \\
&= \frac{1}{2} \left[ \text{vec} \left( \frac{\partial \mathbf{J}_R^{-1}}{\partial \theta_m} \right) \right]^T \text{vec} \left( \frac{\partial \mathbf{J}_R}{\partial \theta_k} \right) + \frac{1}{2} \left[ \text{vec}(\mathbf{J}_R^{-1}) \right]^T \text{vec} \left( \frac{\partial^2 \mathbf{J}_R}{\partial \theta_k \partial \theta_m} \right).
\end{aligned} \tag{3.98}$$

To develop eqn. (3.98) further, the following two partial derivatives have to be derived:

$$(1) \frac{\partial \mathbf{J}_R^{-1}}{\partial \theta_m}, \text{ and } (2) \frac{\partial^2 \mathbf{J}_R}{\partial \theta_k \partial \theta_m}.$$

**(1) Derivation of  $\frac{\partial \mathbf{J}_R^{-1}}{\partial \theta_m}$**

Applying the rule for taking the first derivative of an inverse of a regular symmetric matrix, given by Koch [1990] as:

$$\frac{\partial \mathbf{A}^{-1}}{\partial a_i} = -\mathbf{A}^{-1} \frac{\partial \mathbf{A}}{\partial a_i} \mathbf{A}^{-1}, \tag{3.99}$$

and using the results of eqn. (3.89), we get

$$\frac{\partial \mathbf{J}_R^{-1}}{\partial \theta_m} = -\mathbf{J}_R^{-1} \frac{\partial \mathbf{J}_R}{\partial \theta_m} \mathbf{J}_R^{-1} = \mathbf{J}_R^{-1} \left\{ \text{tr} \left( \mathbf{R} \mathbf{T}_i \mathbf{R} \mathbf{T}_j \mathbf{R} \mathbf{T}_m \right) \right\}_{i,j=1}^r \mathbf{J}_R^{-1}. \tag{3.100}$$

**(2) Derivation of  $\frac{\partial^2 \mathbf{J}_R}{\partial \theta_k \partial \theta_m}$**

Taking derivatives of all terms in eqn. (3.89) with respect to the m-th variance component,  $\theta_m$ , we get

$$\frac{\partial^2 \mathbf{J}_R}{\partial \theta_k \partial \theta_m} = \frac{\partial}{\partial \theta_m} \left( \frac{\partial \mathbf{J}_R}{\partial \theta_k} \right) = - \frac{\partial \left\{ {}_m \text{tr}(\mathbf{RT}_i \mathbf{RT}_j \mathbf{RT}_k) \right\}_{i,j=1}^r}{\partial \theta_m} = - \frac{\partial \left\{ {}_m f_{ij} \right\}_{i,j=1}^r}{\partial \theta_m} = - \left\{ {}_m \frac{\partial f_{ij}}{\partial \theta_m} \right\}_{i,j=1}^r. \quad (3.101)$$

The first partial derivative of  $f_{ij}$  with respect to the m-th variance component,  $\theta_m$ , is computed by applying the chain rule, which results in

$$\frac{\partial f_{ij}}{\partial \theta_m} = \left[ \mathbf{D}f_{ij}(\theta_m) \right]_{1 \times 1} = \frac{\partial f_{ij}}{\partial \text{vec} \mathbf{R}} \frac{\partial \text{vec} \mathbf{R}}{\partial \theta_m} = \left[ \mathbf{D}f_{ij}(\mathbf{R}) \right]_{1 \times m^2} \left[ \mathbf{D}\mathbf{R}(\theta_m) \right]_{m^2 \times 1}. \quad (3.102)$$

Derivation of an expression for the derivative  $\frac{\partial f_{ij}}{\partial \theta_m}$ , eqn. (3.102), will be broken down

into two steps. In the first step the derivative matrix  $\frac{\partial f_{ij}}{\partial \text{vec} \mathbf{R}} = \left[ \mathbf{D}f_{ij}(\mathbf{R}) \right]_{1 \times m^2}$  will be

derived. In the second step an expression for the derivative matrix

$\frac{\partial \text{vec} \mathbf{R}}{\partial \theta_m} = \left[ \mathbf{D}\mathbf{R}(\theta_m) \right]_{m^2 \times 1}$  will be found.

(2a) Derivation of  $\frac{\partial f_{ij}}{\partial \text{vec} \mathbf{R}} = [\mathbf{D}f_{ij}(\mathbf{R})]_{1 \times m^2}$

The differential and the Jacobian (derivative) matrix of the scalar function  $f_{ij}$  of the matrix  $\mathbf{R}$  can be computed by means of the “identification theorem for matrix functions”. For a scalar function of a matrix this theorem implies that, in order to find the derivative matrix  $\mathbf{D}f_{ij}(\mathbf{R})$ , the differential  $df_{ij}$  has to be converted into the following form:

$$df_{ij} = \text{tr}(\mathbf{X}^T d\mathbf{R}) = (\text{vec} \mathbf{X})^T d(\text{vec} \mathbf{R}) = [\mathbf{D}f_{ij}(\mathbf{R})]_{1 \times m^2} d(\text{vec} \mathbf{R})_{m^2 \times 1}, \quad (3.103)$$

where  $[\mathbf{D}f_{ij}(\mathbf{R})]_{1 \times m^2}$  is the Jacobian (derivative) matrix, and  $\mathbf{X}$  is some matrix such that

$$[\mathbf{D}f_{ij}(\mathbf{R})]_{1 \times m^2} = (\text{vec} \mathbf{X})^T = \frac{\partial f_{ij}}{\partial \text{vec} \mathbf{R}}_{m^2 \times 1}. \quad (3.104)$$

Differentiating both sides of the following equation:

$$f_{ij}(\mathbf{R}) = \text{tr}(\mathbf{R} \mathbf{T}_i \mathbf{R} \mathbf{T}_j \mathbf{R} \mathbf{T}_k), \quad (3.105)$$

and applying the matrix identity (see, e.g., [Searle, 1982]):

$$\text{tr}(\mathbf{A}^T \mathbf{B}) = (\text{vec} \mathbf{A})^T \text{vec} \mathbf{B}, \quad (3.106)$$

yields the following expression for the differential  $df_{ij}$ :



$$\begin{aligned}
df_{ij} &= d\left[\text{tr}\left(\mathbf{R}\mathbf{T}_i\mathbf{R}\mathbf{T}_j\mathbf{R}\mathbf{T}_k\right)\right] = \text{tr}\left[d\left(\mathbf{R}\mathbf{T}_i\mathbf{R}\mathbf{T}_j\mathbf{R}\mathbf{T}_k\right)\right] \\
&= \text{tr}\left[d\mathbf{R}\mathbf{T}_i\mathbf{R}\mathbf{T}_j\mathbf{R}\mathbf{T}_k + \mathbf{R}\mathbf{T}_i d\mathbf{R}\mathbf{T}_j\mathbf{R}\mathbf{T}_k + \mathbf{R}\mathbf{T}_i\mathbf{R}\mathbf{T}_j d\mathbf{R}\mathbf{T}_k\right] \\
&= \text{tr}\left(\mathbf{T}_i\mathbf{R}\mathbf{T}_j\mathbf{R}\mathbf{T}_k d\mathbf{R}\right) + \text{tr}\left(\mathbf{T}_j\mathbf{R}\mathbf{T}_k\mathbf{R}\mathbf{T}_i d\mathbf{R}\right) + \text{tr}\left(\mathbf{T}_k\mathbf{R}\mathbf{T}_i\mathbf{R}\mathbf{T}_j d\mathbf{R}\right) \\
&= \left[\text{vec}\left(\mathbf{T}_k\mathbf{R}\mathbf{T}_j\mathbf{R}\mathbf{T}_i\right)\right]^T d(\text{vec}\mathbf{R}) + \left[\text{vec}\left(\mathbf{T}_i\mathbf{R}\mathbf{T}_k\mathbf{R}\mathbf{T}_j\right)\right]^T d(\text{vec}\mathbf{R}) + \left[\text{vec}\left(\mathbf{T}_j\mathbf{R}\mathbf{T}_i\mathbf{R}\mathbf{T}_k\right)\right]^T d(\text{vec}\mathbf{R}) \\
&= \left[\text{vec}\left(\mathbf{T}_k\mathbf{R}\mathbf{T}_j\mathbf{R}\mathbf{T}_i + \mathbf{T}_i\mathbf{R}\mathbf{T}_k\mathbf{R}\mathbf{T}_j + \mathbf{T}_j\mathbf{R}\mathbf{T}_i\mathbf{R}\mathbf{T}_k\right)\right]^T d(\text{vec}\mathbf{R}).
\end{aligned} \tag{3.107}$$

Applying once again the “identification theorem for matrix functions”, eqn. (3.103), to the expression for the differential  $df_{ij}$ , eqn. (3.107), we finally find the derivative matrix  $\mathbf{D}f_{ij}$  to be:

$$\frac{\partial f_{ij}}{\partial \text{vec}\mathbf{R}} = \left[\mathbf{D}f_{ij}(\mathbf{R})\right]_{1 \times m^2} = \left[\text{vec}\left(\mathbf{T}_k\mathbf{R}\mathbf{T}_j\mathbf{R}\mathbf{T}_i + \mathbf{T}_i\mathbf{R}\mathbf{T}_k\mathbf{R}\mathbf{T}_j + \mathbf{T}_j\mathbf{R}\mathbf{T}_i\mathbf{R}\mathbf{T}_k\right)\right]_{1 \times m^2}^T. \tag{3.108}$$

(2b) Derivation of  $\frac{\partial \text{vec}\mathbf{R}}{\partial \theta_m} = \left[\mathbf{D}\mathbf{R}(\theta_m)\right]_{m^2 \times 1}$

As proven before (see eqn. (3.83)), this derivative can be expressed as:

$$\frac{\partial \text{vec}\mathbf{R}}{\partial \theta_m} = \left[\mathbf{D}\mathbf{R}(\theta_m)\right]_{m^2 \times 1} = -\left[\text{vec}\left(\mathbf{R}\mathbf{T}_m\mathbf{R}\right)\right]_{m^2 \times 1}. \tag{3.109}$$

Coming back to the original objective, i.e., derivation of an expression for the first partial derivative of the scalar function  $f_{ij}$  with respect to the  $m$ -th variance component,  $\theta_m$ , we get:

$$\begin{aligned}
\frac{\partial f_{ij}}{\partial \theta_m} &= \left[ \mathbf{D}f_{ij}(\theta_m) \right]_{1 \times 1} = \frac{\partial f_{ij}}{\partial \text{vec} \mathbf{R}} \frac{\partial \text{vec} \mathbf{R}}{\partial \theta_m} = \left[ \mathbf{D}f_{ij}(\mathbf{R}) \right]_{1 \times m^2} \left[ \mathbf{D}\mathbf{R}(\theta_m) \right]_{m^2 \times 1} \\
&= \left[ \text{vec}(\mathbf{T}_k \mathbf{R} \mathbf{T}_j \mathbf{R} \mathbf{T}_i + \mathbf{T}_i \mathbf{R} \mathbf{T}_k \mathbf{R} \mathbf{T}_j + \mathbf{T}_j \mathbf{R} \mathbf{T}_i \mathbf{R} \mathbf{T}_k) \right]_{1 \times m^2}^T \left[ -\text{vec}(\mathbf{R} \mathbf{T}_m \mathbf{R}) \right]_{m^2 \times 1} \\
&= - \left[ \text{vec}(\mathbf{T}_k \mathbf{R} \mathbf{T}_j \mathbf{R} \mathbf{T}_i) \right]^T \text{vec}(\mathbf{R} \mathbf{T}_m \mathbf{R}) - \left[ \text{vec}(\mathbf{T}_i \mathbf{R} \mathbf{T}_k \mathbf{R} \mathbf{T}_j) \right]^T \text{vec}(\mathbf{R} \mathbf{T}_m \mathbf{R}) \\
&\quad - \left[ \text{vec}(\mathbf{T}_j \mathbf{R} \mathbf{T}_i \mathbf{R} \mathbf{T}_k) \right]^T \text{vec}(\mathbf{R} \mathbf{T}_m \mathbf{R}).
\end{aligned} \tag{3.110}$$

Applying the matrix identity given in eqn. (3.106) to eqn. (3.110) we get:

$$\begin{aligned}
\frac{\partial f_{ij}}{\partial \theta_m} &= \left[ \mathbf{D}f_{ij}(\theta_m) \right]_{1 \times 1} = - \left[ \text{vec}(\mathbf{T}_k \mathbf{R} \mathbf{T}_j \mathbf{R} \mathbf{T}_i) \right]^T \text{vec}(\mathbf{R} \mathbf{T}_m \mathbf{R}) - \left[ \text{vec}(\mathbf{T}_i \mathbf{R} \mathbf{T}_k \mathbf{R} \mathbf{T}_j) \right]^T \text{vec}(\mathbf{R} \mathbf{T}_m \mathbf{R}) \\
&\quad - \left[ \text{vec}(\mathbf{T}_j \mathbf{R} \mathbf{T}_i \mathbf{R} \mathbf{T}_k) \right]^T \text{vec}(\mathbf{R} \mathbf{T}_m \mathbf{R}) \\
&= - \text{tr} \left[ (\mathbf{T}_k \mathbf{R} \mathbf{T}_j \mathbf{R} \mathbf{T}_i)^T \mathbf{R} \mathbf{T}_m \mathbf{R} \right] - \text{tr} \left[ (\mathbf{T}_i \mathbf{R} \mathbf{T}_k \mathbf{R} \mathbf{T}_j)^T \mathbf{R} \mathbf{T}_m \mathbf{R} \right] - \text{tr} \left[ (\mathbf{T}_j \mathbf{R} \mathbf{T}_i \mathbf{R} \mathbf{T}_k)^T \mathbf{R} \mathbf{T}_m \mathbf{R} \right] \\
&= - \text{tr}(\mathbf{T}_i \mathbf{R} \mathbf{T}_j \mathbf{R} \mathbf{T}_k \mathbf{R} \mathbf{T}_m \mathbf{R}) - \text{tr}(\mathbf{T}_j \mathbf{R} \mathbf{T}_k \mathbf{R} \mathbf{T}_i \mathbf{R} \mathbf{T}_m \mathbf{R}) - \text{tr}(\mathbf{T}_k \mathbf{R} \mathbf{T}_i \mathbf{R} \mathbf{T}_j \mathbf{R} \mathbf{T}_m \mathbf{R}).
\end{aligned} \tag{3.111}$$

Finally, writing the results of eqn. (3.111) in the matrix form, we find the second derivative of the information matrix,  $\mathbf{J}_R$ , with respect to the k-th and m-th variance components,  $\theta_k$  and  $\theta_m$ , to be:

$$\begin{aligned}
\frac{\partial^2 \mathbf{J}_R}{\partial \theta_k \partial \theta_m} &= - \left\{ \sum_{i,j=1}^m \frac{\partial f_{ij}}{\partial \theta_m} \right\}^r \\
&= \left\{ \sum_{i,j=1}^m \text{tr}(\mathbf{T}_i \mathbf{R} \mathbf{T}_j \mathbf{R} \mathbf{T}_k \mathbf{R} \mathbf{T}_m \mathbf{R}) + \text{tr}(\mathbf{T}_j \mathbf{R} \mathbf{T}_k \mathbf{R} \mathbf{T}_i \mathbf{R} \mathbf{T}_m \mathbf{R}) + \text{tr}(\mathbf{T}_k \mathbf{R} \mathbf{T}_i \mathbf{R} \mathbf{T}_j \mathbf{R} \mathbf{T}_m \mathbf{R}) \right\}^r \quad (3.112) \\
&= \left\{ \sum_{i,j=1}^m \text{tr}(\mathbf{R} \mathbf{T}_i \mathbf{R} \mathbf{T}_j \mathbf{R} \mathbf{T}_k \mathbf{R} \mathbf{T}_m + \mathbf{R} \mathbf{T}_j \mathbf{R} \mathbf{T}_k \mathbf{R} \mathbf{T}_i \mathbf{R} \mathbf{T}_m + \mathbf{R} \mathbf{T}_k \mathbf{R} \mathbf{T}_i \mathbf{R} \mathbf{T}_j \mathbf{R} \mathbf{T}_m) \right\}^r.
\end{aligned}$$

Applying the results of eqn. (3.100) to the expression for the second partial derivative of the natural logarithm of the noninformative prior probability density function, eqn. (3.98), we get:

$$\begin{aligned}
\frac{1}{2} \frac{\partial^2 (\ln |\mathbf{J}_R|)}{\partial \theta_k \partial \theta_m} &= \frac{1}{2} \left[ \text{vec} \left( \frac{\partial \mathbf{J}_R^{-1}}{\partial \theta_m} \right) \right]^T \text{vec} \left( \frac{\partial \mathbf{J}_R}{\partial \theta_k} \right) + \frac{1}{2} \left[ \text{vec}(\mathbf{J}_R^{-1}) \right]^T \text{vec} \left( \frac{\partial^2 \mathbf{J}_R}{\partial \theta_k \partial \theta_m} \right) \\
&= -\frac{1}{2} \left[ \text{vec} \left( \mathbf{J}_R^{-1} \frac{\partial \mathbf{J}_R}{\partial \theta_m} \mathbf{J}_R^{-1} \right) \right]^T \text{vec} \left( \frac{\partial \mathbf{J}_R}{\partial \theta_k} \right) + \frac{1}{2} \left[ \text{vec}(\mathbf{J}_R^{-1}) \right]^T \text{vec} \left( \frac{\partial^2 \mathbf{J}_R}{\partial \theta_k \partial \theta_m} \right). \quad (3.113)
\end{aligned}$$

Applying the matrix identity given in eqn. (3.106) to eqn. (3.113), taking into account that the matrices  $\mathbf{J}_R^{-1}$ ,  $\frac{\partial \mathbf{J}_R}{\partial \theta_k}$ , and  $\frac{\partial \mathbf{J}_R}{\partial \theta_m}$  are symmetric, and making use of the previously derived expressions (eqn. (3.43), eqn. (3.89), and eqn. (3.112)), we find that the second partial derivative of the natural logarithm of the noninformative prior probability density function is given as:

$$\begin{aligned}
\frac{1}{2} \frac{\partial^2 (\ln |\mathbf{J}_R|)}{\partial \theta_k \partial \theta_m} &= -\frac{1}{2} \left[ \text{vec} \left( \mathbf{J}_R^{-1} \frac{\partial \mathbf{J}_R}{\partial \theta_m} \mathbf{J}_R^{-1} \right) \right]^T \text{vec} \left( \frac{\partial \mathbf{J}_R}{\partial \theta_k} \right) + \frac{1}{2} \left[ \text{vec}(\mathbf{J}_R^{-1}) \right]^T \text{vec} \left( \frac{\partial^2 \mathbf{J}_R}{\partial \theta_k \partial \theta_m} \right) \\
&= -\frac{1}{2} \text{tr} \left( \mathbf{J}_R^{-1} \frac{\partial \mathbf{J}_R}{\partial \theta_m} \mathbf{J}_R^{-1} \frac{\partial \mathbf{J}_R}{\partial \theta_k} \right) + \frac{1}{2} \text{tr} \left( \mathbf{J}_R^{-1} \frac{\partial^2 \mathbf{J}_R}{\partial \theta_k \partial \theta_m} \right) \\
&= \frac{1}{2} \text{tr} \left[ \mathbf{J}_R^{-1} \left( \frac{\partial^2 \mathbf{J}_R}{\partial \theta_k \partial \theta_m} - \frac{\partial \mathbf{J}_R}{\partial \theta_m} \mathbf{J}_R^{-1} \frac{\partial \mathbf{J}_R}{\partial \theta_k} \right) \right]
\end{aligned} \tag{3.114}$$

where:

$$\mathbf{J}_R = \left\{ \frac{1}{m} \text{tr}(\mathbf{R}\mathbf{T}_i \mathbf{R}\mathbf{T}_j) \right\}_{i,j=1}^r,$$

$$\frac{\partial \mathbf{J}_R}{\partial \theta_k} = - \left\{ \text{tr}(\mathbf{R}\mathbf{T}_i \mathbf{R}\mathbf{T}_j \mathbf{R}\mathbf{T}_k) \right\}_{i,j=1}^r,$$

$$\frac{\partial \mathbf{J}_R}{\partial \theta_m} = - \left\{ \text{tr}(\mathbf{R}\mathbf{T}_i \mathbf{R}\mathbf{T}_j \mathbf{R}\mathbf{T}_m) \right\}_{i,j=1}^r, \text{ and}$$

$$\frac{\partial^2 \mathbf{J}_R}{\partial \theta_k \partial \theta_m} = \left\{ \text{tr}(\mathbf{R}\mathbf{T}_i \mathbf{R}\mathbf{T}_j \mathbf{R}\mathbf{T}_k \mathbf{R}\mathbf{T}_m + \mathbf{R}\mathbf{T}_j \mathbf{R}\mathbf{T}_k \mathbf{R}\mathbf{T}_i \mathbf{R}\mathbf{T}_m + \mathbf{R}\mathbf{T}_k \mathbf{R}\mathbf{T}_i \mathbf{R}\mathbf{T}_j \mathbf{R}\mathbf{T}_m) \right\}_{i,j=1}^r.$$

The Hessian matrix of the natural logarithm of the noninformative prior probability density function follows with:

$$\mathbf{H}_{N0} = \left\{ \frac{1}{m} \text{tr} \left[ \mathbf{J}_R^{-1} \left( \frac{\partial^2 \mathbf{J}_R}{\partial \theta_k \partial \theta_m} - \frac{\partial \mathbf{J}_R}{\partial \theta_m} \mathbf{J}_R^{-1} \frac{\partial \mathbf{J}_R}{\partial \theta_k} \right) \right] \right\}_{k,m=1}^r. \tag{3.115}$$

Taking into account the previous developments (eqn. (2.57), eqn. (3.45), and eqn. (3.115)), the Hessian matrix of the natural logarithm of the posterior probability density function,  $l_N$ , is given as follows:

$$\mathbf{H}_N = \mathbf{H}_{N0} + \mathbf{H}_R$$

$$= \left\{ \frac{1}{2} \text{tr} \left[ \mathbf{J}_R^{-1} \left( \frac{\partial^2 \mathbf{J}_R}{\partial \theta_k \partial \theta_m} - \frac{\partial \mathbf{J}_R}{\partial \theta_m} \mathbf{J}_R^{-1} \frac{\partial \mathbf{J}_R}{\partial \theta_k} \right) \right] \right\}_{k,m=1}^r + \left\{ \frac{1}{2} \text{tr}(\mathbf{R}\mathbf{T}_k \mathbf{R}\mathbf{T}_m) - \mathbf{l}^T \mathbf{R}\mathbf{T}_k \mathbf{R}\mathbf{T}_m \mathbf{l} \right\}_{k,m=1}^r, \quad (3.116)$$

where  $\mathbf{H}_R$  is the marginal likelihood Hessian matrix (see eqn. (2.57)).

The information matrix is subsequently derived from the Hessian matrix, eqn. (3.116), as:

$$\mathbf{J}_N = -\mathbf{E}(\mathbf{H}_N) = -\mathbf{E}(\mathbf{H}_{N0}) - \mathbf{E}(\mathbf{H}_R) = \mathbf{J}_{N0} + \mathbf{J}_R, \quad (3.117)$$

where  $\mathbf{J}_R$  is the marginal likelihood information matrix (see eqn. (2.61)):

$$\mathbf{J}_R = \left\{ \frac{1}{2} \text{tr}(\mathbf{R}\mathbf{T}_k \mathbf{R}\mathbf{T}_m) \right\}_{k,m=1}^r. \quad (3.118)$$

Since the expectations in eqn. (3.117) are taken with respect to the vector of observations,

$\mathbf{l}$ , the expected value of the Hessian matrix  $\mathbf{H}_{N0}$  is:

$$\mathbf{E}(\mathbf{H}_{N0}) = \mathbf{H}_{N0}. \quad (3.119)$$

Accounting for the results of eqn. (3.114), eqn. (3.117), and eqn. (3.119), we can finally express the information matrix,  $\mathbf{J}_N$ , as:

$$\begin{aligned}
\mathbf{J}_N &= \mathbf{J}_{N0} + \mathbf{J}_R \\
&= \left\{ m \frac{1}{2} \text{tr} \left[ \mathbf{J}_R^{-1} \left( \frac{\partial \mathbf{J}_R}{\partial \theta_m} \mathbf{J}_R^{-1} \frac{\partial \mathbf{J}_R}{\partial \theta_k} - \frac{\partial^2 \mathbf{J}_R}{\partial \theta_k \partial \theta_m} \right) \right] \right\}_{k,m=1}^r + \left\{ m \frac{1}{2} \text{tr}(\mathbf{R}\mathbf{T}_k\mathbf{R}\mathbf{T}_m) \right\}_{k,m=1}^r \\
&= \left\{ m \frac{1}{2} \text{tr} \left( \mathbf{J}_R^{-1} \frac{\partial \mathbf{J}_R}{\partial \theta_m} \mathbf{J}_R^{-1} \frac{\partial \mathbf{J}_R}{\partial \theta_k} \right) \right\}_{k,m=1}^r + \left\{ m - \frac{1}{2} \text{tr} \left( \mathbf{J}_R^{-1} \frac{\partial^2 \mathbf{J}_R}{\partial \theta_k \partial \theta_m} \right) \right\}_{k,m=1}^r + \left\{ m \frac{1}{2} \text{tr}(\mathbf{R}\mathbf{T}_k\mathbf{R}\mathbf{T}_m) \right\}_{k,m=1}^r \\
&= \left\{ m \frac{1}{2} \text{tr} \left[ \left( \left\{ m \frac{1}{2} \text{tr}(\mathbf{R}\mathbf{T}_i\mathbf{R}\mathbf{T}_j) \right\}_{i,j=1}^r \right)^{-1} \left\{ m \text{tr}(\mathbf{R}\mathbf{T}_i\mathbf{R}\mathbf{T}_j\mathbf{R}\mathbf{T}_m) \right\}_{i,j=1}^r \right. \right. \\
&\quad \times \left. \left. \left( \left\{ m \frac{1}{2} \text{tr}(\mathbf{R}\mathbf{T}_i\mathbf{R}\mathbf{T}_j) \right\}_{i,j=1}^r \right)^{-1} \left\{ m \text{tr}(\mathbf{R}\mathbf{T}_i\mathbf{R}\mathbf{T}_j\mathbf{R}\mathbf{T}_k) \right\}_{i,j=1}^r \right) \right] \right\}_{k,m=1}^r \\
&\quad + \left\{ m - \frac{1}{2} \text{tr} \left[ \left( \left\{ m \frac{1}{2} \text{tr}(\mathbf{R}\mathbf{T}_i\mathbf{R}\mathbf{T}_j) \right\}_{i,j=1}^r \right)^{-1} \right. \right. \\
&\quad \times \left. \left. \left\{ m \text{tr}(\mathbf{R}\mathbf{T}_i\mathbf{R}\mathbf{T}_j\mathbf{R}\mathbf{T}_k\mathbf{R}\mathbf{T}_m + \mathbf{R}\mathbf{T}_j\mathbf{R}\mathbf{T}_k\mathbf{R}\mathbf{T}_i\mathbf{R}\mathbf{T}_m + \mathbf{R}\mathbf{T}_k\mathbf{R}\mathbf{T}_i\mathbf{R}\mathbf{T}_j\mathbf{R}\mathbf{T}_m) \right\}_{i,j=1}^r \right] \right\}_{k,m=1}^r \\
&\quad + \left\{ m \frac{1}{2} \text{tr}(\mathbf{R}\mathbf{T}_k\mathbf{R}\mathbf{T}_m) \right\}_{k,m=1}^r.
\end{aligned} \tag{3.120}$$

### 3.2.3 Application of the Fisher Scoring Method

Application of the Fisher scoring method yields the following expression:

$$\begin{aligned}
\hat{\theta}_{(m+1)} &= \hat{\theta}_{(m)} + \left[ \mathbf{J}_N(\hat{\theta}_{(m)}) \right]^{-1} \mathbf{s}_N(\hat{\theta}_{(m)}) \\
&= \hat{\theta}_{(m)} + \left[ \mathbf{J}_{N0}(\hat{\theta}_{(m)}) + \mathbf{J}_R(\hat{\theta}_{(m)}) \right]^{-1} \left[ \mathbf{s}_{N0}(\hat{\theta}_{(m)}) + \mathbf{s}_R(\hat{\theta}_{(m)}) \right],
\end{aligned} \tag{3.121}$$

where:  $\hat{\theta}_{(m)}$  is the GML estimate of the vector of variance components from the m-th

iterative step,

$\mathbf{J}_N$  is the information matrix, eqn. (3.120), and

$\mathbf{s}_N$  is the vector of GML scores, eqn. (3.92).

Taking into account that the vector of marginal likelihood scores (see eqn. (3.27) and eqn. (3.29)) can be expressed as:

$$\mathbf{s}_R = -\mathbf{J}_R \boldsymbol{\theta} + \mathbf{q}_R, \quad (3.122)$$

and subtracting and adding the product  $\mathbf{J}_{N0}(\hat{\theta}_{(m)})\hat{\theta}_{(m)}$  from the vector of GML scores,

$\mathbf{s}_N(\hat{\theta}_{(m)})$ , we get:

$$\begin{aligned} \hat{\theta}_{(m+1)} &= \hat{\theta}_{(m)} + \left[ \mathbf{J}_{N0}(\hat{\theta}_{(m)}) + \mathbf{J}_R(\hat{\theta}_{(m)}) \right]^{-1} \\ &\times \left[ \mathbf{s}_{N0}(\hat{\theta}_{(m)}) + \mathbf{q}_R(\hat{\theta}_{(m)}) - \mathbf{J}_R(\hat{\theta}_{(m)})\hat{\theta}_{(m)} - \mathbf{J}_{N0}(\hat{\theta}_{(m)})\hat{\theta}_{(m)} + \mathbf{J}_{N0}(\hat{\theta}_{(m)})\hat{\theta}_{(m)} \right] \\ &= \left[ \mathbf{J}_{N0}(\hat{\theta}_{(m)}) + \mathbf{J}_R(\hat{\theta}_{(m)}) \right]^{-1} \left[ \mathbf{s}_{N0}(\hat{\theta}_{(m)}) + \mathbf{q}_R(\hat{\theta}_{(m)}) + \mathbf{J}_{N0}(\hat{\theta}_{(m)})\hat{\theta}_{(m)} \right] \\ &= \left[ \mathbf{J}_N(\hat{\theta}_{(m)}) \right]^{-1} \mathbf{q}_N(\hat{\theta}_{(m)}). \end{aligned} \quad (3.123)$$

In the limit, eqn. (3.121) yields the GML estimator (with the noninformative prior),  $\hat{\theta}$ , of the vector of variance components. The asymptotic covariance matrix of such GML estimator (see Appendix B) is given by

$$\text{var}(\hat{\theta}) \approx \left[ \mathbf{J}_N(\hat{\theta}) \right]^{-1}. \quad (3.124)$$

### **3.3 Application of the Principle of Mixed Estimation to the Dispersion-Mean Model**

In this chapter the dispersion-mean model is analyzed in the context of the introduction of prior information by means of weighted constraints on the vector of variance-covariance components. A link between the dispersion-mean model with weighted constraints and the GML estimation method is subsequently explored; the posterior probability density function that would result in the estimating equations of the dispersion-mean model with weighted constraints if the Fisher scoring method were applied to the GML optimization problem, eqn. (3.2), with such posterior probability density function, is found.

#### ***3.3.1 Principle of the Dispersion-Mean Model***

The general mixed model, eqn. (2.1), can be reformulated in such a way that it is linear in terms of the vector of variance-covariance components,  $\theta$ . Such a model, which was - according to Searle et al. [1992] - first introduced by Pukelsheim [1974], is called the dispersion-mean model. Similarly to the criteria-based methods (see Section 2.4), the objective is to find a quadratic estimator,  $\mathbf{I}^T \mathbf{M} \mathbf{I}$  (where  $\mathbf{M}$  is a symmetric matrix to be determined), of a linear combination of the variance-covariance components,  $\mathbf{p}^T \theta$ , subject to the constraint of translation invariance, eqn. (2.92). Under these assumptions, the general mixed model, eqn. (2.1), can be reformulated in the following way [Searle et al., 1992]:



with a projection matrix defined as

$$\mathbf{N} = \mathbf{I} - \mathbf{A}\mathbf{A}^+ = \mathbf{I} - \mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T = \mathbf{N}^T = \mathbf{N}^2, \quad (3.125)$$

and using an identity,

$$\text{vec}(\mathbf{ABC}) = (\mathbf{C}^T \otimes \mathbf{A})\text{vec}\mathbf{B} \quad (3.126)$$

we get,

$$\begin{aligned} \mathbf{I}^T\mathbf{M}\mathbf{I} &= \mathbf{I}^T\mathbf{N}\mathbf{M}\mathbf{N}\mathbf{I} = \text{vec}(\mathbf{I}^T\mathbf{N}\mathbf{M}\mathbf{N}\mathbf{I}) = [(\mathbf{N}\mathbf{I})^T \otimes \mathbf{I}^T\mathbf{N}]\text{vec}\mathbf{M} \\ &= (\text{vec}\mathbf{M})^T(\mathbf{N}\mathbf{I} \otimes \mathbf{N}\mathbf{I}) = (\text{vec}\mathbf{M})^T\mathbf{l}. \end{aligned} \quad (3.127)$$

It can be proven [Searle et al., 1992] that the expected value of the vector  $\mathbf{l}$  can be expressed as:

$$\mathbf{E}(\mathbf{l}) = \mathcal{A}\boldsymbol{\theta}, \text{ or} \quad (3.128)$$

$$\mathbf{l} = \mathcal{A}\boldsymbol{\theta} + \mathbf{v}, \quad (3.129)$$

where:

$$\mathcal{A} = \left\{ \text{vec}(\mathbf{N}\mathbf{T}_i\mathbf{N}) \right\}_{i=1}^r, \text{ and} \quad (3.130)$$

$\mathbf{v}$  is a vector of errors. The model described by eqn. (3.129) is called a dispersion-mean model.

It can also be proven [Searle et al., 1992] that if the vector of observations,  $\mathbf{l}$ , is normally distributed, then the covariance matrix of the vector of observables,  $\mathbf{l}$ , in the dispersion-mean model can be expressed as

$$\text{var}(\mathbf{t}) = \mathbf{C}_t = (\mathbf{N}\mathbf{C}_\theta\mathbf{N} \otimes \mathbf{N}\mathbf{C}_\theta\mathbf{N})(\mathbf{I} + \mathbf{S}), \quad (3.131)$$

where  $\mathbf{S}$  is a vec permutation matrix (see [Searle et al., 1992]). A generalized inverse of this covariance matrix, eqn. (3.131), being a weight matrix of the dispersion-mean model, is given by Searle et al. [1992] as:

$$\mathbf{P}_t = \mathbf{C}_t^- = \frac{1}{4}(\mathbf{I} + \mathbf{S}) \left[ (\mathbf{N}\mathbf{C}_\theta\mathbf{N})^- \otimes (\mathbf{N}\mathbf{C}_\theta\mathbf{N})^- \right]. \quad (3.132)$$

Application of the method of least squares to the dispersion-mean model, eqn. (3.129), with the weight matrix given by eqn. (3.132) yields the following normal equations [Searle et al., 1992]:

$$\mathcal{A}^T \mathbf{C}_t^- \mathcal{A} \theta = \mathcal{A}^T \mathbf{C}_t^- \mathbf{t}, \quad (3.133)$$

with

$$\mathcal{A}^T \mathbf{C}_t^- \mathcal{A} = \left\{ \frac{1}{2} \text{tr}(\mathbf{R}\mathbf{T}_i \mathbf{R}\mathbf{T}_j) \right\}_{i,j=1}^r = \mathbf{J}_R, \text{ and} \quad (3.134)$$

$$\mathcal{A}^T \mathbf{C}_t^- \mathbf{t} = \left\{ \frac{1}{2} \mathbf{1}^T \mathbf{R}\mathbf{T}_i \mathbf{R}\mathbf{1} \right\}_{i=1}^r = \mathbf{q}_R, \quad (3.135)$$

where:  $\mathbf{J}_R$  is the REML information matrix (see eqn. (2.61)), and

$\mathbf{q}_R$  is the vector of quadratic forms in the REML scoring equations, eqn. (2.63).

Since the elements of both the dispersion-mean model design matrix,  $\mathcal{A}$ , and the covariance matrix,  $\mathbf{C}_t$ , are functions of the vector of variance-covariance components,  $\theta$ ,

the normal equations, eqn. (3.133), have to be solved iteratively. The solution to the normal equations, in the  $m$ -th iterative step, is then:

$$\hat{\theta}_{(m+1)} = \left[ \mathbf{J}_R(\hat{\theta}_{(m)}) \right]^{-1} \mathbf{q}_R(\hat{\theta}_{(m)}). \quad (3.136)$$

These iterative equations, eqn. (3.136), are exactly the same as the REML scoring equations, eqn. (2.63), which result from application of the Fisher scoring method to the marginal log-likelihood function, eqn. (2.45).

### 3.3.2 The Dispersion-Mean Model With Weighted Constrains

In the case of the Generalized Least Squares estimation of the vector of location parameters,  $\mathbf{x}$ , in the model,

$$\mathbf{l} = \mathbf{A}\mathbf{x} + \mathbf{v}, \quad \mathbf{v} \sim (\mathbf{0}, \mathbf{C}_1), \quad (3.137)$$

where  $\mathbf{C}_1$  is a known positive definite covariance matrix, prior information about the vector of location parameters,  $\mathbf{x}$ , can be introduced by means of pseudo-observations, which are sometimes called the weighted constraints [Wells, 1990]. The resulting model, called a parametric model with weighted constraints, is then of the following form:

$$\begin{aligned} \mathbf{l} &= \mathbf{A}\mathbf{x} + \mathbf{v} \quad \text{with} \quad \mathbf{P}_1 = \mathbf{C}_1^{-1} \\ \mathbf{x}_0 &= \mathbf{I}\mathbf{x} + \delta \quad \text{with} \quad \mathbf{P}_{\mathbf{x}_0} = \mathbf{C}_{\mathbf{x}_0}^{-1}, \end{aligned} \quad (3.138)$$

where  $\delta$  is a sub-vector of errors corresponding to the vector of pseudo-observations  $\mathbf{x}_0$ .

Application of the method of least squares to the parametric model with weighted constraints, eqn. (3.138), yields the following normal equations (see, e.g., [Merminod and Rizos, 1989]):

$$\left[ \mathbf{A}^T \mathbf{C}_l^{-1} \mathbf{A} + \mathbf{C}_{x_0}^{-1} \right] \mathbf{x} = \mathbf{A}^T \mathbf{C}_l^{-1} \mathbf{l} + \mathbf{C}_{x_0}^{-1} \mathbf{x}_0. \quad (3.139)$$

In the similar fashion, the existing prior information about the vector of variance-covariance components,  $\theta$ , can be introduced in the dispersion-mean model, eqn. (3.129), by means of pseudo-observations of the variance-covariance components or, in other words, weighted constraints on the vector  $\theta$ . The resulting dispersion-mean model with weighted constraints reads:

$$\begin{aligned} \mathbf{l} &= \mathbf{A}\theta + \mathbf{v} \quad \text{with } \mathbf{P}_l = \mathbf{C}_l^{-1} \\ \theta_0 &= \mathbf{I}\theta + \delta \quad \text{with } \mathbf{P}_{\theta_0} = \mathbf{C}_{\theta_0}^{-1}, \end{aligned} \quad (3.140)$$

where  $\mathbf{C}_{\theta_0}$  is the covariance matrix associated with the vector of pseudo-observations of variance-covariance components,  $\theta_0$ .

With

$$\mathbf{A}_* = \begin{bmatrix} \mathbf{A} \\ \mathbf{I} \end{bmatrix}, \quad \mathbf{v}_* = \begin{bmatrix} \mathbf{v} \\ \delta \end{bmatrix}, \quad \text{and} \quad \mathbf{l}_* = \begin{bmatrix} \mathbf{l} \\ \theta_0 \end{bmatrix}, \quad (3.141)$$

the dispersion-mean model with weighted constraints reads

$$\mathbf{l}_* = \mathbf{A}_* \theta + \mathbf{v}_* \quad \text{with} \quad \mathbf{P} = \begin{bmatrix} \mathbf{C}_l^{-1} & 0 \\ 0 & \mathbf{C}_{\theta_0}^{-1} \end{bmatrix}. \quad (3.142)$$

Application of the least squares methodology to the dispersion-mean model with weighted constraints, eqn. (3.142), leads to the following normal equations:

$$\mathbf{A}_*^T \mathbf{P} \mathbf{A}_* \theta = \mathbf{A}_*^T \mathbf{p} l_*, \quad (3.143)$$

where:

$$\mathbf{A}_*^T \mathbf{P} \mathbf{A}_* = \mathbf{A}^T \mathbf{C}_t^{-1} \mathbf{A} + \mathbf{C}_{\theta_0}^{-1}, \text{ and} \quad (3.144)$$

$$\mathbf{A}_*^T \mathbf{p} l_* = \mathbf{A}^T \mathbf{C}_t^{-1} \mathbf{l} + \mathbf{C}_{\theta_0}^{-1} \theta_0, \quad (3.145)$$

which, after some rearrangement, can be written as:

$$\left( \mathbf{J}_R + \mathbf{C}_{\theta_0}^{-1} \right) \theta = \mathbf{q}_R + \mathbf{C}_{\theta_0}^{-1} \theta_0. \quad (3.146)$$

Since the elements of both the information matrix,  $\mathbf{J}_R$ , and the vector of quadratic forms,  $\mathbf{q}_R$ , are functions of the vector of variance-covariance components,  $\theta$ , the normal equations, eqn. (3.146), have to be solved iteratively. The solution to the normal equations, in the  $m$ -th iterative step, reads:

$$\hat{\theta}_{(m+1)} = \left[ \mathbf{J}_R(\hat{\theta}_{(m)}) + \mathbf{C}_{\theta_0}^{-1} \right]^{-1} \left[ \mathbf{q}_R(\hat{\theta}_{(m)}) + \mathbf{C}_{\theta_0}^{-1} \theta_0 \right] = \left[ \mathbf{J}_D(\hat{\theta}_{(m)}) \right]^{-1} \mathbf{q}_D(\hat{\theta}_{(m)}). \quad (3.147)$$

### 3.3.3 Conditions for Equivalence with GML Estimation

As discussed in Section 2.3, application of the Bayes' Theorem, eqn. (2.65), to the problem of estimation of the vector of variance-covariance components,  $\theta$ , with the marginal likelihood function, eqn. (2.44), assigned to the data and some prior probability density function,  $p(\theta)$ , assigned to the vector  $\theta$ , results in the following posterior probability density function:

$$p(\theta|\mathbf{l}) \propto p(\theta)L_R, \text{ with} \quad (3.148)$$

$$L_R \propto |\mathbf{C}_\theta|^{-\frac{1}{2}} |\mathbf{A}^T \mathbf{C}_\theta^{-1} \mathbf{A}|^{-\frac{1}{2}} \exp\left(-\frac{1}{2} \mathbf{l}^T \mathbf{R} \mathbf{l}\right). \quad (3.149)$$

As discussed in Subsection 2.3.4, application of the Fisher scoring method to the natural logarithm of a posterior probability density function yields the GML estimate of the vector of variance-covariance components. The task now is to find such prior probability density function,  $p(\theta)$ , that will produce the iterative equations of the dispersion-mean model with weighted constraints, eqn. (3.147), when the Fisher scoring method is applied to the natural logarithm of the posterior probability density function,

$$l_D = \ln[p(\theta)L_R] = \ln[p(\theta)] + l_R = l_{D0} + l_R. \quad (3.150)$$

Application of the Fisher scoring method to the function  $l_D$ , eqn. (3.150), results in the following expression (cf. eqn. (3.24) and eqn. (3.121)):

$$\begin{aligned} \hat{\theta}_{(m+1)} &= \hat{\theta}_{(m)} + \left[ \mathbf{J}_D(\hat{\theta}_{(m)}) \right]^{-1} \mathbf{s}_D(\hat{\theta}_{(m)}) \\ &= \hat{\theta}_{(m)} + \left[ \mathbf{J}_{D0}(\hat{\theta}_{(m)}) + \mathbf{J}_R(\hat{\theta}_{(m)}) \right]^{-1} \left[ \mathbf{s}_{D0}(\hat{\theta}_{(m)}) + \mathbf{s}_R(\hat{\theta}_{(m)}) \right]. \end{aligned} \quad (3.151)$$

where the vector of scores,  $\mathbf{s}_D$ , is defined as:

$$\mathbf{s}_D = \frac{\partial l_D}{\partial \theta} = \frac{\partial l_{D0}}{\partial \theta} + \frac{\partial l_R}{\partial \theta} = \mathbf{s}_{D0} + \mathbf{s}_R, \quad (3.152)$$

with the vector of marginal likelihood scores (see eqn. (2.56) and eqn. (2.62)) given as:

$$\mathbf{s}_R = \left\{ c - \frac{1}{2} \text{tr}(\mathbf{R} \mathbf{T}_k) + \frac{1}{2} \mathbf{l}^T \mathbf{R} \mathbf{T}_k \mathbf{R} \mathbf{l} \right\}_{k=1}^r = -\mathbf{J}_R \theta + \mathbf{q}_R, \text{ and} \quad (3.153)$$

the information matrix,  $\mathbf{J}_D$ , defined as:

$$\mathbf{J}_D = \frac{\partial^2 l_D}{\partial \theta \partial \theta} = \frac{\partial^2 l_{D0}}{\partial \theta \partial \theta} + \frac{\partial^2 l_R}{\partial \theta \partial \theta} = \mathbf{J}_{D0} + \mathbf{J}_R. \quad (3.154)$$

Introducing the results of eqn. (3.153) into the scoring equations, eqn. (3.151), and adding

and subtracting the product  $\mathbf{J}_{D0}(\hat{\theta}_{(m)})\hat{\theta}_{(m)}$  from the vector of scores,  $\mathbf{s}_D$ , we get:

$$\begin{aligned} \hat{\theta}_{(m+1)} &= \hat{\theta}_{(m)} + \left[ \mathbf{J}_{D0}(\hat{\theta}_{(m)}) + \mathbf{J}_R(\hat{\theta}_{(m)}) \right]^{-1} \\ &\times \left[ \mathbf{s}_{D0}(\hat{\theta}_{(m)}) + \mathbf{J}_{D0}(\hat{\theta}_{(m)})\hat{\theta}_{(m)} - \mathbf{J}_{D0}(\hat{\theta}_{(m)})\hat{\theta}_{(m)} - \mathbf{J}_R(\hat{\theta}_{(m)})\hat{\theta}_{(m)} + \mathbf{q}_R(\hat{\theta}_{(m)}) \right] \quad (3.155) \\ &= \left[ \mathbf{J}_R(\hat{\theta}_{(m)}) + \mathbf{J}_{D0}(\hat{\theta}_{(m)}) \right]^{-1} \left[ \mathbf{q}_R(\hat{\theta}_{(m)}) + \mathbf{s}_{D0}(\hat{\theta}_{(m)}) + \mathbf{J}_{D0}(\hat{\theta}_{(m)})\hat{\theta}_{(m)} \right]. \end{aligned}$$

If the scoring equations, eqn. (3.155), were to be equivalent to the normal equations of the dispersion-mean model with weighted constraints, eqn. (3.147), the following conditions would have to be fulfilled:

$$\mathbf{J}_{D0}(\hat{\theta}_{(m)}) = \mathbf{C}_{\theta_0}^{-1}, \text{ and} \quad (3.156)$$

$$\mathbf{s}_{D0}(\hat{\theta}_{(m)}) + \mathbf{J}_{D0}(\hat{\theta}_{(m)})\hat{\theta}_{(m)} = \mathbf{C}_{\theta_0}^{-1}\theta_0. \quad (3.157)$$

Substituting  $\mathbf{C}_{\theta_0}^{-1}$  for  $\mathbf{J}_{D0}(\hat{\theta}_{(m)})$  in eqn. (3.157) we get:

$$\mathbf{s}_{D0}(\hat{\theta}_{(m)}) + \mathbf{C}_{\theta_0}^{-1}\hat{\theta}_{(m)} = \mathbf{C}_{\theta_0}^{-1}\theta_0. \quad (3.158)$$

The relationship given in eqn. (3.158) will hold true for each  $\hat{\theta}_{(m)}$  if the following differential equation is fulfilled:

$$\frac{\partial l_{D0}}{\partial \theta} = \mathbf{C}_{\theta_0}^{-1}(\theta_0 - \theta) . \quad (3.159)$$

Solving this differential equation results in the following natural logarithm of the prior probability density function:

$$l_{D0} = \ln p(\theta) = -\frac{1}{2}\theta^T \mathbf{C}_{\theta_0}^{-1}\theta + \theta^T \mathbf{C}_{\theta_0}^{-1}\theta_0 + c , \quad (3.160)$$

where  $c$  is some constant. Since  $c$  may take any value, one may choose it to be such that the following relationship holds:

$$-\frac{1}{2}\theta^T \mathbf{C}_{\theta_0}^{-1}\theta + \theta^T \mathbf{C}_{\theta_0}^{-1}\theta_0 + c = -\frac{1}{2}(\theta - \theta_0)^T \mathbf{C}_{\theta_0}^{-1}(\theta - \theta_0) , \quad (3.161)$$

which results in

$$p(\theta) \propto \exp\left(-\frac{1}{2}\theta^T \mathbf{C}_{\theta_0}^{-1}\theta + \theta^T \mathbf{C}_{\theta_0}^{-1}\theta_0 + c\right) \propto \exp\left[-\frac{1}{2}(\theta - \theta_0)^T \mathbf{C}_{\theta_0}^{-1}(\theta - \theta_0)\right] , \quad (3.162)$$

which in turn implies that the prior distribution of the multivariate normal type:  $N(\theta_0, \mathbf{C}_{\theta_0})$ . It should be noted that, since the variance components are by definition positive, the multivariate normal distribution is not an appropriate choice for the prior for the variance components. Moreover, in the general mixed model, eqn (2.1), it is possible to find prior values of variance-covariance components that lead to negative definite covariance matrices, but nevertheless have positive prior probability associated with them [Knight, 1995].



## CHAPTER 4

### INFLUENCE OF ESTIMATION OF VARIANCE-COVARIANCE COMPONENTS ON THE COVARIANCE MATRIX OF THE ESTIMATED LOCATION PARAMETERS

The usual approach to least squares estimation of the general mixed model (extended Gauss-Markov model) of eqn. (2.1) is to neglect the uncertainty that is introduced by errors associated with the estimation of variance-covariance components (see [Searle et al., 1992]). As discussed in Chapter 1, the covariance matrix of the vector of observations is treated as fixed and errorless, even though it has been estimated with some error - due to the fact that the estimated variance-covariance components are not fixed quantities. The effect of this simplification is that the covariance matrix of the estimated location parameters is always underestimated, resulting in the confidence regions for the estimated location parameters that are too small. This may adversely affect decisions based upon statistical testing of hypotheses involving the least squares residuals and the estimated location parameters.

As reported by Searle et al. [1992], the covariance matrix of the estimated location parameters is always underestimated. Replacing the weight matrix  $\mathbf{P}_\theta$ , computed at the vector of true variance-covariance components, with the weight matrix  $\mathbf{P}_{\hat{\theta}}$ , computed at

the estimated vector of variance-covariance components, leads to an estimator  $\hat{\mathbf{x}}_{\hat{\theta}}$  that is consistent but has larger variance than  $\hat{\mathbf{x}}_{\theta}$ . To account for this increase of variance Searle et al. [1992] propose the use of the bootstrap method or the Taylor series expansion. The solution, however, is not given.

In this section the effect of estimation of variance-covariance components on the covariance matrix of the estimated location parameters will be discussed.

First, let us observe that the information matrix for the ML method of estimation, given by eqn. (2.25), is a block diagonal matrix composed of the information matrices for the vector of location parameters and variance-covariance components. In turn, the inverse of this information matrix is also a block diagonal matrix composed of the separate inverses of the two blocks, and hence has no cross elements between the location parameters and the variance-covariance components.

Then, for the experiment fulfilling the Hartley-Rao conditions [Hartley and Rao, 1967], as given in Subsection 2.2.1, the inverse of the information matrix given by eqn. (2.25) is the asymptotic covariance matrix of the vector of unknown parameters - including both the location parameters and the variance components. Thus, asymptotically (as the number of observations increases to infinity) estimating variance components has no effect on the covariance matrix of the estimated location parameters [Knight, 1997].

To assess the effect of using the estimated covariance matrix (computed using the estimated variance-covariance components) rather than the exact covariance matrix (computed using the true variance-covariance components) on the covariance matrix of the estimated location parameters, in the least squares adjustment, the following development is proposed by Knight [1997].

First, the vector of observations,  $\mathbf{l}$ , is transformed to make its covariance matrix identity. Then, the vector of location parameters,  $\mathbf{x}$ , is transformed to make  $\mathbf{A}^T\mathbf{A} = \mathbf{I}$  and translated to zero. The resulting model reads

$$\mathbf{l} = \mathbf{A}\mathbf{x} + \mathbf{v} \quad \text{with} \quad \mathbf{C}_1 = \mathbf{I}, \quad (4.1)$$

with the least squares estimate of the vector of location parameters is given as

$$\hat{\mathbf{x}} = \mathbf{A}^T\mathbf{l}, \quad (4.2)$$

and the vector of least squares residuals being

$$\hat{\mathbf{v}} = (\mathbf{I} - \mathbf{A}\mathbf{A}^T)\mathbf{l}. \quad (4.3)$$

Suppose next that least squares estimation is based on a wrong covariance matrix  $\mathbf{I} + \mathbf{W}$ , where  $\mathbf{W}$  is small. The resulting least squares estimate of the vector of location parameters reads:

$$\hat{\mathbf{x}}_{\mathbf{W}} = \left[ \mathbf{A}^T(\mathbf{I} + \mathbf{W})^{-1}\mathbf{A} \right]^{-1} \mathbf{A}^T(\mathbf{I} + \mathbf{W})^{-1}\mathbf{l}. \quad (4.4)$$

Applying repeatedly

$$(\mathbf{I} - \mathbf{Z})^{-1} = \sum_{k=0}^{\infty} (-\mathbf{Z})^k \quad (4.5)$$

and neglecting the terms of order two or higher in  $\mathbf{W}$  we get

$$\begin{aligned} \hat{\mathbf{x}}_{\mathbf{W}} &= [\mathbf{A}^T(\mathbf{I} - \mathbf{W})\mathbf{A}]^{-1} \mathbf{A}^T(\mathbf{I} - \mathbf{W})\mathbf{l} = (\mathbf{I} - \mathbf{A}^T\mathbf{W}\mathbf{A})^{-1}(\mathbf{A}^T - \mathbf{A}^T\mathbf{W})\mathbf{l} \\ &= (\mathbf{I} + \mathbf{A}^T\mathbf{W}\mathbf{A})(\mathbf{A}^T - \mathbf{A}^T\mathbf{W})\mathbf{l} = [\mathbf{A}^T - \mathbf{A}^T\mathbf{W}(\mathbf{I} - \mathbf{A}\mathbf{A}^T)]\mathbf{l} \\ &= \hat{\mathbf{x}} - \mathbf{A}^T\mathbf{W}\hat{\mathbf{v}}. \end{aligned} \quad (4.6)$$

The covariance matrix of  $\hat{\mathbf{x}}_{\mathbf{W}}$  follows with

$$\text{var}(\hat{\mathbf{x}}_{\mathbf{W}}) = \text{E}(\hat{\mathbf{x}}_{\mathbf{W}}\hat{\mathbf{x}}_{\mathbf{W}}^T) = \text{E}(\hat{\mathbf{x}}\hat{\mathbf{x}}^T - \hat{\mathbf{x}}\hat{\mathbf{v}}^T\mathbf{W}^T\mathbf{A} - \mathbf{A}^T\mathbf{W}\hat{\mathbf{v}}\hat{\mathbf{x}}^T + \mathbf{A}^T\mathbf{W}\hat{\mathbf{v}}\hat{\mathbf{v}}^T\mathbf{W}^T\mathbf{A}). \quad (4.7)$$

Neglecting the fourth term (being quadratic in  $\mathbf{W}$ ) and noting that for fixed  $\mathbf{W}$  expectations of the second and the third term are zero (because  $\hat{\mathbf{v}}$  and  $\hat{\mathbf{x}}$  are uncorrelated) we finally get:

$$\text{var}(\hat{\mathbf{x}}_{\mathbf{W}}) = \text{E}(\hat{\mathbf{x}}\hat{\mathbf{x}}^T) = \text{var}(\hat{\mathbf{x}}). \quad (4.8)$$

This proves that using the estimated covariance matrix of observations, in place of the true covariance matrix, has no effect on the covariance matrix of the least-squares estimated vector of location parameters - up to the first order terms.

## CHAPTER 5

### **ANALYSIS OF HOMOGENEITY OF THE DATA: DETECTION OF INFLUENTIAL OBSERVATIONS**

As far as the estimation of variance-covariance components is concerned, analysis of geodetic data is generally handicapped by small redundancy. As a result, undetected outliers will have a disproportionately high influence on the estimates, resulting in distorted values of the estimated variance-covariance components. In addition, because of the aforementioned problem with small redundancy, some observations (not necessarily being the outliers of the least squares adjustment) may have excessively high influence on the estimated components due to, e.g., misspecification of the error model or nonhomogeneity of the data. In the extreme cases, this problem may manifest itself in negative estimates of variances or negative definiteness of the resulting covariance matrix of observations.

The existing approach to minimizing the influence of outlying observations on the process of estimation of variance-covariance components for geodetic data [Caspary, 1987], [Chrzanowski et al., 1994] is to detect and remove potential outliers during the least squares adjustment process, performed prior to the estimation of variance-covariance components and using the empirical weights of observations. Since the covariance matrix of observations may be in error - as it is only approximate, possibly based on manufacturer's specifications and pre-analysis of sources of errors - some of the outliers

may not be detected by the statistical testing procedures. Once the outlying observations have been removed, the next step is to estimate the unknown variance-covariance components. The third step in the detection of outliers is to readjust the network after estimating the variance-covariance components, using the estimated components to compute the covariance matrix of observations. The data is screened again for outliers, and if new ones are found, then the outlying observations are removed and the estimation of variance-covariance components is repeated. This iterative procedure is continued until all outliers are removed from the data. The drawback of this method is that the outliers not detected in the first step will influence the estimation of variance-covariance components resulting in unrealistic values and, hence, overestimated covariance matrix of the observations. This in effect will make detection of additional outliers in the subsequent steps increasingly difficult, since the confidence regions used, for example, in the t-test of studentized residuals will be unrealistically large as well.

Since, as mentioned above, the number of degrees of freedom in geodetic networks is usually comparatively small, one or otherwise a small number of observations may have a disproportionately high influence on one or more of the estimated variance-covariance components and their variances, in some cases almost uniquely determining their value. In other words, removal of such influential observation would drastically change the results of estimation. Such influential observations must be thoroughly analyzed to determine whether they belong to the same error model as other observations in the given group.

The presence of influential observations may also signal the presence of undetected outliers in the data.

This chapter addresses the analysis of data in terms of the presence of highly influential observations and undetected outliers, and their influence on the results of the estimation of variance-covariance components.

### 5.1 Influence Function

The influence of a single observation on the vector of the estimated variance-covariance components may be described by the influence function. The influence function expresses the effect that adding one observation with value  $y$  has on a functional (estimate)  $\hat{\theta}$  at a distribution  $F$ , and is defined by Hampel et al. [1986] as:

$$\mathbf{IF}[y; \hat{\theta}(F)] = \lim_{s \rightarrow 0} \frac{\hat{\theta}((1-s)F + s\delta_y) - \hat{\theta}(F)}{s}, \quad (5.1)$$

where  $\delta_y$  denotes the pointmass 1 at  $y$ .

The influence function, eqn. (5.1), is defined in terms of asymptotic values of the estimator. There exist, however, the finite-sample versions of the influence function. These approximations can be found in, e.g., [Hampel et al., 1986] and [Chatterjee and Hadi, 1988] and include among others:

- (1) the empirical influence function, which for an estimator  $\{\hat{\theta}_m; m \geq 1\}$  and a sample  $(l_1, l_2, \dots, l_{m-1})$  is defined as

$$\mathbf{EIF}_m(y) = \hat{\boldsymbol{\theta}}_m(l_1, l_2, \dots, l_{m-1}, y), \quad (5.2)$$

(2) the sensitivity function, which for an estimator  $\{\hat{\boldsymbol{\theta}}_m; m \geq 1\}$  and a sample  $(l_1, l_2, \dots, l_{m-1})$  is defined as:

$$\mathbf{SF}_m(y) = m[\hat{\boldsymbol{\theta}}_m(l_1, l_2, \dots, l_{m-1}, y) - \hat{\boldsymbol{\theta}}_{m-1}(l_1, l_2, \dots, l_{m-1})], \text{ and} \quad (5.3)$$

(3) the sample influence function, which for an estimator  $\{\hat{\boldsymbol{\theta}}_m; m \geq 1\}$  and a sample  $(l_1, l_2, \dots, l_{m-1})$  is defined as:

$$\mathbf{SIF}_m(y) = (m-1)[\hat{\boldsymbol{\theta}}_m(l_1, l_2, \dots, l_{m-1}, y) - \hat{\boldsymbol{\theta}}_{m-1}(l_1, l_2, \dots, l_{m-1})]. \quad (5.4)$$

Maximum influence of outliers on the estimate  $\hat{\boldsymbol{\theta}}$  can be quantified by the unstandardized gross-error sensitivity of the estimate  $\hat{\boldsymbol{\theta}}$ , which is given by Hampel et al. [1986] as:

$$\gamma_u^* = \sup_y \|\mathbf{IF}[y; \hat{\boldsymbol{\theta}}(F)]\|, \quad (5.5)$$

where  $\|\cdot\|$  denotes the Euclidean norm. The unstandardized gross-error sensitivity, eqn. (5.5), is not invariant to changes of scale of the parameters. To overcome that problem, Hampel et al. [1986] propose two alternative measures of maximum sensitivity of the estimate  $\hat{\boldsymbol{\theta}}$  to outliers, namely:

(1) the self-standardized sensitivity which is defined as

$$\gamma_s^* = \sup_y \left\{ \mathbf{IF}[y; \hat{\boldsymbol{\theta}}(F)]^T \left\{ \text{var}[\hat{\boldsymbol{\theta}}(F)] \right\}^{-1} \mathbf{IF}[y; \hat{\boldsymbol{\theta}}(F)] \right\}^{\frac{1}{2}}, \quad (5.6)$$

(if  $\text{var}[\hat{\boldsymbol{\theta}}(F)]$  is nonsingular, else by  $\infty$ ) where  $\text{var}[\hat{\boldsymbol{\theta}}(F)]$  is the asymptotic covariance matrix of the estimator  $\hat{\boldsymbol{\theta}}$ , and



(2) the information-standardized sensitivity which is defined as

$$\gamma_i^* = \sup_x \left\{ \mathbf{IF}[x; \hat{\theta}(F)]^T \{ \mathbf{J}[\hat{\theta}(F)] \}^{-1} \mathbf{IF}[x; \hat{\theta}(F)] \right\}^{\frac{1}{2}}, \quad (5.7)$$

if  $\mathbf{J}[\hat{\theta}(F)]$  exists, where  $\mathbf{J}[\hat{\theta}(F)]$  is the information matrix of the estimator  $\hat{\theta}$ .

In the case of the maximum likelihood type estimation, the influence function is shown by Hampel et al. [1986] to be

$$\mathbf{IF}[y; \hat{\theta}(F_*)] = \{ \mathbf{J}[\hat{\theta}(F_*)] \}^{-1} s[y; \hat{\theta}(F_*)], \quad (5.8)$$

where:  $s[y; \hat{\theta}(F_*)]$  is the scores function,

$\hat{\theta}$  is the maximum likelihood type estimator,

$\theta_*$  is a fixed parameter value, and

$F_*$  is the corresponding distribution,  $F_* = F(\theta_*)$ .

Taking into account the formulae for the vector of scores (see eqn. (2.23), eqn. (2.56), eqn. (3.11), and eqn. (3.92)) for all maximum likelihood type estimators considered in this thesis (ML, REML, and GML) the influence function is unbounded in terms of the vector of observations. For the REML and GML estimators this is due to the influence of observations on the vector of quadratic forms of observations

$$\mathbf{q}_R(\hat{\theta}) = \left\{ c \frac{1}{2} \mathbf{l}^T \mathbf{R}_{\hat{\theta}} \mathbf{T}_1 \mathbf{R}_{\hat{\theta}} \mathbf{l} \right\}_{i=1}^r, \quad (5.9)$$

and for the ML estimator through

$$s_{M_{\hat{\theta}}}(\hat{\mathbf{x}}, \hat{\boldsymbol{\theta}}) = \left\{ c - \frac{1}{2} \text{tr}(\mathbf{C}_{\hat{\boldsymbol{\theta}}}^{-1} \mathbf{T}_k) + \frac{1}{2} (\mathbf{I} - \mathbf{A}\hat{\mathbf{x}})^T \mathbf{C}_{\hat{\boldsymbol{\theta}}}^{-1} \mathbf{T}_k \mathbf{C}_{\hat{\boldsymbol{\theta}}}^{-1} (\mathbf{I} - \mathbf{A}\hat{\mathbf{x}}) \right\}_{k=1}^r. \quad (5.10)$$

As a result, an arbitrarily large outlier can cause an arbitrarily large change in the vector of estimated variance or variance-covariance components.

Approximating the influence function by the finite-sample sensitivity function:

$$\mathbf{SF}_i(\mathbf{l}_i) = m(\hat{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}^{(i)}), \quad (5.11)$$

where  $\hat{\boldsymbol{\theta}}^{(i)}$  is the vector of estimated variance or variance-covariance components when the  $i$ -th observation is removed, we get the following expression for the self-standardized gross-error sensitivity function:

$$\gamma_s(\mathbf{l}_i) = m \left\{ (\hat{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}^{(i)})^T [\text{var}(\hat{\boldsymbol{\theta}})]^{-1} (\hat{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}^{(i)}) \right\}^{\frac{1}{2}}. \quad (5.12)$$

## 5.2 Computational Aspects of Detection of Influential Observations

As far as ML, REML, and GML estimation of variance or variance-covariance components is concerned, one of the most computationally expensive tasks is the computation of the matrix  $\mathbf{R}_{\hat{\boldsymbol{\theta}}}$ :

$$\mathbf{R}_{\hat{\boldsymbol{\theta}}} = \mathbf{C}_{\hat{\boldsymbol{\theta}}}^{-1} \left[ \mathbf{I} - \mathbf{A} (\mathbf{A}^T \mathbf{C}_{\hat{\boldsymbol{\theta}}}^{-1} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{C}_{\hat{\boldsymbol{\theta}}}^{-1} \right]. \quad (5.13)$$

In the case of ML estimation of variance-covariance components, the matrix  $\mathbf{R}_{\hat{\boldsymbol{\theta}}}$  takes part in computation of the vector of quadratic forms of observations:

$$\mathbf{q}_R = \left\{ \frac{1}{c} \mathbf{l}^T \mathbf{R}_{\hat{\theta}} \mathbf{T}_i \mathbf{R}_{\hat{\theta}} \mathbf{l} \right\}_{i=1}^r, \quad (5.14)$$

in the iterative procedure given by eqn. (2.28). When the REML estimation method is used, the matrix  $\mathbf{R}_{\hat{\theta}}$  takes part both in computation of the vector of REML scores, through the vector of quadratic forms of observations,  $\mathbf{q}_R$ , and in computation of the information matrix  $\mathbf{J}_R$ :

$$\mathbf{J}_R = \left\{ \frac{1}{m} \text{tr}(\mathbf{R}_{\hat{\theta}} \mathbf{T}_i \mathbf{R}_{\hat{\theta}} \mathbf{T}_j) \right\}_{i,j=1}^r. \quad (5.15)$$

For both GML type estimators considered in this thesis, i.e., GMLE with the inverted gamma prior and GMLE with the noninformative prior, the matrix  $\mathbf{R}_{\hat{\theta}}$  takes part both in computation of vectors of GML scores (see eqn. (3.11) and eqn. (3.92)), and in computation of the GML information matrices (see eqn. (3.23) and eqn. (3.120)). As far as the dispersion-mean model with weighted constraints is concerned, the matrix  $\mathbf{R}_{\hat{\theta}}$  takes part both in computation of the vector of quadratic forms,  $\mathbf{q}_R$ , and the REML information matrix  $\mathbf{J}_R$  - in the iterative procedure given by eqn. (3.147). As a result, an efficient algorithm allowing quick and effective computation of matrix  $\mathbf{R}_{(i)}$  (where  $\mathbf{R}_{(i)}$  denotes the  $\mathbf{R}_{\hat{\theta}}$  matrix computed for a sample with  $i$ -th observation removed) from the elements of the matrix  $\mathbf{R}_{\hat{\theta}}$  would considerably decrease the computational burden involved in the detection of influential observations. In other words, it would make possible the computation of  $\hat{\theta}^{(i)}$ , for each observation  $l_i$ , without the need to perform the estimation process each time an  $i$ -th observation is removed. Such an algorithm, however,

would allow us to compute only one step estimates of the vector of variance components, thus making the interpretation process somewhat more difficult.

An algorithm for expressing the elements of the matrix  $\mathbf{R}_{(i)}$  in terms of the elements of the matrix  $\mathbf{R}_{\hat{\theta}}$  (for the simpler case involving only the variances, i.e., with the covariance matrix  $\mathbf{C}_{\hat{\theta}}$  being a diagonal matrix) is derived in the following subsections.

### *5.2.1 The Effect of Omitting an Observation on the Inverse of the Matrix of Normal Equations*

To simplify all subsequent derivations, without any loss of generality, the  $i$ -th observation is assumed to be the last observation in the vector of observations,  $\mathbf{l}$ .

For the diagonal weight matrix the following relationship holds:

$$\mathbf{A}^T \mathbf{P} \mathbf{A} = \mathbf{A}_{(i)}^T \mathbf{P}_{(i)} \mathbf{A}_{(i)} + \mathbf{a}_i p_{ii} \mathbf{a}_i^T, \quad (5.16)$$

where:  $\mathbf{A}_{(i)}$  denotes the  $\mathbf{A}$  matrix with the  $i$ -th row omitted,

$\mathbf{P}_{(i)}$  denotes the  $\mathbf{P}$  matrix with the  $i$ -th row and  $i$ -th column omitted,

$\mathbf{a}_i$  denotes the  $i$ -th column of the  $\mathbf{A}^T$  matrix, and

$p_{ii}$  denotes the  $i$ -th diagonal element of the  $\mathbf{P}$  matrix.

Applying the Sherman-Morrison-Woodbury theorem [Chatterjee and Hadi, 1988],

$$(\mathbf{U} + \mathbf{V}\mathbf{X}\mathbf{W}^T)^{-1} = \mathbf{U}^{-1} - \mathbf{U}^{-1}\mathbf{V}(\mathbf{X}^{-1} + \mathbf{W}^T\mathbf{U}^{-1}\mathbf{V})^{-1}\mathbf{W}^T\mathbf{U}^{-1}, \quad (5.17)$$

to the expression for the matrix of normal equations, eqn. (5.16), and substituting

$$\mathbf{U} = \mathbf{A}^T\mathbf{P}\mathbf{A}, \quad (5.18)$$

$$\mathbf{V} = -\mathbf{a}_i, \quad (5.19)$$

$$\mathbf{W}^T = \mathbf{a}_i^T, \text{ and} \quad (5.20)$$

$$\mathbf{X} = p_{ii}, \quad (5.21)$$

one gets the inverse of the matrix of normal equations, with the  $i$ -th observation removed from the sample, as:

$$\left(\mathbf{A}_{(i)}^T\mathbf{P}_{(i)}\mathbf{A}_{(i)}\right)^{-1} = \left(\mathbf{A}^T\mathbf{P}\mathbf{A} - \mathbf{a}_i p_{ii} \mathbf{a}_i^T\right)^{-1} = \left(\mathbf{A}^T\mathbf{P}\mathbf{A}\right)^{-1} + \frac{\left(\mathbf{A}^T\mathbf{P}\mathbf{A}\right)^{-1} \mathbf{a}_i \mathbf{a}_i^T \left(\mathbf{A}^T\mathbf{P}\mathbf{A}\right)^{-1}}{1/p_{ii} - \mathbf{a}_i^T \left(\mathbf{A}^T\mathbf{P}\mathbf{A}\right)^{-1} \mathbf{a}_i}. \quad (5.22)$$

### 5.2.2 The Effect of Omitting an Observation on the $\mathbf{R}_{\hat{\theta}}$ Matrix

The problem at hand is to find a functional relationship between the elements of the  $\mathbf{R}_{(i)}$  matrix and the elements of the  $\mathbf{R}_{\hat{\theta}}$  matrix.

The  $\mathbf{R}_{\hat{\theta}}$  matrix can be expressed as the difference between the weight matrix  $\mathbf{P}$  and the matrix  $\mathbf{M}$ :

$$\mathbf{R}_{\hat{\theta}} = \mathbf{P} \left[ \mathbf{I} - \mathbf{A} \left(\mathbf{A}^T\mathbf{P}\mathbf{A}\right)^{-1} \mathbf{A}^T \mathbf{P} \right] = \mathbf{P} - \mathbf{P}\mathbf{A} \left(\mathbf{A}^T\mathbf{P}\mathbf{A}\right)^{-1} \mathbf{A}^T \mathbf{P} = \mathbf{P} - \mathbf{P}\mathbf{G}\mathbf{P} = \mathbf{P} - \mathbf{M}. \quad (5.23)$$

The  $\mathbf{R}_{(i)}$  matrix can in turn be expressed as:

$$\begin{aligned}
\mathbf{R}_{(i)} &= \mathbf{P}_{(i)} \left[ \mathbf{I} - \mathbf{A}_{(i)} \left( \mathbf{A}_{(i)}^T \mathbf{P}_{(i)} \mathbf{A}_{(i)} \right)^{-1} \mathbf{A}_{(i)}^T \mathbf{P}_{(i)} \right] \\
&= \mathbf{P}_{(i)} - \mathbf{P}_{(i)} \mathbf{A}_{(i)} \left( \mathbf{A}_{(i)}^T \mathbf{P}_{(i)} \mathbf{A}_{(i)} \right)^{-1} \mathbf{A}_{(i)}^T \mathbf{P}_{(i)} \\
&= \mathbf{P}_{(i)} - \mathbf{P}_{(i)} \mathbf{G}_{(i)} \mathbf{P}_{(i)} = \mathbf{P}_{(i)} - \mathbf{M}_{(i)}.
\end{aligned} \tag{5.24}$$

To find a functional relationship between the elements of  $\mathbf{R}_{(i)}$  and  $\mathbf{R}_{\hat{\theta}}$ , such relationship will be first derived for the elements of matrices  $\mathbf{G}_{(i)}$  and  $\mathbf{G}$ , and then for the elements of matrices  $\mathbf{M}_{(i)}$  and  $\mathbf{M}$ .

The relationship between the elements of the  $\mathbf{G}$  and  $\mathbf{G}_{(i)}$  matrices is derived as follows.

The matrices  $\mathbf{G}$  and  $\mathbf{G}_{(i)}$  are given as:

$$\mathbf{G} = \mathbf{A} \left( \mathbf{A}^T \mathbf{P} \mathbf{A} \right)^{-1} \mathbf{A}^T, \text{ and} \tag{5.25}$$

$$\mathbf{G}_{(i)} = \mathbf{A}_{(i)} \left( \mathbf{A}_{(i)}^T \mathbf{P}_{(i)} \mathbf{A}_{(i)} \right)^{-1} \mathbf{A}_{(i)}^T. \tag{5.26}$$

The (r,c)-th element of the matrix  $\mathbf{G}$  can be expressed as:

$$g_{rc} = \mathbf{a}_c^T \left( \mathbf{A}^T \mathbf{P} \mathbf{A} \right)^{-1} \mathbf{a}_r. \tag{5.27}$$

The (r,c)-th element of the matrix  $\mathbf{G}_{(i)}$  is then:

$$\begin{aligned}
g_{rc(i)} &= \mathbf{a}_c^T (\mathbf{A}_{(i)}^T \mathbf{P}_{(i)} \mathbf{A}_{(i)})^{-1} \mathbf{a}_r \\
&= \mathbf{a}_c^T \left[ (\mathbf{A}^T \mathbf{P} \mathbf{A})^{-1} + \frac{(\mathbf{A}^T \mathbf{P} \mathbf{A})^{-1} \mathbf{a}_i \mathbf{a}_i^T (\mathbf{A}^T \mathbf{P} \mathbf{A})^{-1}}{\frac{1}{p_{ii}} - \mathbf{a}_i^T (\mathbf{A}^T \mathbf{P} \mathbf{A})^{-1} \mathbf{a}_i} \right] \mathbf{a}_r \\
&= \mathbf{a}_c^T (\mathbf{A}^T \mathbf{P} \mathbf{A})^{-1} \mathbf{a}_r + \frac{\left[ \mathbf{a}_c^T (\mathbf{A}^T \mathbf{P} \mathbf{A})^{-1} \mathbf{a}_i \right] \left[ \mathbf{a}_i^T (\mathbf{A}^T \mathbf{P} \mathbf{A})^{-1} \mathbf{a}_r \right]}{\frac{1}{p_{ii}} - \mathbf{a}_i^T (\mathbf{A}^T \mathbf{P} \mathbf{A})^{-1} \mathbf{a}_i} \\
&= g_{rc} + \frac{g_{ic} g_{ri}}{\frac{1}{p_{ii}} - g_{ii}}.
\end{aligned} \tag{5.28}$$

Having found the functional relationship between the elements of  $\mathbf{G}$  and  $\mathbf{G}_{(i)}$ , eqn. (5.28), one can derive a functional relationship between the elements of the matrices  $\mathbf{M}$  and  $\mathbf{M}_{(i)}$ .

The matrices  $\mathbf{M}$  and  $\mathbf{M}_{(i)}$  are given as:

$$\mathbf{M} = \mathbf{P} \mathbf{G} \mathbf{P}, \text{ and} \tag{5.29}$$

$$\mathbf{M}_{(i)} = \mathbf{P}_{(i)} \mathbf{G}_{(i)} \mathbf{P}_{(i)}. \tag{5.30}$$

The (r,c)-th element of the matrix  $\mathbf{M}$  is thus expressed as:

$$m_{rc} = p_{rr} g_{rc} p_{cc}, \text{ and} \tag{5.31}$$

the (r,c)-th element of the matrix  $\mathbf{M}_{(i)}$  as:

$$\begin{aligned}
m_{rc(i)} &= p_{rr} g_{rc(i)} p_{cc} \\
&= p_{rr} \left[ g_{rc} + \frac{g_{ic} g_{ri}}{1/p_{ii} - g_{ii}} \right] p_{cc} \\
&= m_{rc} + p_{rr} \frac{g_{ic} g_{ri}}{1/p_{ii} - g_{ii}} p_{cc}.
\end{aligned} \tag{5.32}$$

Applying the results of eqn. (5.32) to the expression for the  $\mathbf{R}_{(i)}$  matrix, eqn. (5.24), a functional relationship between the elements of the  $\mathbf{R}_{\hat{\theta}}$  and  $\mathbf{R}_{(i)}$  matrices is found. As seen in eqn. (5.23) and eqn. (5.24) the matrices  $\mathbf{R}_{\hat{\theta}}$  and  $\mathbf{R}_{(i)}$  can be expressed as:

$$\mathbf{R}_{\hat{\theta}} = \mathbf{P} - \mathbf{M}, \text{ and} \tag{5.33}$$

$$\mathbf{R}_{(i)} = \mathbf{P}_{(i)} - \mathbf{M}_{(i)}. \tag{5.34}$$

The (r,c)-th element and the (c,c)-th element of the matrix  $\mathbf{R}_{\hat{\theta}}$  can therefore be expressed as:

$$r_{rc} = -p_{rr} g_{rc} p_{cc}, \text{ and} \tag{5.35}$$

$$r_{cc} = p_{cc} - p_{cc} g_{cc} p_{cc}. \tag{5.36}$$

The (r,c)-th element and the (c,c)-th element of the matrix  $\mathbf{R}_{(i)}$  are then:



$$\begin{aligned}
r_{rc(i)} &= -p_{rr}g_{rc(i)}p_{cc} \\
&= -p_{rr}\left[g_{rc} + \frac{g_{ic}g_{ri}}{1/p_{ii} - g_{ii}}\right]p_{cc} \\
&= r_{rc} - p_{rr}\frac{g_{ic}g_{ri}}{1/p_{ii} - g_{ii}}p_{cc} \\
&= r_{rc} - \frac{(p_{cc}g_{ic}p_{ii})(p_{ii}g_{ri}p_{rr})}{p_{ii} - p_{ii}g_{ii}p_{ii}} \\
&= r_{rc} - \frac{r_{ci}r_{ir}}{r_{ii}} \\
&= r_{rc} - \Delta r_{rc}, \text{ and}
\end{aligned} \tag{5.37}$$

$$\begin{aligned}
r_{cc(i)} &= p_{cc} - p_{cc}g_{cc(i)}p_{cc} \\
&= p_{cc} - p_{cc}\left[g_{cc} + \frac{(g_{ic})^2}{1/p_{ii} - g_{ii}}\right]p_{cc} \\
&= r_{cc} - p_{cc}\frac{(g_{ic})^2}{1/p_{ii} - g_{ii}}p_{cc} \\
&= r_{cc} - \frac{(p_{cc}g_{ic}p_{ii})(p_{cc}g_{ri}p_{ii})}{p_{ii} - p_{ii}g_{ii}p_{ii}} \\
&= r_{cc} - \frac{(r_{ci})^2}{r_{ii}} \\
&= r_{cc} - \Delta r_{cc}.
\end{aligned} \tag{5.38}$$

Finally, the expression for the matrix  $\mathbf{R}_{(i)}$  is found as:

$$\mathbf{R}_{(i)} = \left\{ {}_m r_{rc(i)} \right\}_{r,c=1}^{m-1} = \left\{ {}_m r_{rc} - \frac{r_{ci}r_{ir}}{r_{ii}} \right\}_{r,c=1}^{m-1}, \tag{5.39}$$

where,

$$\mathbf{R}_{\hat{\theta}} = \left\{ \sum_{r,c=1}^m r_{rc} \right\} . \quad (5.40)$$

### 5.3 Influence of a Single Observation on the Asymptotic Covariance Matrix of the Vector of the Estimated Variance-Covariance Components

The influence of the  $i$ -th observation on the asymptotic covariance matrix of the estimated variance or variance-covariance components can be measured by comparing values of some scalar function of the covariance matrices  $\text{var}(\hat{\theta}_{(i)})$  and  $\text{var}(\hat{\theta})$ . If the ratio of determinants of these covariance matrices is used, then the resulting measure is similar to COVRATIO (see [Belsley et al., 1980]) or  $\text{VR}_i$  (see [Chatterjee and Hadi, 1988]) which measures the influence of an observation on the covariance matrix of the vector of estimated location parameters in the least squares regression. If the ratio of determinants of the asymptotic covariance matrices is used, then such measure (called here VR) is given by:

$$\text{VR}(l_i) = \frac{\det[\text{var}(\hat{\theta}_{(i)})]}{\det[\text{var}(\hat{\theta})]} . \quad (5.41)$$

Other measures of the influence of an  $i$ -th observation on the asymptotic covariance matrix of the estimated variance or variance-covariance components include the ratio of the determinants of the information matrices (resulting in a measure similar to the Andrews-Pregibon statistic [Andrews and Pregibon, 1978]), and a logarithm of a ratio of volumes of

the  $r$ -dimensional asymptotic error ellipsoids (resulting in a measure similar to the Cook-Weisberg statistic [Cook and Weisberg, 1980]). Other possible measures include a ratio of maximum eigenvalues and the ratio of traces of the asymptotic covariance matrices.

## **CHAPTER 6**

### **INFLUENCE OF ESTIMATION OF VARIANCE-COVARIANCE COMPONENTS ON SPATIAL DEFORMATION TREND ANALYSIS**

In this chapter the concept of spatial deformation trend analysis will be reviewed, with the special emphasis put on the Iterative Weighted Similarity Transformation procedure. Subsequently, the effect that the estimation of variance-covariance components has on the spatial deformation trend analysis process, and, in particular on the Iterative Weighted Similarity Transformation, will be discussed.

#### **6.1 Principle of Spatial Deformation Trend Analysis**

There are a number of alternative procedures for detection of unstable reference and object points (see [Chrzanowski, 1981]). Most of them rely on statistical testing of some datum invariant quantities, in order to confirm stability of the network points. This chapter will concentrate on the Iterative Weighted Similarity Transformation process [Chen, 1983], which is a part of the UNB Generalized Method of Deformation Analysis [Chrzanowski et al., 1986].

When deformation measurements are performed by geodetic methods, two main types of geodetic monitoring networks are distinguished [Chrzanowski et al., 1986]: absolute

networks and relative networks. The absolute networks are comprised of (1) points established outside the deformable body constituting a reference network which in turn is used to determine absolute displacements of the object points, and of (2) object points. The relative networks, are such networks in which all the surveyed stations and observed points are located on or within the deformable object.

In the case of the absolute networks the main objective of deformation trend analysis is to identify the stable reference points. Once the stable reference points have been identified, they may subsequently serve as the computational base for determination of the true displacement field of the object points.

As far as the relative networks are concerned, the main goal of deformation trend analysis is determination of the deformation pattern in space or both in space and time domains. The trend analysis for the relative networks is more complex than for the absolute networks, since in addition to the possible single point movements, all the points may experience relative movements caused by deformation of the object itself.

The Iterative Weighted Similarity Transformation may be used both for the analysis of stability of the reference points and for the trend analysis of the relative networks.

The concept of spatial deformation trend analysis and, in particular, the Iterative Weighted Similarity Transformation procedure, is inherently connected with the theory of free

networks and the problem of datum defects. A free network is a network that does not contain enough information to be located in space. It suffers from datum defects, i.e., it can be freely translated, rotated or scaled in space. In the case of deformation surveys by geodetic methods one has to assume that the monitoring network is a free network, since no point can be assumed to be stable prior to the analysis [Chen et al., 1990b]. If a monitoring network is a free network [Chen, 1983] then the matrix of normal equations is singular, and, hence, there is an infinite number of solutions for the vector of location parameters,  $\mathbf{x}$ . In other words, if  $\mathbf{x}_1$  is a solution then  $\mathbf{x}_2 = \mathbf{x}_1 + \mathbf{H} \mathbf{t}$  is also a solution, with  $\mathbf{H}$  being a matrix of coefficients defined by the equations of the similarity (Helmert) transformation, and  $\mathbf{t}$  being an arbitrary vector. A Helmert transformation matrix for a three dimensional geodetic network with undefined location, orientation and scale, coordinated in a left handed coordinate system, is given by [Casparly, 1987] as:

$$\mathbf{H}^T = \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 0 & \dots & \dots & \dots & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & \dots & \dots & \dots & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & \dots & \dots & \dots & 0 & 0 & 1 \\ 0 & z_1^0 & -y_1^0 & 0 & z_2^0 & -y_2^0 & \dots & \dots & \dots & 0 & z_m^0 & -y_m^0 \\ -z_1^0 & 0 & x_1^0 & -z_2^0 & 0 & x_2^0 & \dots & \dots & \dots & -z_m^0 & 0 & x_m^0 \\ y_1^0 & -x_1^0 & 0 & y_2^0 & -x_2^0 & 0 & \dots & \dots & \dots & y_m^0 & -x_m^0 & 0 \\ x_1^0 & y_1^0 & z_1^0 & x_2^0 & y_2^0 & z_2^0 & \dots & \dots & \dots & x_m^0 & y_m^0 & z_m^0 \end{bmatrix}, \quad (6.1)$$

where the first three rows of the matrix  $\mathbf{H}^T$  correspond to translations along  $x$ ,  $y$  and  $z$ , respectively, the next three rows to rotations about  $x$ ,  $y$  and  $z$ , respectively, and the last

row to the scale of the network. The coordinates  $x_i^0, y_i^0$  and  $z_i^0$  refer to the center of gravity of the network.

The Helmert matrix  $\mathbf{H}$  has a property of being a null matrix of the first order design matrix [Chen, 1983]:

$$\mathbf{A} \mathbf{H} = \mathbf{0}. \quad (6.2)$$

The datum parameters are the quantities that define the location of a free network in space. The number and type of datum parameters required to define a free network corresponds to the number and type of its datum defects. To overcome a problem of singularity of the normal equations, the datum equations have to be added to the observation equations, one for each datum defect of the network. In such case, the Gauss-Markov model with datum constraints [Casparly, 1987], [Chen et al., 1990b], reads:

$$\begin{aligned} \mathbf{l} &= \mathbf{A} \mathbf{x} + \mathbf{v}, \quad \mathbf{v} \sim (\mathbf{0}, \mathbf{C}_1 = \sigma_0^2 \mathbf{Q}), \\ \mathbf{D}^T \mathbf{x} &= \mathbf{0}, \end{aligned} \quad (6.3)$$

where:

vector of observations  $\mathbf{l}$  is normally distributed,

$\mathbf{C}_1$  is the covariance matrix of observations,

$\sigma_0^2$  is the unknown variance factor,

$\mathbf{Q}$  is the cofactor matrix of observations,

$\mathbf{D}^T \mathbf{x} = \mathbf{0}$  are the datum equations,

$\text{rank}(\mathbf{D}) = \text{rank}(\mathbf{H}) = \text{number of datum defects}$ , and

$\mathbf{H}$  is a matrix of coefficients defined by the equations of the similarity (Helmert) transformation.

Application of the least squares procedure to the Gauss-Markov model with datum constraints, eqn. (6.3), results in the following solution [Chen et al., 1990b] for the estimated vector of location parameters (vector of adjusted coordinates):

$$\hat{\mathbf{x}} = (\mathbf{A}^T \mathbf{Q}^{-1} \mathbf{A} + \mathbf{D} \mathbf{D}^T)^{-1} \mathbf{A}^T \mathbf{Q}^{-1} \mathbf{l}, \quad (6.4)$$

with the cofactor matrix of the estimated location parameters (adjusted coordinates) given as

$$\mathbf{Q}_{\hat{\mathbf{x}}} = (\mathbf{A}^T \mathbf{Q}^{-1} \mathbf{A} + \mathbf{D} \mathbf{D}^T)^{-1} - \mathbf{H} (\mathbf{H}^T \mathbf{D} \mathbf{D}^T \mathbf{H})^{-1} \mathbf{H}^T. \quad (6.5)$$

There are two alternative methods (alternative to the solution method given by eqn. (6.4) and eqn. (6.5)) for solving the Gauss-Markov model with datum constraints, eqn. (6.3). The principle of the first of them [Chen et al., 1990b] is to introduce pseudo-observations with small variances in lieu of the datum equations. The principle of the second method [Casparly, 1987] is to deal only with a subset of any  $u^*$  ( $u^* = \text{rank}(\mathbf{A})$ ) linearly independent columns of the first order design matrix  $\mathbf{A}$ . The resulting design matrix,  $\mathbf{A}^*$ , created by deleting  $d$  columns from matrix  $\mathbf{A}$  (with  $d$  being the number of datum defects:



$d = \dim(\mathbf{A}) - \text{rank}(\mathbf{A})$ ), has a rank defect of zero, and the normal equations are no longer singular.

When two epochs of measurements are analyzed, the vector of displacements can be thought of as the difference between the vector of adjusted coordinates from epoch 2,  $\hat{\mathbf{x}}_2$ , and the vector of adjusted coordinates from epoch 1,  $\hat{\mathbf{x}}_1$ ,

$$\hat{\mathbf{d}} = \hat{\mathbf{x}}_2 - \hat{\mathbf{x}}_1. \quad (6.6)$$

The cofactor matrix of the vector of displacements is then simply a sum of the cofactor matrices of the vectors of adjusted coordinates from epoch 1 and epoch 2:

$$\mathbf{Q}_{\hat{\mathbf{d}}} = \mathbf{Q}_{\hat{\mathbf{x}}_1} + \mathbf{Q}_{\hat{\mathbf{x}}_2}. \quad (6.7)$$

Owing to the fact that we are dealing with free networks, the vector of displacements, eqn. (6.6), and its cofactor matrix, eqn. (6.7), may be biased by the preselected datum. This may result from, e.g., the network having different types of datum defects in epoch 1 and epoch 2, unstable points being used in definition of the datum in at least one of the epochs, or different datum constraints being used in both epochs (e.g., due to the damage of some points). As a result, the vector of displacements may not show a true picture of displacements. To alleviate this problem a transformation to a common datum is required. The Weighted Similarity Transformation [Chen, 1983] procedure is capable of performing such a task. It transforms the vector of adjusted coordinates and its cofactor matrix from one datum to another one, without the need to repeat the least squares adjustment process. If  $\hat{\mathbf{x}}_u$  is any solution for the estimated vector of coordinates and  $\mathbf{Q}_{\hat{\mathbf{x}}_u}$  is its

cofactor matrix, then the transformed vector of the estimated coordinates (transformed to the datum defined by the datum equations  $\mathbf{D}^T \mathbf{x} = \mathbf{0}$ ), and its cofactor matrix read:

$$\hat{\mathbf{x}} = \mathbf{S} \hat{\mathbf{x}}_u, \quad (6.8)$$

$$\mathbf{Q}_{\hat{\mathbf{x}}} = \mathbf{S} \mathbf{Q}_{\hat{\mathbf{x}}_u} \mathbf{S}^T, \quad (6.9)$$

with the similarity transformation matrix defined as:

$$\mathbf{S} = \mathbf{I} - \mathbf{H}(\mathbf{D}^T \mathbf{H})^{-1} \mathbf{D}^T = \mathbf{I} - \mathbf{H}(\mathbf{H}^T \mathbf{W} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{W} \quad (6.10)$$

where:  $\mathbf{W}$  is a datum weight matrix defining the new datum, and

$$\mathbf{D}^T = \mathbf{H}^T \mathbf{W} \text{ is a transpose of the datum equations matrix.}$$

If all of the points in the network are of the same importance in defining the datum, then  $\mathbf{W} = \mathbf{I}$ , and the resulting solution is the inner constrained solution.

Since, as aforementioned, the vector of displacement components may be biased by a change of the datum definition between two epochs, a method for finding such a location of the datum that results in the least biased vector of displacements is required. A method for finding such a datum location, called the Iterative Weighted Similarity Transformation (IWST) method, was first proposed by Chen [1983]. It uses a principle of Weighted Similarity Transformation (eqn. (6.8), eqn. (6.9) and eqn. (6.10)). The objective of the method is to solve for such a datum weight matrix  $\mathbf{W}$  (cf. eqn. (6.10)) that will minimize the first norm of the vector of displacements:

$$\|\hat{\mathbf{d}}\|_1 = \min. \quad (6.11)$$

The optimization problem given in eqn. (6.11) is, in general, non-linear and the solution requires iterations [Chen, 1983]. The iterative process is always started with the datum weight matrix  $\mathbf{W}$  being the identity matrix, i.e., the vector of displacements and its cofactor matrix are transformed to the inner constraints datum. At each subsequent iteration step the weight matrix is redefined as:

$$\mathbf{W} = \left\{ \frac{1}{|d_i|} \right\}_{i=1}^u, \quad (6.12)$$

where  $d_i$  is the  $i$ -th displacement component, and

$u$  is the number of displacement components.

The Weighted Similarity Transformation is then performed with the redefined weight matrix given by eqn. (6.12). The transformation process (eqn. (6.8), eqn. (6.9), eqn. (6.10) and eqn. (6.12)) is repeated until convergence is achieved.

Once the Iterative Weighted Similarity Transformation has converged, the transformed displacement vector of each point is tested against its confidence region, using the following statistic:

$$T = \frac{\tilde{\mathbf{d}}_j^T \mathbf{Q}_{\tilde{\mathbf{d}}_j}^{-1} \tilde{\mathbf{d}}_j}{\hat{\sigma}_{0p}^2 u_d}, \quad (6.13)$$

where:  $\tilde{\mathbf{d}}_j$  is a sub-vector of displacement components of point  $j$ ,

$Q_{\tilde{\mathbf{d}}_j}$  is a cofactor sub-matrix of the displacement sub-vector  $\tilde{\mathbf{d}}_j$ ,

$u_d$  is the dimension of the displacement sub-vector  $\tilde{\mathbf{d}}_j$ ,

$\hat{\sigma}_{0p}^2$  is the pooled estimated variance factor from epoch 1 and epoch 2, and

$df_p$  is the number of its degrees of freedom.

The statistic T is compared against the value of  $F(1-\alpha; u_d, df_p)$ , where  $\alpha$  is the significance level, arising from the F distribution. If the statistic T is larger than the F value then a given point is flagged as unstable. This is equivalent to graphically comparing the displacement of each point against its confidence region (error ellipsoid, error ellipse or error bar) at a specified significance level  $\alpha$ .

## **6.2 Influence of Estimation of Variance-Covariance Components on the Iterative Weighted Similarity Transformation**

As mentioned above, statistical decisions regarding stability of the points of a monitoring network are made by comparing the value of statistic T, eqn. (6.13), against the F value:  $F(1-\alpha; u_d, df_p)$ . If the covariance matrices of observations for both epochs were estimated through the process of estimation of variance-covariance components (in the general mixed model, eqn. (2.1)) or variance components (in the variance components model, eqn. (3.1)) then the statistic T reads:

$$\mathbf{T} = \frac{\tilde{\mathbf{d}}_j^T \mathbf{C}_{\tilde{\mathbf{d}}_j}^{-1} \tilde{\mathbf{d}}_j}{u_d}. \quad (6.14)$$

In the limit, once IWST converges, the covariance matrix of the vector of transformed displacement components is given as:

$$\mathbf{C}_{\tilde{\mathbf{d}}} = \mathbf{S} \mathbf{C}_{\hat{\mathbf{d}}} \mathbf{S}^T = \mathbf{S} (\mathbf{C}_{\hat{\mathbf{x}}_1} + \mathbf{C}_{\hat{\mathbf{x}}_2}) \mathbf{S}^T, \quad (6.15)$$

where  $\mathbf{S}$  is the final similarity transformation matrix,

$$\mathbf{S} = \mathbf{I} - \mathbf{H} (\mathbf{H}^T \mathbf{W} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{W}, \text{ with} \quad (6.16)$$

$$\mathbf{W} = \left\{ \begin{matrix} 1 \\ d \\ |\tilde{\mathbf{d}}_i| \end{matrix} \right\}_{i=1}^u, \text{ and} \quad (6.17)$$

$\mathbf{C}_{\hat{\mathbf{x}}_1}$  and  $\mathbf{C}_{\hat{\mathbf{x}}_2}$  are the covariance matrices of the vectors of adjusted coordinates from epoch 1 and epoch 2, respectively, where

$$\mathbf{C}_{\hat{\mathbf{x}}} = (\mathbf{A}^T \mathbf{C}_{\hat{\boldsymbol{\theta}}}^{-1} \mathbf{A} + \mathbf{D} \mathbf{D}^T)^{-1} - \mathbf{H} (\mathbf{H}^T \mathbf{D} \mathbf{D}^T \mathbf{H})^{-1} \mathbf{H}^T. \quad (6.18)$$

The transformed vector of displacement components is given as:

$$\tilde{\mathbf{d}} = \mathbf{S} \hat{\mathbf{d}} = \mathbf{S} (\hat{\mathbf{x}}_2 - \hat{\mathbf{x}}_1), \quad (6.19)$$

where the vector of adjusted coordinates is expressed as

$$\hat{\mathbf{x}} = (\mathbf{A}^T \mathbf{C}_{\hat{\boldsymbol{\theta}}}^{-1} \mathbf{A} + \mathbf{D} \mathbf{D}^T)^{-1} \mathbf{A}^T \mathbf{C}_{\hat{\boldsymbol{\theta}}}^{-1} \mathbf{l}. \quad (6.20)$$

The estimation of variance or variance-covariance components affects primarily the covariance matrix of the vector of adjusted coordinates, eqn. (6.18), through the changes in the covariance matrix of observations, due to such factors as the estimation method, the choice of the error model or the amount of prior information. The vector of adjusted coordinates (in both epochs), eqn. (6.20), is also affected.

The final similarity transformation matrix, eqn. (6.16), is affected by changes in the vector of displacement components, eqn. (6.6), through the datum weight matrix, eqn. (6.12). Changes in the initial vector of displacement components,  $\hat{\mathbf{d}}$ , being the result of changes in the vectors of adjusted coordinates - for both epochs, will cause changes in the datum weight matrix,  $\mathbf{W}$ , propagating through the iterative process of IWST.

The similarity transformation matrix, eqn. (6.16), is also influenced by changes in the vector of adjusted coordinates which are used in computing the rotation and scale elements in the Helmert matrix  $\mathbf{H}$  (see eqn. (6.1)). This effect is, however, likely to be insignificant as the changes of the adjusted coordinates are small in comparison with the diameter of a typical geodetic network.

As a result, both the final covariance matrix of the transformed displacement components, eqn. (6.15), and the vector of the transformed displacement components, eqn. (6.19), will

be affected, thus changing the values of the T statistic, eqn. (6.14), and therefore influencing statistical decisions concerning stability of the reference points.

The influence of various aspects of the estimation of variance components on the results of the spatial deformation trend analysis will be further investigated in Chapter 7, based on the analysis of the Mactaquac monitoring network [Chrzanowski and Secord, 1987 and 1990], [Chrzanowski et al., 1989].

## CHAPTER 7

### NUMERICAL EXAMPLES

In this chapter the proposed methods and procedures, derived in Chapters 3 and 5 are tested on the data from periodic surveys of a horizontal monitoring network at the Mactaquac hydroelectric generating station near Fredericton, New Brunswick, [Chrzanowski and Secord, 1987 and 1990], [Chrzanowski et al., 1989]. First, the GML estimating equations with the noninformative and the inverted gamma priors and the estimating equations resulting from application of the least squares principle to the dispersion-mean model with weighted constraints, all derived in Chapter 3, are applied to periodic observations of the Mactaquac network. The results are compared with the results of REML estimation. Subsequently, the influence of the choice of the estimation method, the amount of prior information, and the choice of the error model, on the results of the spatial deformation trend analysis process is quantified - based on the results of the 1991 and 1993 surveys of the Mactaquac monitoring network. Finally, the algorithm for detecting influential observations, derived in Chapter 5, is applied to the 1989 Mactaquac survey data set.



**7.1 Estimation of Variance Components for the Mactaquac Monitoring Network:  
Campaigns of 1986, 1989, 1991 and 1993**

In this section the results of the estimation of variance components for a geodetic horizontal monitoring network (Fig. 7.1), located in the vicinity of the Mactaquac power generating station (see [Chrzanowski and Secord, 1987 and 1990] and [Chrzanowski et al., 1989]) are presented. Four campaigns are analyzed (see Table 7.1). The 1986 and 1989 campaigns involve only distance measurements. The 1991 and 1993 campaigns involve both distance and direction measurements.

Table 7.1 Summary of the Mactaquac monitoring network: campaigns of 1986, 1989,  
1991 and 1993

<b>Year</b>	<b>Number of Stations</b>	<b>Number of Distances</b>	<b>Number of Directions</b>	<b>Instruments Used</b>
1986	28	252	-	Kern ME 3000
1989	28	254	-	Kern ME 3000
1991	32	193	83	Tellurometer MA 200 Kern E 2
1993	32	182	80	Tellurometer MA 200 Kern E 2

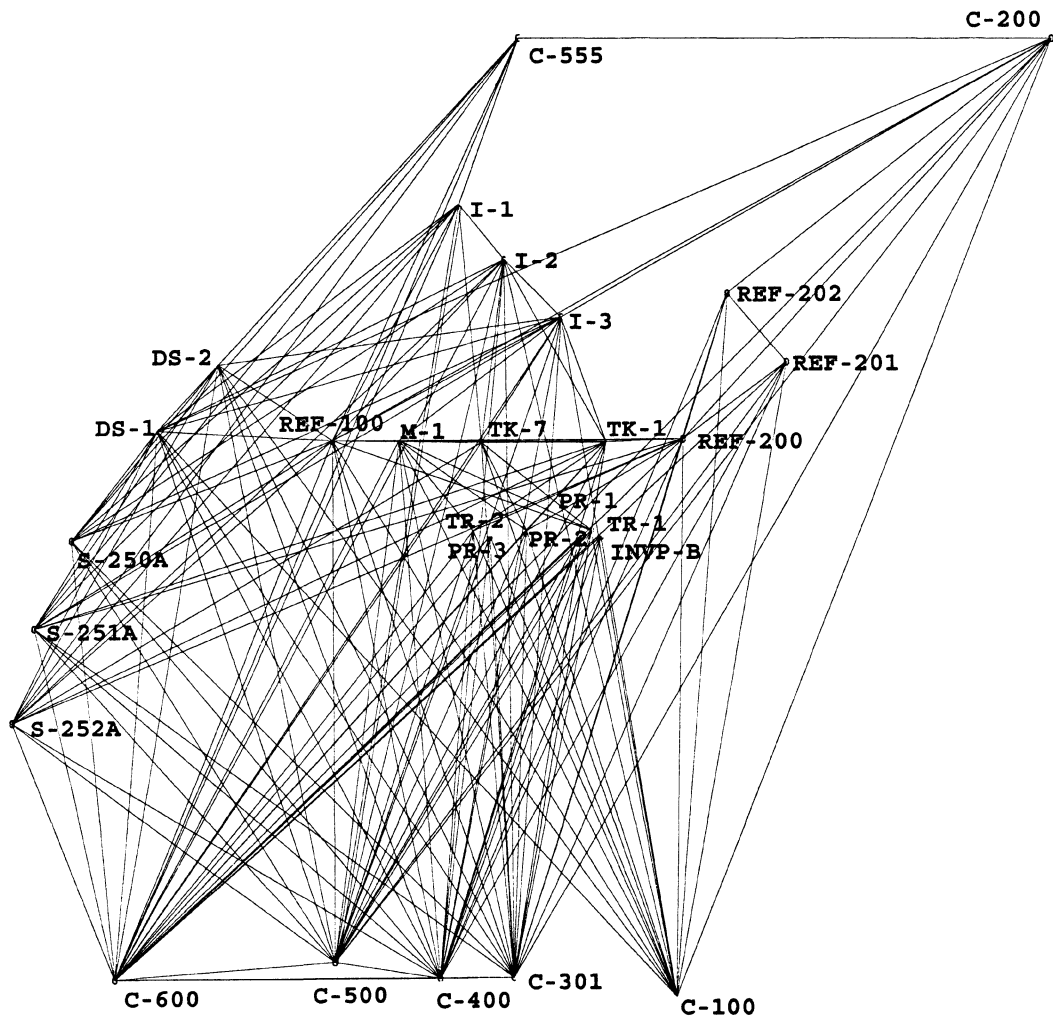


Figure 7.1 Mactaquac Horizontal Monitoring Network

For all four campaigns the weighting scheme for distance observations [Chrzanowski et al., 1994] is based on the following model:

$$\sigma_S^2 = a_S^2 + b_S^2 L^2, \quad (7.1)$$

where  $\sigma_S^2$  is the variance of the distance observation,  $a_S^2$  and  $b_S^2$  are the unknown variance components, and  $L$  is an approximate fixed value for the true distance being measured. It should be noted that even though the exact true distances are not known, the effect of using the approximate values in eqn. (7.1) (being either the measured distances themselves or the distances derived from the initial least squares adjustment) will be negligible for all practical purposes, as the measurements of distances in geodetic networks are performed with very high accuracy, and since the error model, eqn. (7.1), is only an approximation of the stochastic model of the electro-optical distance measurement process.

As far as the campaigns of 1991 and 1993 are concerned, the variances of direction observations are computed from the following model:

$$\sigma_D^2 = a_D^2, \quad (7.2)$$

where  $a_D^2$  is the unknown variance component (the variance of a direction measurement).

The empirical, *a priori*, values of variance components for all four epochs are given in Table 7.2 [Chrzanowski et al., 1994]. It should be mentioned here, that the values given in Table 7.2 were arrived at by means of the trial and error approach and did not result from the rigorous analysis of the sources of errors [Chrzanowski, 1995]. Moreover, some of the observations were assessed to be the outliers and were removed from the analyzed

data sets. In particular, the 1989 data set used by the author differs from the 1989 data set used to obtain the variance components given in Table 7.2, as the number of removed outlying observations is not the same. As a result, the estimated values of the variance components, presented later in this section, are not directly comparable with the *a priori* values given in Table 7.2. It should be noted that the empirical, *a priori*, values of variance components are used in this section solely for the purpose of introducing prior information.

Table 7.2 The empirical, *a priori*, values of variance components  
[Chrzanowski et al., 1994]

Epoch	$a_S^2$ [mm <sup>2</sup> ]	$b_S^2$ [ppm <sup>2</sup> ]	$a_D^2$ [sec <sup>2</sup> ]
1986	0.25	9.0	–
1989	0.09	16.0	–
1991	0.09	4.0	0.81
1993	0.09	4.0	0.64

In this section, the following four estimation methods are used with the Mactaquac data:

1. the REML method,
2. the GML method with the inverted gamma prior,
3. the GML method with the noninformative prior, and
4. the dispersion-mean model with weighted constraints.

As far as the REML estimation method is concerned, the estimating equation presented in Chapter 2 (eqn. 2.63) is applied. In all other three cases the estimating equations derived in Chapter 3 (eqn. 3.34, eqn. 3.123, and eqn. 3.147) are used.

The software "MINQE", version 1.2, [Chrzanowski et al., 1994] is used for REML estimation, while a modified and partly rewritten, by the author, version of the "MINQE" software is used in the other three cases.

### ***7.1.1 Conformity with the Hartley-Rao Conditions***

In Chapter 2 (Subsection 2.2.1) the Hartley-Rao conditions [Hartley and Rao, 1967], required for establishing asymptotic properties of the ML estimates of variance components, are given. In this subsection conformity of the Mactaquac monitoring network with the Hartley-Rao conditions is investigated.

Hartley and Rao [1967] analyze a series of experiments for which the number of observations increases to infinity, and for which the Hartley-Rao conditions 1, 2 and 3 (see Subsection 2.2.1, page 30) hold. They establish that in the limit the ML estimate of the vector of variance components, provided it is computed at a global maximum of the likelihood function, is weakly consistent, asymptotically efficient and asymptotically normally distributed with the mean equal to the true value of the vector of variance components ( $\theta_0$ ) and the covariance matrix equal to the inverse of the information matrix computed at  $\theta_0$ .

Hartley and Rao [1967] consider the variance components model given by eqn (2.3). They further restrict this model by requiring that all matrices  $U_i$  have in each row one element equal to 1 and all the remaining elements equal to 0. One of the implications of this restriction is that the matrices  $U_i^T U_i$  are diagonal. In this subsection, the general mixed model given by eqn. (2.1), restricted to the variance components models implied by eqn. (7.1) and eqn. (7.2), will be considered.

Let us first consider applicability of the Hartley-Rao conditions to the Mactaquac network for the first two epochs (1986 and 1989). As far as the epochs of 1986 and 1989 are concerned, the network consists of distances only, and with the stochastic model given by eqn. (7.1), the variance components model reads:

$$\mathbf{l} = \mathbf{Ax} + \mathbf{v}, \quad \mathbf{v} \sim (\mathbf{0}, \mathbf{C}_\theta = a_S^2 \mathbf{T}_1 + b_S^2 \mathbf{T}_2), \quad \mathbf{v} \sim N_m(\mathbf{0}, \mathbf{C}_\theta), \quad (7.3)$$

with

$$\mathbf{T}_1 = \mathbf{I}_m, \text{ and} \quad (7.4)$$

$$\mathbf{T}_2 = \left\{ {}_d L_i^2 \right\}_{i=1}^m. \quad (7.5)$$

Using the notation of eqn. (2.3) and eqn. (2.4) we can express this variance components model as:

$$\mathbf{l} = \mathbf{Ax} + \mathbf{U}_1 \mathbf{b}_1 + \mathbf{U}_2 \mathbf{b}_2, \quad (7.6)$$

where

$$\mathbf{U}_1 = \mathbf{I}_m, \quad (7.7)$$

$$\mathbf{U}_2 = \left\{ {}_d L_i \right\}_{i=1}^m, \quad (7.8)$$

with  $\mathbf{b}_1$  and  $\mathbf{b}_2$  being vectors of  $m$  independent random variables from  $N(0, a_S^2)$  and  $N(0, b_S^2)$ , respectively, and hence, the covariance matrix of the vector of observations given by

$$\mathbf{C}_\theta = a_S^2 \mathbf{U}_1 \mathbf{U}_1^T + b_S^2 \mathbf{U}_2 \mathbf{U}_2^T. \quad (7.9)$$

The first Hartley-Rao condition requires that as the total number of observations and the number of observations corresponding to each random factor (variance component) increases to infinity all positive elements of the diagonal matrices  $\mathbf{U}_i^T \mathbf{U}_i$  remain smaller than some constant  $R$ . Using the variance components model for epochs 1986 and 1989, given by equations (7.6) to (7.9), we find that there always exists  $R > 0$  such that

$$\mathbf{U}_1^T \mathbf{U}_1 = \mathbf{I}_m \leq R \mathbf{I}_m, \text{ and} \quad (7.10)$$

$$\mathbf{U}_2^T \mathbf{U}_2 = \left\{ {}_d L_i^2 \right\}_{i=1}^m \leq R \mathbf{I}_m, \quad (7.11)$$

and is given by:

$$R \geq \max(1, L_i^2), \quad i = 1, 2, \dots, m. \quad (7.12)$$

The second Hartley-Rao condition, which requires that all elements of the inverse of the matrix of normal equations are smaller than  $R/m$  (where  $R$  is some constant and  $m$  is the total number of observations), ensures consistent estimation of all elements of the location

vector. This condition is satisfied if all repeated observations in the network are distributed evenly, i.e., there is no such situation where, e.g., only one distance is remeasured an increasing number of times.

The third Hartley-Rao condition is the condition for estimability of the vector of location parameters and the vector of variance components. It requires that the base  $\mathbf{W}$  of the adjoined matrix  $\mathbf{M}=[\mathbf{A}|\mathbf{U}_1|\dots|\mathbf{U}_r]$  is of the form  $\mathbf{W}=[\mathbf{A}|\mathbf{U}^*]$ , where  $\mathbf{U}^*$  contains at least one column from each  $\mathbf{U}_i$ . This implies that all column vectors of the design matrix  $\mathbf{A}$  are linearly independent, and that at least one column vector of each matrix  $\mathbf{U}_i$  is linearly independent from all column vectors of  $\mathbf{A}$ .

For the Mactaquac network of epochs 1986 and 1989 the adjoined matrix  $\mathbf{M}$  is of the following form:

$$\mathbf{M}=[\mathbf{A}|\mathbf{U}_1|\mathbf{U}_2], \quad (7.13)$$

where  $\mathbf{U}_1$  and  $\mathbf{U}_2$  are given by eqn. (7.7) and eqn. (7.8), respectively. The column vectors of the design matrix  $\mathbf{A}$  are all linearly independent if the datum defects have been eliminated and the network does not suffer from the configuration defects, i.e., there are no points in the network that are determined by a single distance observation. One of the many possible forms of  $\mathbf{U}^*$  is:

$$\mathbf{U}^* = \begin{bmatrix} \{L_i\}_{i=1}^p & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{m-p} \end{bmatrix}, \quad (7.14)$$



where the matrix  $\{L_i\}_{i=1}^p$  corresponds to the first  $p$  observations.

For epochs 1991 and 1993 the Mactaquac network consists of both distance and direction observations. Using the stochastic models for distance and direction observations, given by eqn. (7.1) and eqn. (7.2), we find the variance components model to be:

$$\mathbf{l} = \mathbf{Ax} + \mathbf{v}, \quad \mathbf{v} \sim (\mathbf{0}, \mathbf{C}_\theta = a_S^2 \mathbf{T}_1 + b_S^2 \mathbf{T}_2 + a_D^2 \mathbf{T}_3), \quad \mathbf{v} \sim N_m(\mathbf{0}, \mathbf{C}_\theta), \quad (7.15)$$

with

$$\mathbf{T}_1 = \begin{bmatrix} \mathbf{I}_{m_1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}, \quad (7.16)$$

$$\mathbf{T}_2 = \begin{bmatrix} \{L_i\}_{i=1}^{m_1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}, \text{ and} \quad (7.17)$$

$$\mathbf{T}_3 = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{m_2} \end{bmatrix}, \quad (7.18)$$

where  $m_1$  and  $m_2$  is the number of distance and direction observations, respectively ( $m_1 + m_2 = m$ ). Equivalently, using Hartley and Rao's [1967] notation, we can express this variance components model as:

$$\mathbf{l} = \mathbf{Ax} + \mathbf{U}_1 \mathbf{b}_1 + \mathbf{U}_2 \mathbf{b}_2 + \mathbf{U}_3 \mathbf{b}_3, \quad (7.19)$$

where

$$\mathbf{U}_1 = \begin{bmatrix} \mathbf{I}_{m_1} \\ \mathbf{0} \end{bmatrix}, \quad (7.20)$$

$$\mathbf{U}_2 = \begin{bmatrix} \{L_i\}_{i=1}^{m_1} \\ \mathbf{0} \end{bmatrix}, \text{ and} \quad (7.21)$$

$$\mathbf{U}_3 = \begin{bmatrix} \mathbf{0} \\ \mathbf{I}_{m_2} \end{bmatrix}, \quad (7.22)$$

with  $\mathbf{b}_1$  and  $\mathbf{b}_2$  being vectors of  $m_1$  independent random variables from  $N(0, a_S^2)$  and  $N(0, b_S^2)$ , respectively, and  $\mathbf{b}_3$  being a vector of  $m_2$  independent random variables from  $N(0, a_D^2)$ . The covariance matrix of the vector of observations is given by

$$\mathbf{C}_\theta = a_S^2 \mathbf{U}_1 \mathbf{U}_1^T + b_S^2 \mathbf{U}_2 \mathbf{U}_2^T + a_D^2 \mathbf{U}_3 \mathbf{U}_3^T. \quad (7.23)$$

The first Hartley-Rao condition, requiring that all positive elements of the diagonal matrices  $\mathbf{U}_i^T \mathbf{U}_i$  are smaller than some constant  $R$ , is verified as follows. From the variance components model for epochs 1991 and 1993 (see equations (7.15) to (7.23)), we see that if  $R$  is chosen as

$$R \geq \max(1, L_i^2), \quad i = 1, 2, \dots, m_1, \quad (7.24)$$

it will satisfy the first Hartley-Rao condition given by:

$$\mathbf{U}_1^T \mathbf{U}_1 = \mathbf{I}_{m_1} \leq R \mathbf{I}_{m_1}, \quad (7.25)$$

$$\mathbf{U}_2^T \mathbf{U}_2 = \{L_i^2\}_{i=1}^{m_1} \leq R \mathbf{I}_{m_1}, \text{ and} \quad (7.26)$$

$$\mathbf{U}_3^T \mathbf{U}_3 = \mathbf{I}_{m_2} \leq R \mathbf{I}_{m_2}. \quad (7.27)$$

As for the epochs of 1986 and 1898, the second Hartley-Rao condition is satisfied for the epochs of 1991 and 1993 if all repeated observations in the network are distributed evenly.

The third Hartley-Rao condition is verified as follows. For epochs 1991 and 1993 the adjoined matrix  $\mathbf{M}$  is given as:

$$\mathbf{M}=[\mathbf{A}|\mathbf{U}_1|\mathbf{U}_2|\mathbf{U}_3], \quad (7.28)$$

where  $\mathbf{U}_1$ ,  $\mathbf{U}_2$  and  $\mathbf{U}_3$  are given by eqn. (7.20), eqn. (7.21) and eqn. (7.22), respectively.

Again, the column vectors of the design matrix  $\mathbf{A}$  are all linearly independent if the datum defects have been eliminated and the network does not suffer from the configuration defects. There are no configuration defects in the geodetic network of the type analyzed here if, and only if, each point of the network is determined by at least two observations (2 distances, or 2 directions, or 1 distance and 1 direction). The matrix  $\mathbf{U}^*$  of a dimension  $m$  by  $m$  may be expressed, e.g., as:

$$\mathbf{U}^* = \begin{bmatrix} \left\{ \begin{smallmatrix} \mathbf{d} \\ \mathbf{L}_i \end{smallmatrix} \right\}_{i=1}^p & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{m_1-p} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_{m_2} \end{bmatrix}, \quad (7.29)$$

where the matrix  $\left\{ \begin{smallmatrix} \mathbf{d} \\ \mathbf{L}_i \end{smallmatrix} \right\}_{i=1}^p$  corresponds to the first  $p$  distance observations.

It should be noted that the Hartley-Rao conditions are formulated for the variance components model (cf. eqn. (2.3)) for which all elements of matrices  $\mathbf{U}_i$  (and hence of  $\mathbf{T}_i$ ) are known fixed numbers. As discussed before (see Section 7.1), even though the exact

true distances  $L_i$  (used in eqn. (7.5) and eqn. (7.17)) are not known the approximate values being, e.g., the distances derived from the initial least squares adjustment of a network with an approximate covariance matrix of observations, can be treated as known fixed numbers as they do not change with additional observations (they will stay the same as  $m_1 \rightarrow \infty$ ).

Based on the presented arguments, it may be concluded that the Mactaquac monitoring network, for all four campaigns, fulfills the Hartley-Rao conditions - under the assumptions specified above. As a result, the asymptotic properties of the ML estimates of variance components, as established by Hartley and Rao [1967], may be applied for analyzing the results of the estimation of variance components, provided the number of observations is sufficiently large.

### ***7.1.2 Discussion of the Results***

A complete listing of the results of the estimation of variance components for the Mactaquac monitoring network, for all four epochs and all four analyzed methods of estimation, is given in Appendix C in Tables C.1 to C.22.

As far as the REML (Section C.1) and the GML with the noninformative prior (Section C.3) methods of estimation are concerned, the values of the estimated variance components, their asymptotic standard deviations and the speed of convergence are given.

For the GML with the inverted gamma prior (Section C.2) and the dispersion-mean model with weighted constraints (Section C.4) estimation methods the prior values of the components (means of the prior distribution), the values of the standard deviations of the priors, the values of the estimated variance components, their asymptotic standard deviations and the speed of convergence are given. For both these methods the results are tabulated for the values of the standard deviations of the priors ranging from 1 to 16 times the value of the prior.

The results of the estimation, in a graphical form, are also given in Figures C.1 to C.20 in Section C.5 and C.6. In order to facilitate comparison of different methods of estimation, the charts are organized on an epoch by epoch basis.

The values of the estimated components, for all four campaigns and all four methods of estimation, together with the prior values (means of the prior distribution) are shown in Figures C.1 to C.10 (Section C.5). To simplify interpretation by expressing the results in more intuitive units (i.e., millimeters, parts per million, and seconds of arc) the square root values of the variance components are given.

To facilitate the assessment of the impact of both the method of estimation and the amount of prior information on the level of uncertainty associated with the estimates

(tightness of the confidence intervals) the ratios of the estimated components to their asymptotic standard deviations are given in Figures C.11 to C.20 (Section C.6).

Both the REML and the GML with the noninformative prior estimation methods do not introduce any prior information about the variance components. From the examination of the results given in Tables C.1 to C.6 and C.12 to C.17 and Figures C.1 to C.20, the following conclusions can be drawn:

1. As far as the rate of convergence is concerned, both methods gave comparable results (see Tables C.1 to C.6 and C.12 to C.17).
2. The GML estimated variance components were consistently larger than their counterparts estimated with the REML method (see Figures C.1 to C.2 and C.5 to C.10).
3. The asymptotic standard deviations of the GML estimated variance components were consistently larger than their corresponding values obtained from the REML estimation method (see Tables C.1 to C.6 and C.12 to C.17).
4. The ratios of the estimated components to their asymptotic standard deviations were consistently smaller for the GML method than for the REML method (see Figures C.11 to C.12 and C.15 to C.20). The only exception was the  $b_{\xi}^2$  component for the 1993 campaign, as shown in Figure C.19. Thus, introduction of the noninformative prior appears to have decreased the significance of the estimated components.

Both the GML with the inverted gamma prior estimation method and the dispersion-mean model with weighted constraints make a provision for introduction of prior information concerning the unknown variance components. When comparing the results produced by these two methods with the REML results the following was observed:

1. The rate of convergence varied widely between all three estimation methods. In most cases introduction of prior information reduced the speed of convergence (see Tables C.1 to C.6, C.7 to C.11 and C.18 to C.22).
2. When very low weights were assigned to the prior values (means of the prior distribution), the variance components estimated using the dispersion-mean model with weighted constraints were virtually identical to the REML estimated values (see Figures C.1 to C.2 and C.5 to C.10). The results of GML estimation, even when high standard deviations were assigned to the prior values (means of the prior distribution), in general differed from the REML estimates.
3. The same observation applies to the asymptotic standard deviations of the estimated variance components (see Tables C.1 to C.6, C.7 to C.11 and C.18 to C.22).
4. The following trends were observed when analyzing the uncertainty of the estimated variance components:
  - a) Introduction of prior information for the  $a_S^2$  component (see Figures C.11, C.15 and C.18), by means of the GML method with the inverted gamma prior, reduced its significance. No similar trend could be observed for the dispersion-mean model with weighted constraints.

- b) Introduction of prior information (see Figures C.12, C.16 and C.19) increased, in general, significance of the  $b_S^2$  component.
- c) The effects of introduction of prior information on the uncertainty of the  $a_D^2$  component were mixed (see Figures C.17 and C.20).

When comparing the results of the GML with the inverted gamma prior estimation method with the results obtained from the dispersion-mean model with weighted constraints it was found that:

1. The rate of convergence differed considerably between the two analyzed methods. The increase of the amount of prior information (decrease of the standard deviations of the priors) did not appear to have had any significant impact on the rate of convergence (see Tables C.7 to C.11 and C.18 to C.22).
2. The estimates derived using the dispersion-mean model with weighted constraints were much more strongly attracted towards the prior values (means of the prior distribution) than the GML estimates; even though, initially (i.e., when very high standard deviations are assigned to the priors) the GML estimates were closer to the prior values (see Figures C.1 to C.10).
3. The asymptotic standard deviations of the estimated variance components differed significantly between the two methods. No discernible pattern of differences could, however, be observed (see Tables C.7 to C.11 and C.18 to C.22).



4. Significance of the estimated components was much more strongly affected by the change of the amount of prior information in the dispersion-mean model with weighted constraints than in the GML method. In particular:
- a) For both methods, the ratio of the estimated component to its asymptotic standard deviation decreased with the increase of the amount of prior information for the  $a_{\xi}^2$  component (see Figures C.11, C.13 and C.18). The only exception was the campaign of 1991 (see Figure C.15).
  - b) The ratio increased, and thus the uncertainty of the estimation became smaller when more prior information was added, in the case of  $b_{\xi}^2$  component (see Figures C.12, C.14, C.16 and C.19).
  - c) The influence of the amount of prior information on the uncertainty of the  $a_{\text{D}}^2$  component was much smaller than on the uncertainty of the two variance components corresponding to the distance observations. In general, the uncertainty of the estimation decreased slightly when more prior information was added (see Figures C.17 and C.20).

When estimating the variance components for the Mactaquac data, some computational problems were encountered. They were:

1. Complete lack of convergence, encountered when using:
  - a) the REML and GML with the noninformative prior estimation methods with the 1989 data set (see Table C.2 and C.13), and

- b) the GML method with the inverted gamma prior, and the ratio of the standard deviation of the prior to the prior value (mean of the prior distribution) of 1, with the 1991 data set (see Table C.9).
2. Lack of convergence when the estimation process starts with the default initial values of unknown variance components. This phenomenon occurred when processing the 1989 data, using the GML method with the inverted gamma prior (see Table C.8).
  3. Very slow rate of convergence arising when:
    - a) the GML method with the inverted gamma prior was applied to the 1989 and 1991 data sets (see Tables C.8 and C.9), and
    - b) the dispersion-mean model with weighted constraints was used with the 1989 data set (see Table C.19).

In the case of the 1989 data, where both the REML method and the GML method with the noninformative prior failed to produce any results, introduction of prior information about the unknown variance components forced the iterative process to converge.

Lack of convergence when using the default initial values of variance components was overcome by using better initial values - closer to the final estimates.

Slow convergence encountered when processing the 1989 data set appears to have arisen, to some extent, from the presence of undetected outliers in the data. It is worth to note that both the REML method and the GML method with the noninformative prior failed to

produce any results for the 1989 data set, due to divergence of the computational procedure. However, as indicated in Section 7.3, removal of all formal outliers from the 1989 data set caused the REML estimation process to converge.

## **7.2 Influence of the Estimation of Variance Components on the Spatial Deformation Trend Analysis Process**

The influence of the estimation of variance components on the deformation trend analysis process, in particular on the results of the Iterative Weighted Similarity Transformation (IWST), is analyzed in this section, based on the 1991 and 1993 Mactaquac campaigns (see Section 7.1). The influence of the error model, the method of estimation, and the amount of prior information on the transformed displacement components and their confidence regions and on the IWST results is investigated.

For the 1991 campaign a 3 parameter error model was used for the estimation of variance components. The model comprises the  $a_S^2$  and  $b_S^2$  terms (see eqn. 7.1) - for the distance measurements - and the  $a_D^2$  term (see eqn. 7.2) for the direction measurements. For the 1993 campaign two error models were considered: the 3 parameter model consisting of the  $a_S^2$ ,  $b_S^2$ , and  $a_D^2$  terms, and a 2 parameter model consisting of the  $b_S^2$  and  $a_D^2$  terms.

### ***7.2.1 Influence of the Error Model***

The influence of the choice of the variance components model on the results of the spatial deformation trend analysis is investigated in this subsection - for the REML estimation

method. For the 1991 data the 3 parameter variance components model was used. For the 1993 data either the 2 parameter or the 3 parameter model was utilized.

The results of IWST, for all the reference network points, are shown in Figures 7.2 and 7.3. Figure 7.2 depicts the transformed displacements and their error ellipses for the case when the 2 parameter error model is used for the 1993 data. Figure 7.3 shows the IWST outcome for the 3 parameter case. The displacement field and the error ellipses, for both error models, are compared in Figure 7.4.

It is seen that the reference points REF-200, REF-201 and REF-202, which were assessed to be stable when the 2 parameter model was used, were showing to have undergone significant movements when the 3 parameter model was adopted. On the other hand, points C-100 and C-200, assessed to be unstable under the 3 parameter model, turned out to be stable when the 3 parameter model was utilized. Altogether, the results of IWST, being the verdict whether a point is stable or not, changed for 5 out of 14 reference points with the change of the error model for the 1993 epoch.

It is seen in Figure 7.4 that the displacement vectors were only moderately affected by the change of the error model. The impact of the error model on the error ellipses of the displacement vectors was, however, much more pronounced; the change of the error model caused changes of both the size and the orientation of the error ellipses.

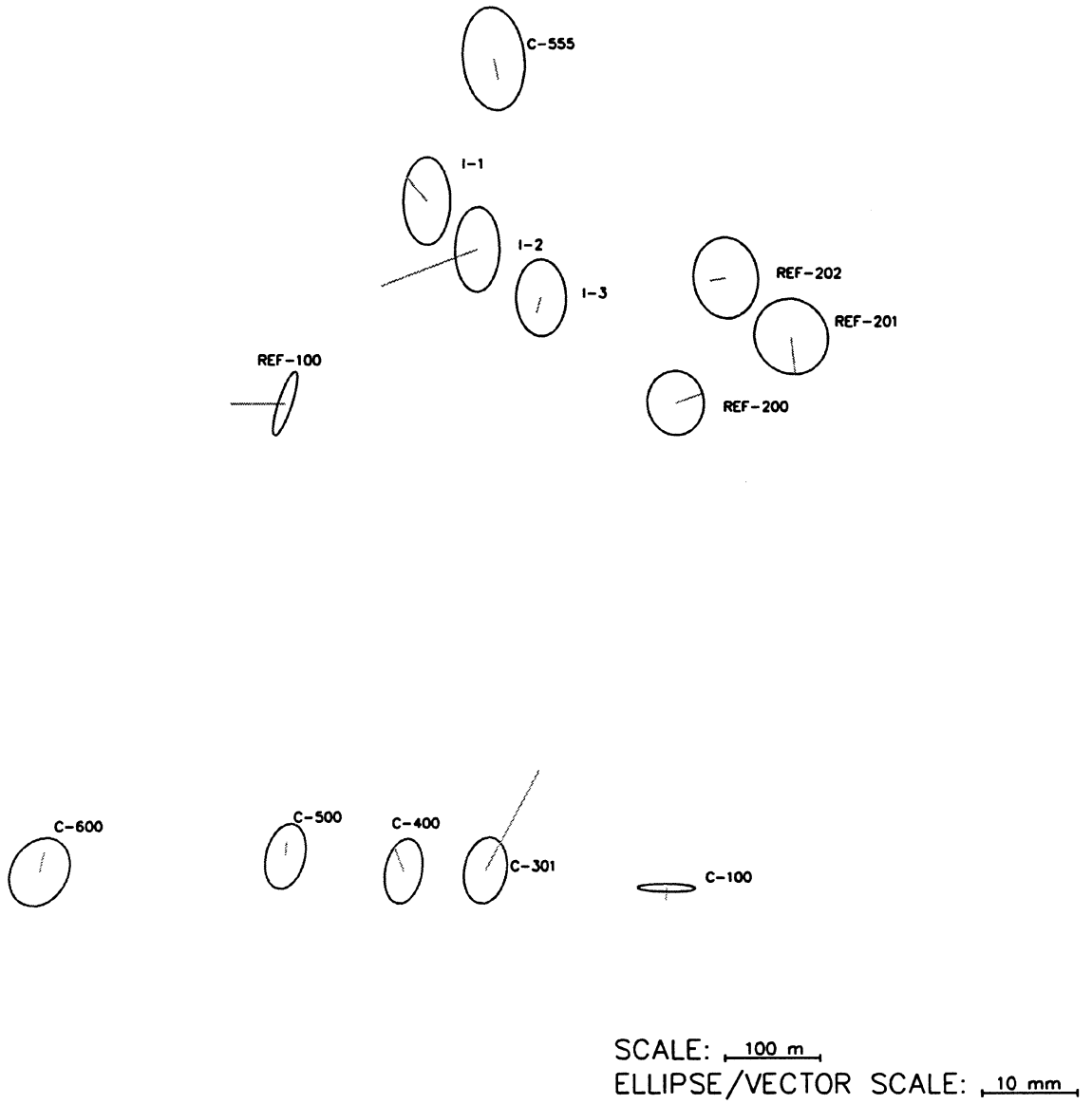


Figure 7.2 Displacement Field and Error Ellipses (95%):  
2 Components Model Used for the 1993 Data,  
REML Estimation Method

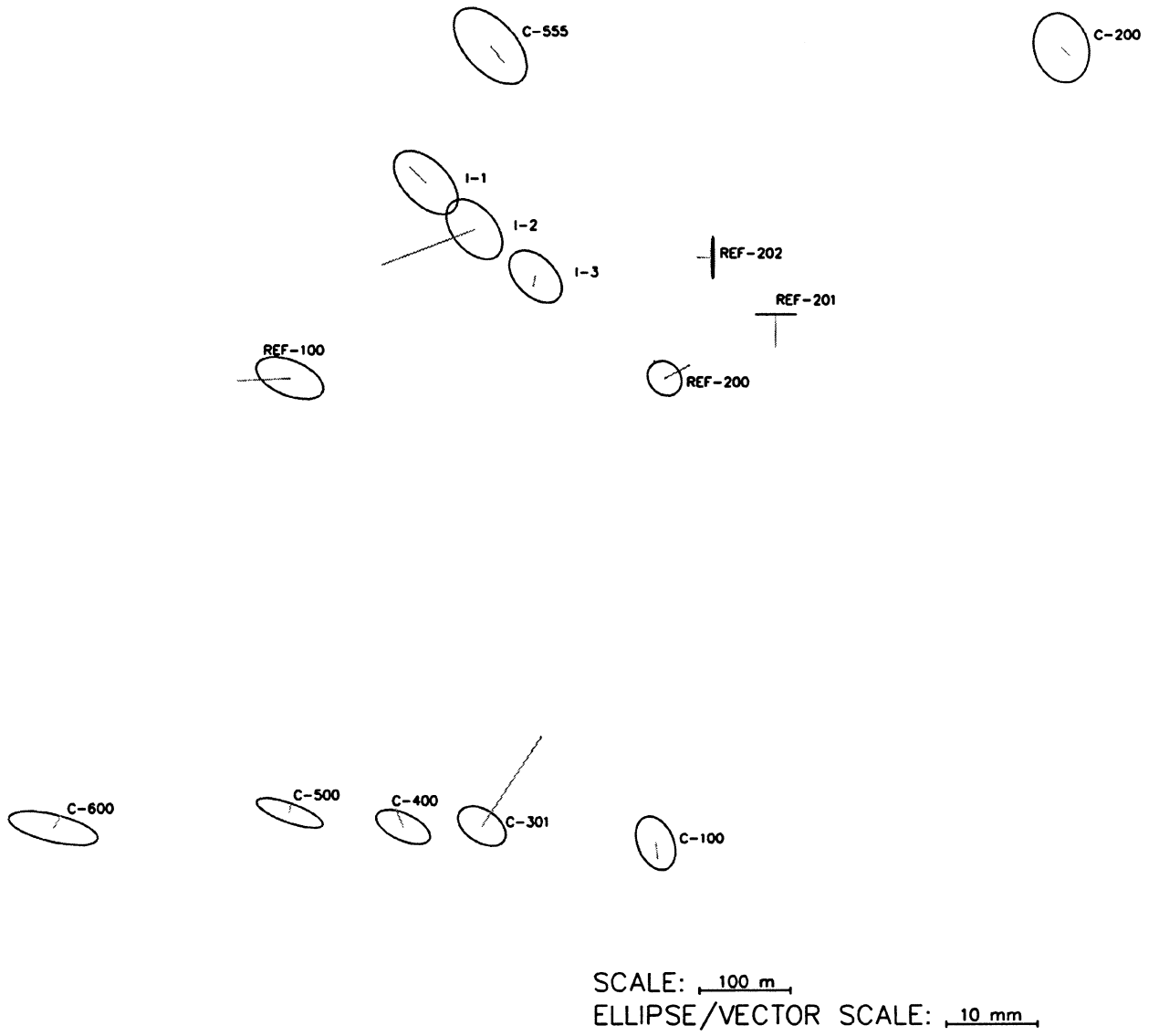


Figure 7.3 Displacement Field and Error Ellipses (95%):  
 3 Components Model Used for the 1993 Data,  
 REML Estimation Method

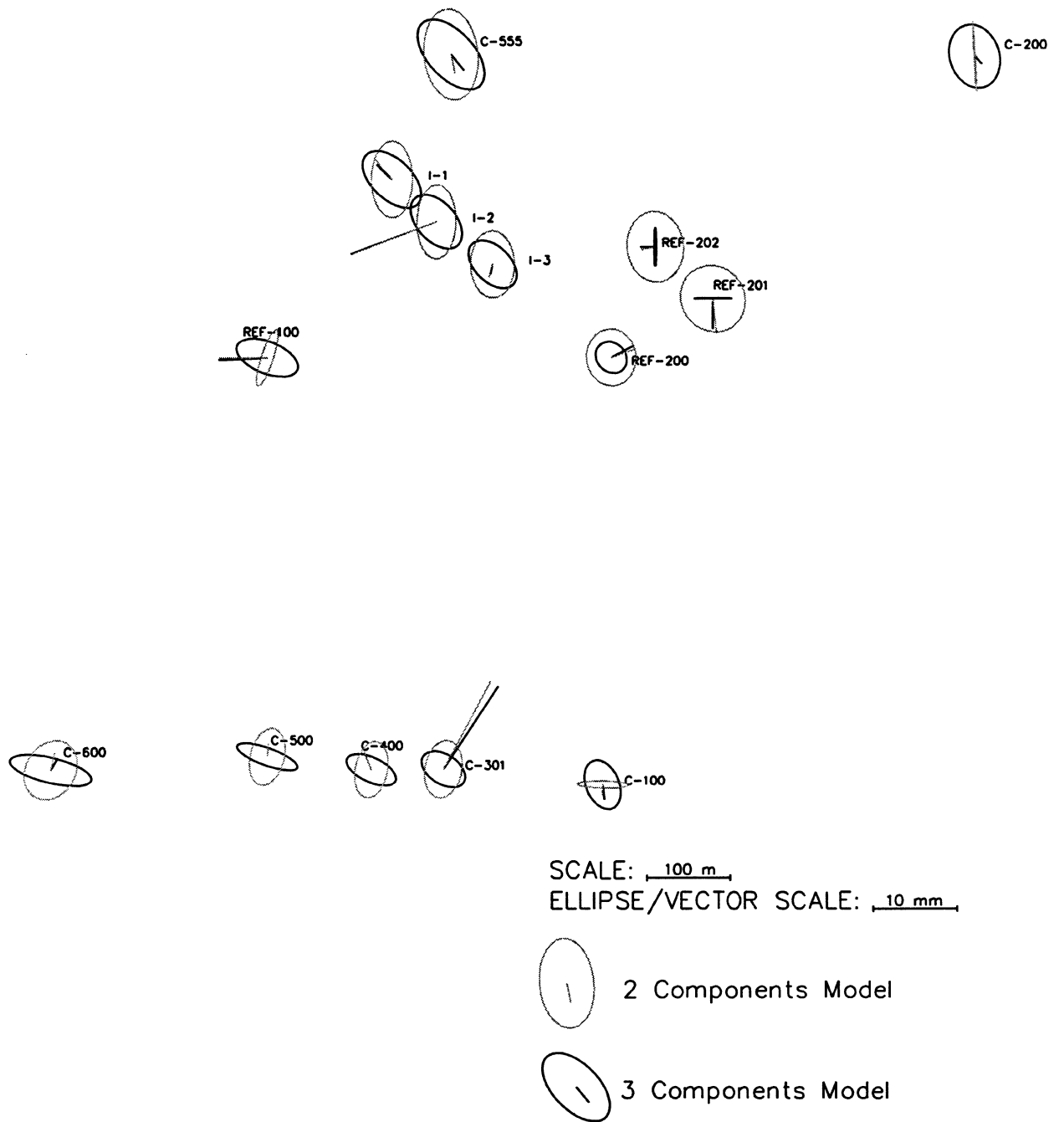


Figure 7.4 Displacement Field and Error Ellipses (95%):  
 2 Components Model versus 3 Components Model

### ***7.2.2 Influence of the Method of Estimation***

The influence of the method chosen for the estimation of variance components on the results of the spatial deformation trend analysis is examined in this subsection. As before, the 3 parameter variance components model was used for the 1991 data while either the 2 parameter or the 3 parameter model was utilized for the 1993 data set.

It was observed that the impact of the choice of the variance components estimation method on the displacement field and the error ellipses resulting from IWST was much less prominent than the impact of the error model. The largest difference occurred between the REML and the GML with the inverted gamma prior (with the ratio of the standard deviation of the prior to the prior value of 2) estimation methods, for the 3 components error model.

The results of IWST for the case when the GML with the inverted gamma prior estimation method was used, are shown in Figure 7.5. The comparison with the REML case is depicted in Figure 7.6.

It is seen that the results of IWST changed only for one reference point: C-500, which changed designation from stable to unstable when the variance components estimation method changed from REML to GML with the inverted gamma prior.



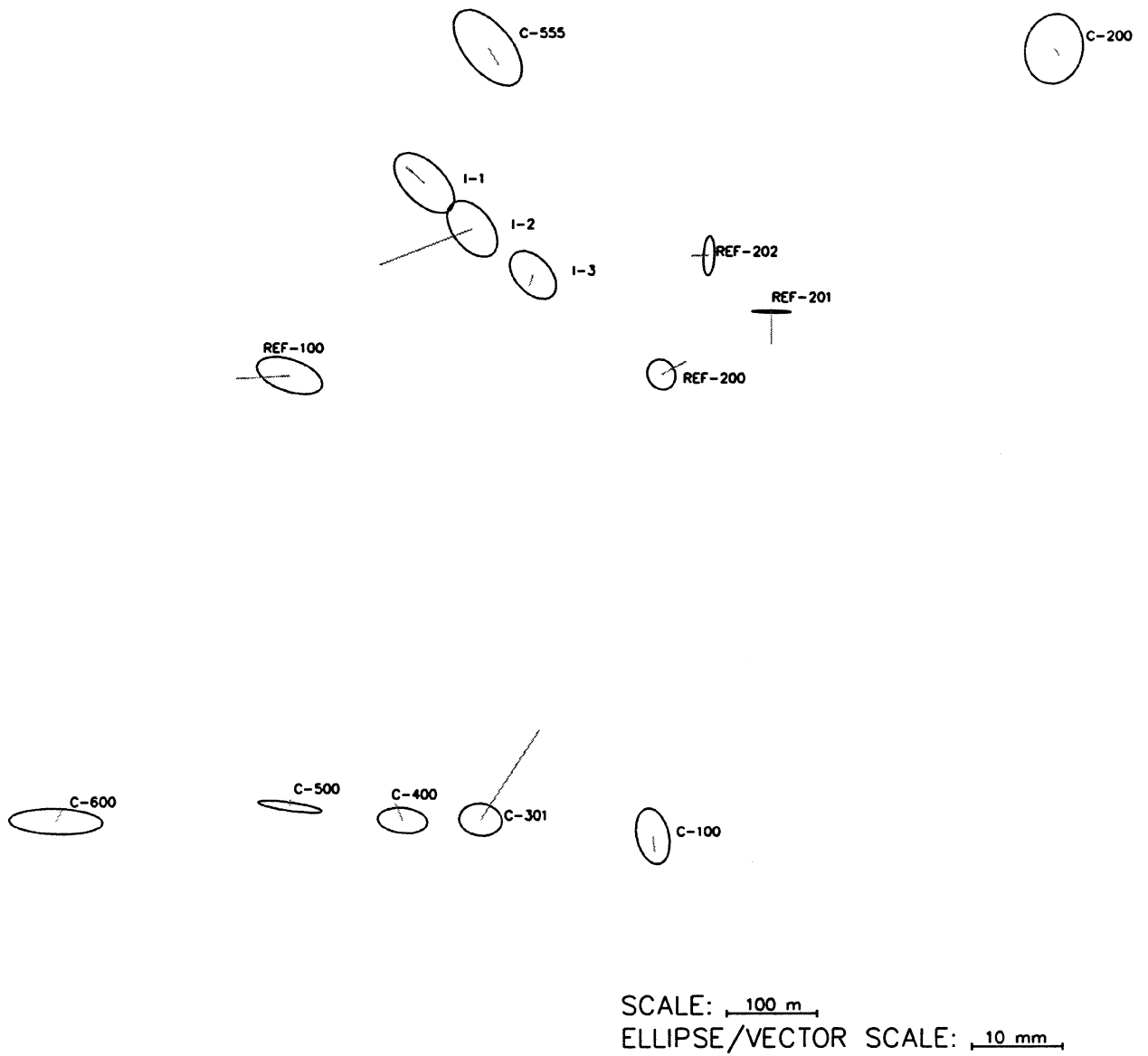


Figure 7.5 Displacement Field and Error Ellipses (95%):  
 3 Components Model Used for the 1993 Data;  
 GML with Inverted Gamma Prior Estimation Method,  
 Ratio of the Standard Deviations of the Priors to the Prior Values = 2

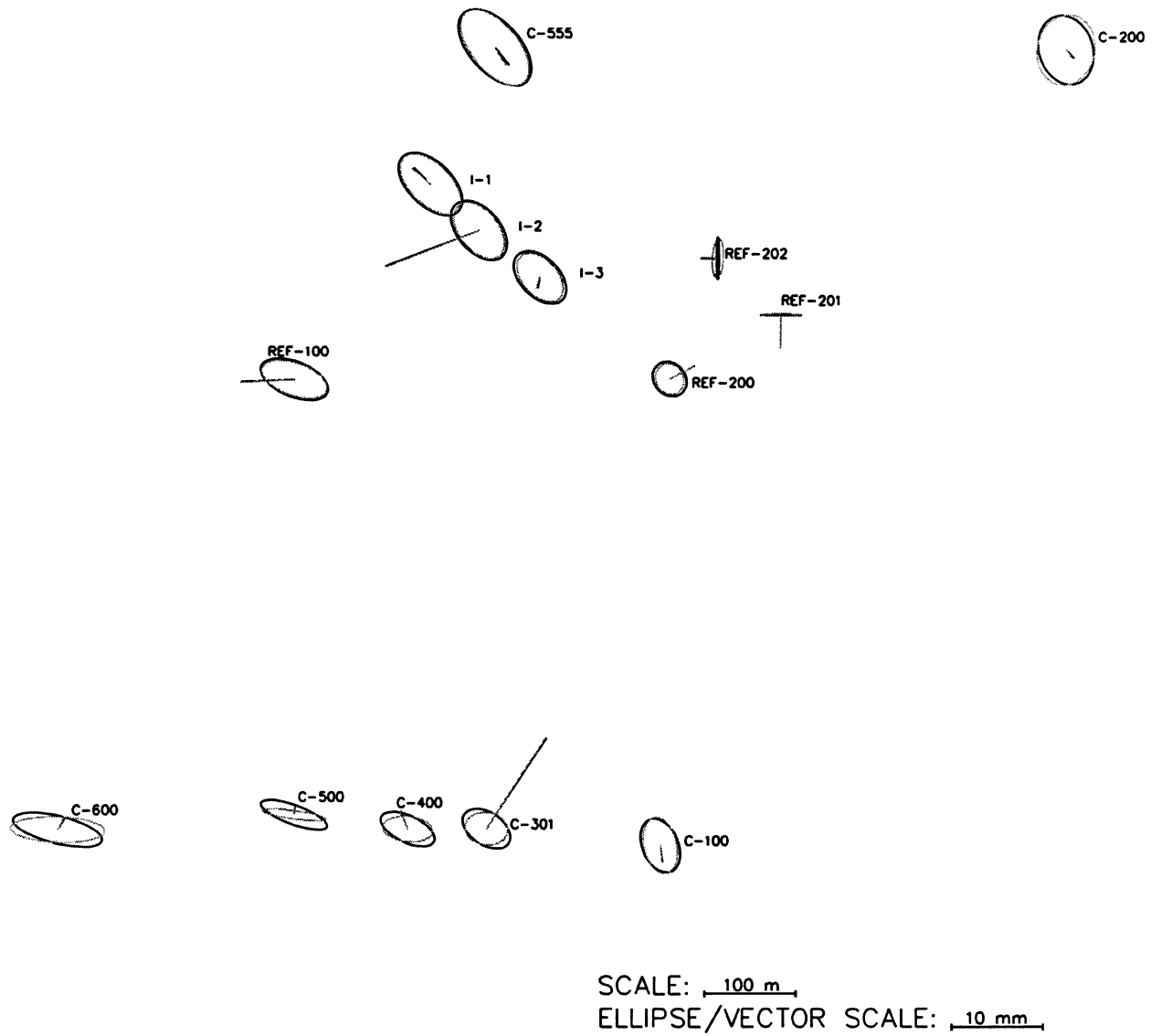


Figure 7.6 Displacement Field and Error Ellipses (95%):  
 REML versus GML with Inverted Gamma Prior

### ***7.2.3 Influence of the Amount of Prior Information***

The way the amount of prior information affects the results of the spatial deformation trend analysis is investigated here. Again, the 3 parameter variance components model was used for the 1991 data while either the 2 parameter or the 3 parameter model was utilized for the 1993 data set.

The variability in the amount of prior information had less impact on the results of IWST than all other analyzed factors. The results of IWST were the same for all ratios of the standard deviations of the priors to the prior values (means of the prior distribution) (ranging from 2 to 16), for both the GML with the inverted gamma prior and the dispersion-mean model with weighted constraints estimation methods.

The largest difference, noticeable mainly in the shape and size of the error ellipses of the transformed displacement components, was seen when the results of IWST were compared for the two cases when the dispersion-mean model with weighted constraints estimation method was used: in the first case with the ratios of the standard deviations of the priors to the prior values (means of the prior distribution) set to 2 (Figure 7.7); in the second case with the ratios set to 16 (Figure 7.8). The IWST results are compared in Figure 7.9.

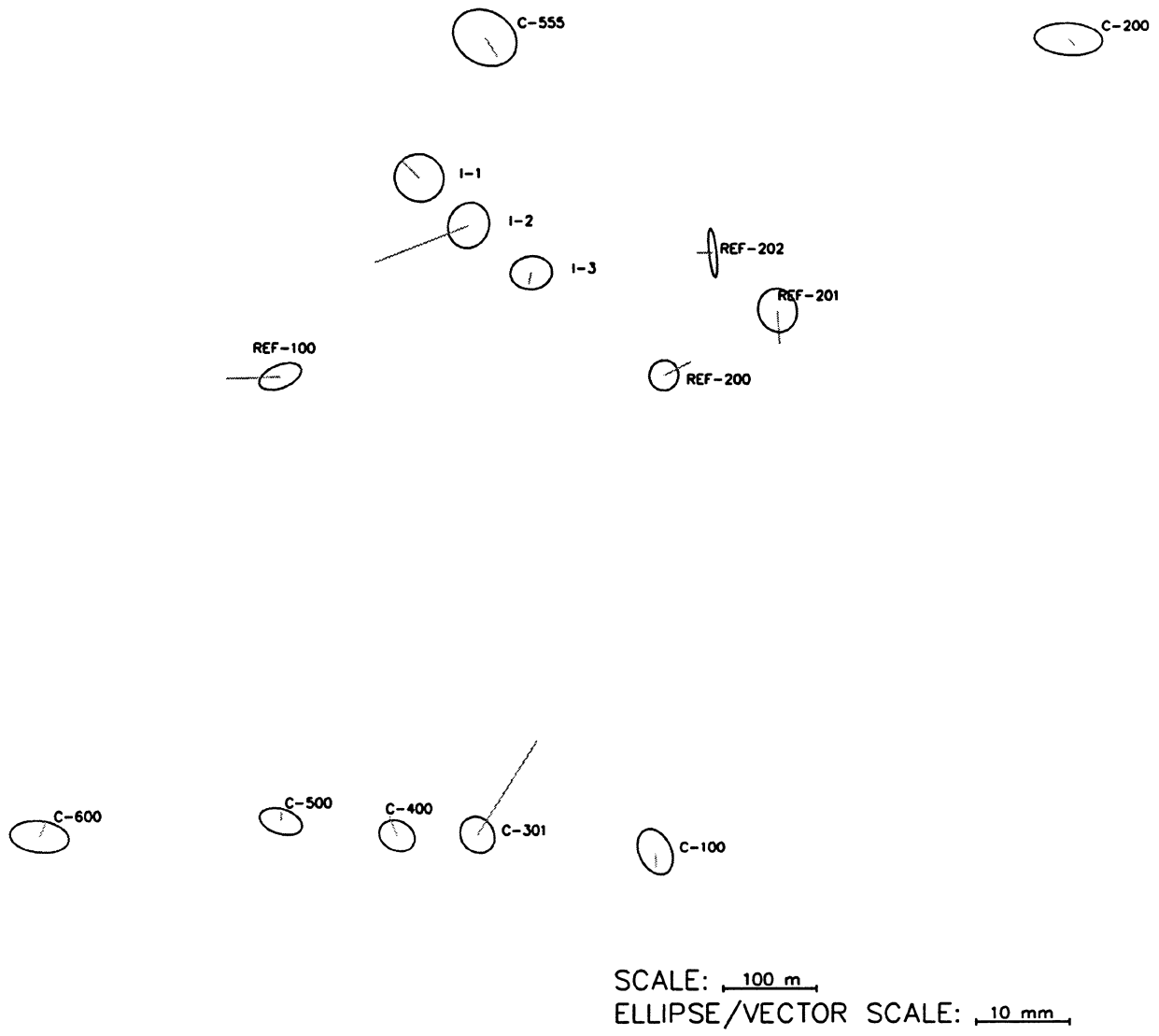


Figure 7.7 Displacement Field and Error Ellipses (95%):  
 3 Components Model Used for the 1993 Data,  
 Dispersion-Mean Model with Weighted Constraints,  
 Ratio of the Standard Deviations of the Priors to the Prior Values = 2

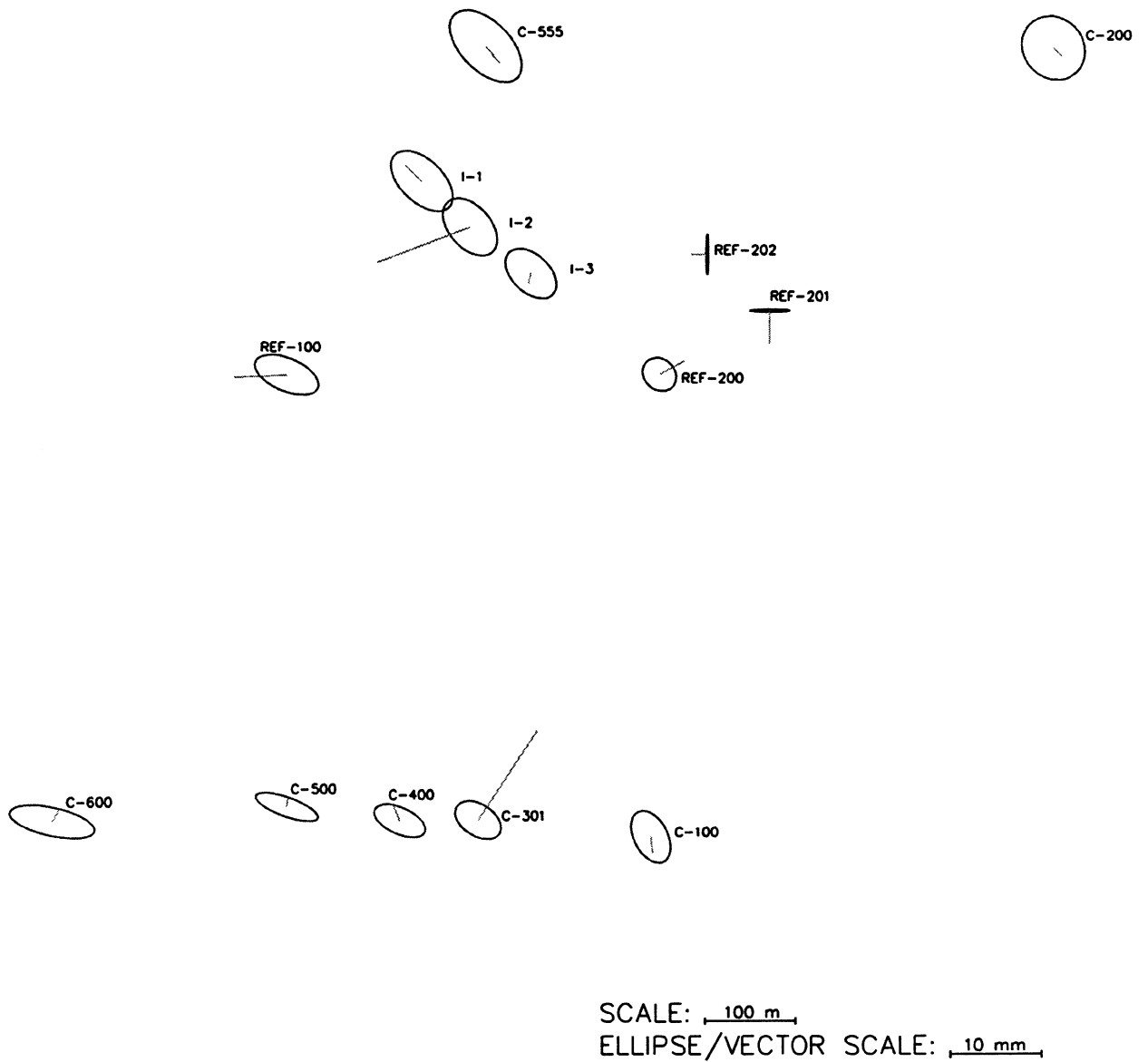


Figure 7.8 Displacement Field and Error Ellipses (95%):  
 3 Components Model Used for the 1993 Data,  
 Dispersion-Mean Model with Weighted Constraints,  
 Ratio of the Standard Deviations of the Priors to the Prior Values = 16

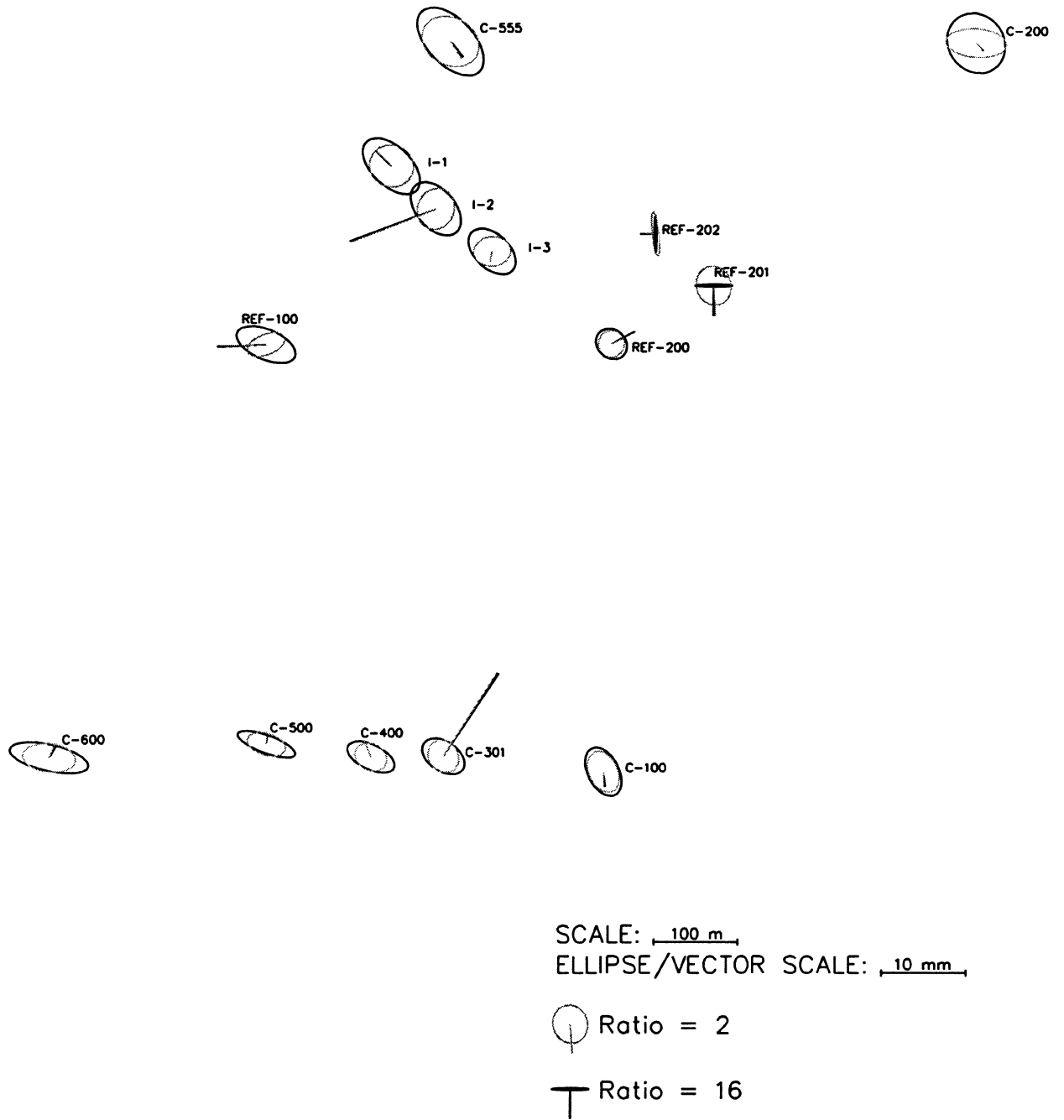


Figure 7.9 Displacement Field and Error Ellipses (95%):  
 Ratio of 2 versus Ratio of 16

#### ***7.2.4 Discussion of the Results***

As far as the influence of the estimation of variance components on the spatial deformation trend analysis process is concerned, the amount of prior information had the smallest effect from all the discussed factors. The change of the ratio of the standard deviations of the priors to the prior values (means of the prior distribution) appears to have had a pronounced effect only in the case of the dispersion-mean model with weighted constraints when using the 3 parameter model for the epoch 1993. In the case of the dispersion-mean model with weighted constraints, the effect of introduction of prior information decreased to almost zero with the increase of the ratio of the standard deviations of the priors to the prior values (means of the prior distribution).

The second in magnitude was the influence of the choice of the method of estimation. All four discussed estimation methods affected the results of the spatial deformation trend analysis in a very similar way.

The choice of the error model had by far the largest influence on the results of the spatial deformation trend analysis process, for all the methods.

As far as the influence of all three analyzed factors on the transformed displacements versus their influence on the elements of the transformed error ellipses is concerned, the confidence regions were in all cases affected much more strongly than the displacements themselves.

### **7.3 Analysis of Homogeneity of the Data: Mactaquac Monitoring Network (Campaign of 1989)**

In this section the campaign of 1989 is chosen to demonstrate the proposed approach to the analysis of homogeneity of the data, by means of the algorithm for detecting influential observations, developed in Chapter 5 of this thesis. The epoch of 1989 is chosen because of the problems encountered when estimating variance components with the 2 parameter model (eqn. 7.1), affecting all four estimation methods. For the REML method and the GML method with the noninformative prior, the estimation process did not produce any results as it diverged (see Subsection 7.1.2). Introduction of prior information by means of either the inverted gamma prior or the weighted constraints on the unknown variance components in the dispersion-mean model produced a solution. Convergence was, however, very slow and in the case of GML estimation with the inverted gamma prior the iterative process was very sensitive to the choice of the approximate starting values.

When the original data set was used, i.e., when no observations were removed, and the empirical, *a priori*, values of the two variance components (see Table 7.2) were used to compute the covariance matrix of observations, the least squares estimation process produced six observations flagged as outliers (see Table 7.3). All six flagged residuals failed the  $\tau$ -test by a small margin. As mentioned above if neither of the flagged observations were removed the REML iterative process diverged for the 2 parameter model.



Table 7.3 Rejected residuals (tau-max criterion used)

From	To	Residual [m]	Critical Point (at 95%)
REF-100	M-1	0.0019	0.0016
C-400	PR-1	0.0071	0.0068
I-3	M-1	0.0040	0.0035
M-1	TK-7	0.0020	0.0018
M-1	REF-100	0.0019	0.0016
TK-7	I-3	0.0029	0.0025

Since the REML iterative process diverged, it was extremely difficult to quantify the influence of each observation, as it changed drastically from one iterative step to another. From the two of the variance components the  $a_{\xi}^2$  term changed very little from one step to another remaining at around  $4 \text{ mm}^2$ , while the  $b_{\xi}^2$  term jumped between  $0.2 \text{ ppm}^2$  and  $1.2 \text{ ppm}^2$  diverging more with each iteration. As a result, the influence of each observation changed between even and odd steps. In the latter case, which resulted in the  $b_{\xi}^2$  component of about  $1.2 \text{ ppm}^2$ , distance observation from TK-1 to PR-1 had the largest positive effect on the  $b_{\xi}^2$  component and its removal increased the value of the  $b_{\xi}^2$  term by up to 60%. The six formal outliers had also a positive effect on the  $b_{\xi}^2$  component - ranging from 2% to 30%. In the former case, which resulted in the  $b_{\xi}^2$  component of about  $0.2 \text{ ppm}^2$ , also the same distance observation (from TK-1 to PR-1) had the largest positive effect on the  $b_{\xi}^2$  component. Removal of this observation

increased the value of the  $b_S^2$  components to about 1 ppm<sup>2</sup> (over 500% increase). The six formal outliers had also a pronounced positive effect on the  $b_S^2$  component - ranging from 17% to 215%. In all cases the variance ratio  $VR(i)$  (eqn. 5.41) remained practically insensitive to the removal of any single observation, showing no more than a few percent change.

In the next step the distance observation from TK-1 to PR-1 was removed and the REML estimation process was repeated. The process converged (albeit very slowly) to the following set of values:

$$a_S^2 = 3.85 \text{ mm}^2 \pm 0.63, \text{ and}$$

$$b_S^2 = 1.14 \text{ ppm}^2 \pm 1.75.$$

Influence of the six formal outliers on the estimation process is shown in Table 7.4. The largest positive influence on the  $b_S^2$  component was exerted by the distance from C-200 to C-600 (46%). This may be explained by the fact that there was a significant discrepancy between the distance from C-200 to C-600 and the distance measured in the opposite direction: from C-600 to C-200, of over 5 mm. The distance from C-600 to C-200 had on the other hand a large negative influence on the  $b_S^2$  component (-183%). The  $a_S^2$  component was largely unaffected by any single observation. The variance ratio  $VR(i)$  remained again practically insensitive to the removal of any single observation.

In the final step, in addition to the distance observation from TK-1 to PR-1, all formal outliers were removed from the data and the REML estimation process was repeated. The estimation process converged in just over 10 iterations to the following set of values:

$$a_{\zeta}^2 = 3.10 \text{ mm}^2 \pm 0.57, \text{ and}$$

$$b_{\zeta}^2 = 1.95 \text{ ppm}^2 \pm 1.69.$$

Observation from M-1 to I-3 had the largest positive influence on the  $b_{\zeta}^2$  component (26%). A large positive influence on the  $b_{\zeta}^2$  component was also exerted by the distance from C-200 to C-600 (22%). The distance from C-600 to C-200 had again the largest negative influence on the  $b_{\zeta}^2$  component (-79%). As before, the  $a_{\zeta}^2$  component was not significantly affected by any single observation. The variance ratio  $VR(l_i)$  was again insensitive to the removal of any single observation.

Table 7.4 Influence of formal outliers on the  $b_{\zeta}^2$  term

From	To	Influence on $b_{\zeta}^2$ Term [%]
REF-100	M-1	39.6
C-400	PR-1	-0.1
I-3	M-1	23.1
M-1	TK-7	6.0
M-1	REF-100	39.6
TK-7	I-3	5.0

If all the formal outliers were initially removed, then the REML iterative process converged, albeit very slowly (over 40 iterations were required), to the following set of values:

$$a_{\xi}^2 = 3.43 \text{ mm}^2 \pm 0.60, \text{ and}$$

$$b_{\xi}^2 = 1.45 \text{ ppm}^2 \pm 1.70.$$

The distance from TK-1 to PR-1 had the largest positive effect on the  $b_{\xi}^2$  component; its removal increased the value of  $b_{\xi}^2$  by 58%. Observations M-1 — I-3 and C-200 — C-600 had also a very large positive influence on the  $b_{\xi}^2$  component (33% and 22% respectively). The distance C-600 - C-200 had the largest negative influence on the  $b_{\xi}^2$  component (-129%). Neither the  $a_{\xi}^2$  component nor the variance ratio  $VR(l_i)$  was significantly affected by any single observation. The results of the estimation could not, however, be accepted as final, as the  $b_{\xi}^2$  component was smaller than its standard deviation and the convergence was very slow. Removal of the distance observation exerting the largest positive influence on the  $b_{\xi}^2$  term resulted in the estimation process of step 3, which suffered neither of the aforementioned deficiencies.

It should be noted that the efficiency of all computations described in this section was significantly improved with the help of the numerical algorithm for recomputation of the  $\mathbf{R}_{\hat{\theta}}$  matrix - derived in Chapter 5 - making the process of detection of influential observations computationally feasible.

## CHAPTER 8

### SUMMARY, CONCLUSIONS AND RECOMMENDATIONS

#### **8.1 Recommended Strategy for Estimation of Variance or Variance-Covariance Components for Geodetic Observations**

In this chapter a strategy for the estimation of variance or variance-covariance components, advocated by the author, is presented. The following two cases are considered.

(1) No prior information about the unknown variance or variance-covariance components is available.

If the general mixed model (eqn. 2.1) is used, then the Restricted Maximum Likelihood (REML) estimation method should be chosen. If the variance components model of eqn.(3.1) is used then either REML or the Generalized Maximum Likelihood (GML) method with the noninformative prior is recommended. From these two methods the REML method is substantially more efficient computationally. The REML method produces the estimates of variance or variance-covariance components that are unbiased and invariant to the choice of the vector of location parameters. The REML estimates are identical to the Iterated BIQUE and the IMINQE(U,I) estimates. The GML estimates are

invariant to the choice of the vector of location parameters. They are, however, no longer unbiased.

(2) Some prior information about the unknown variance or variance-covariance components is available.

In such case either the GML method with the inverted gamma prior or the dispersion-mean model with weighted constraints can be used. The GML method with the inverted gamma prior (applicable only to the variance components model of eqn.(3.1)) takes into account the fact that the variance components are, by definition, positive. The GML (with the inverted gamma prior) estimates do not, however, approach either the REML estimates or the GML (with the noninformative prior) estimates with the increase of the variances of the priors. The dispersion-mean model with weighted constraints is not recommended, as the prior probability density function implied by it, is improper for either the variance or the variance-covariance components models.

## **8.2 Summary of Results and Conclusions**

As far as numerical procedures are concerned, many of the statistical methods for estimation of variance or variance-covariance components for unbalanced data are, under certain restrictions, equivalent. The following relationships exist between these methods:

1. MINQE(U,I) is numerically equivalent to BIQUE, to MIVQUE under the restriction of normality of observations to AUE (under the assumption that the initial,

approximate values of the variance-covariance components are proportional to their true values), to the Helmert-type estimation (for variance components only), and to the first step REML estimation.

2. MINQE(I) is numerically equivalent to the first step ML estimation.
3. REML estimation is numerically equivalent to IMINQE(U,I), to iterated BIQUE, IAUE (which may be regarded as one of the algorithms for solution of REML), and to the iterated Helmert-type estimation (for variance components only).
4. ML estimation is numerically equivalent to IMINQE(I).

The general ANOVA methods are non-unique and lack optimal properties. As a result they should be regarded as inferior to other estimation methods.

While the computational approach of the GML methods is similar both in principle and in scope to that of the maximum likelihood methods (ML and REML), the other two Bayesian methods, i.e., the Bayes estimation and the interval estimation, employ completely different computational procedures. Both the Bayes estimation and the interval estimation require extensive computations and are essentially impractical for cases involving more than a few components.

Introduction of prior information about the unknown variance components into the estimation process can be efficiently accomplished by means of the Generalized Maximum Likelihood (GML) method. If the inverted gamma distribution function is chosen for the

prior, then the GML estimating equations, resulting from application of the Fisher scoring method to the posterior probability density function, constitute a straightforward extension of the REML estimating equations. The estimated components and their asymptotic standard deviations do not, in general, approach the values of their REML estimated counterparts with the increase of the ratio of the standard deviations of the priors to the prior values of the unknown components (means of the prior distribution). It is found that introduction of prior information may cause the estimation process to converge in cases where the REML method diverges.

Prior information can also be introduced by means of the weighted constraints on the unknown variance-covariance components in the dispersion-mean model. Application of the least squares methodology to the resulting dispersion-mean model with weighted constraints yields estimating equations that are equivalent to the GML scoring equations with the multivariate normal prior. It should be noted that the multivariate normal prior is not a proper choice for either the variance or variance-covariance components as it assigns positive probabilities to negative definite covariance matrices.

When no prior information is available, the GML approach can be used with the noninformative prior. The results of the GML estimation with the noninformative prior are similar to those obtained with the REML method. As far as the rate of convergence is concerned, both methods give comparable results. While the GML estimated components



are consistently larger, their significance is in general smaller than that of the REML estimated counterparts.

Using the estimated covariance matrix of observations, in place of the true covariance matrix, has no effect on the covariance matrix of the estimated vector of location parameters, up to the first order terms.

As far as the influence of various aspects of the estimation of variance components on the results of the spatial deformation trend analysis process is concerned, the choice of the error model has by far the largest impact. The effect of prior information and the choice of the estimation method is much smaller.

The effectiveness of the proposed procedure for the detection of influential observations is demonstrated in the thesis. It is shown that observations that are not flagged as outliers by the  $\tau$ -test used in the least squares adjustment process may have very pronounced influence on the results of the variance components estimation.

### **8.3 Recommendations for Related Future Research**

It is recommended that further research be performed in the following areas:

1. The effect of using the estimated covariance matrix of observations, in place of the true covariance matrix, on the accuracy of the estimated location parameters was discussed

in this thesis. It is recommended that the influence of neglecting the errors of the estimated variance-covariance components be investigated.

2. The principle of GML estimation was applied in this research to the variance components model, using either the inverted gamma prior or the noninformative prior. It is recommended that applicability of the GML principle to the more general variance-covariance components models, with other choices of priors, be investigated.
3. It is recommended that the GML estimating equations, derived for the cases involving the inverted gamma and the noninformative priors, be tested with other types of variance components models, using other types of geodetic data.
4. The choice of the error model had by far the largest influence, from all the analyzed aspects of the process of estimation of variance components, on the results of spatial deformation trend analysis. It is, thus, recommended that more research be done in the area of proper formulation and statistical testing of the error model.
5. The procedure for the detection of influential observations, proposed in this research, is limited to the variance components model. It is recommended that possibility of extending this procedure to the more involved variance-covariance components models be investigated.
6. The procedure for the detection of influential observations should be further tested with the data sets involving more than one type of observable and more complicated variance components models.
7. Throughout this research it was assumed that the asymptotic covariance matrix of the estimated variance or variance-covariance components adequately represented the error

structure of the estimated components. Since, especially for the small sample size, such assumption may not always be justified it is recommended that more research be done in the area of interval estimation and statistical testing of the variance-covariance components.

8. Posterior normality of the posterior probability density function (with both the inverted gamma and the noninformative priors) was verified in this research for the Gauss-Markov model with the unknown variance factor. It is recommended that the asymptotic properties of of the posterior probability density function, for both choices of priors, are investigated for more general variance components models.

## REFERENCES

- Andrews, D.F. and Pregibon, D. (1978). "Finding Outliers that Matter." *Journal of the Royal Statistical Society*, (B), 40, 85-93.
- Belsley, D.A., Kuh, E. and Welsch, R.E. (1980). *Regression Diagnostics: Identifying Influential Data Points and Sources of Collinearity*. John Wiley & Sons, New York.
- Bernardo, J.M and Smith, A.F.M. (1993). *Bayesian Theory*. John Wiley & Sons, Chichester.
- Blachut, T., Chrzanowski, A. and Saastamoinen, J. (1979). *Urban Surveying and Mapping*. New York: Springer-Verlag.
- Bossler, J.D. (1972). *Bayesian Inference in Geodesy*. Ph.D. Dissertation. Ohio State University.
- Caspary, W.F. (1987). *Concepts of Network and Deformation Analysis*. The University of New South Wales, School of Surveying, Monograph No. 21, Kensington (Australia)
- Chatterjee, S. and Hadi, A.S. (1988). *Sensitivity Analysis in Linear Regression*. John Wiley & Sons, New York.
- Chen, Y.Q. (1983). *Analysis of Deformation Surveys - a Generalized Method*. Ph.D. Dissertation, Department of Surveying Engineering Technical Report No. 94, University of New Brunswick, Fredericton.
- Chen, Y.Q. and Chrzanowski, A. (1985). "Assessment of Levelling Measurements Using the Theory of MINQE." *Proceedings of the Third International Symposium on the North American Vertical Datum*, held in Rockville, Maryland, U.S.A. on 21-26 April 1985.
- Chen, Y.Q., Chrzanowski, A. and Kavouras, M. (1990a). "Assessment of Observations Using Minimum Norm Quadratic Unbiased Estimation (MINQUE)." *CISM Journal ACSGC*, Vol. 44, No. 1, 9-18.
- Chen, Y.Q., Chrzanowski, A. and Secord, J. (1990b). "A Strategy for the Analysis of the Stability of Reference Points in Deformation Surveys." *CISM Journal ACSGC*, Vol. 44, No. 2, 141-149.

- Chrzanowski, A. (1974). "Pre-analysis and Design of Surveying Projects." *Northpoint*, Ontario Association of Certified Survey Technicians and Technologists, Vol. XI, Nos. 1, 2.
- Chrzanowski, A. (1975). "Pre-analysis and Design of Surveying Projects." *Northpoint*, Ontario Association of Certified Survey Technicians and Technologists, Vol. XII, No. 1.
- Chrzanowski, A. (1981). "A Comparison of Different Approaches into the Analysis of Deformation Measurements." *Proceedings of the FIG XVI International Congress*, Montreux, Switzerland.
- Chrzanowski, A., Chen, Y.Q. and Secord, J. (1986). "Geometrical Analysis of Deformation Surveys." *Proceedings of the Deformation Measurements Workshop*, MIT, Cambridge, Mass., 170 - 206.
- Chrzanowski, A. and Secord, J. (1987). "An Integrated Analysis of Deformation Measurements at The Mactaquac Generating Station." Final Report 1987 12, Prepared for The New Brunswick Electric Power Commission. Department of Surveying Engineering, University of New Brunswick.
- Chrzanowski, A., Secord, J. and Szostak-Chrzanowski, A. (1989). "The 1988 Integrated Analysis of Deformation Measurements at The Mactaquac Generating Station." Final Report, Prepared for The New Brunswick Electric Power Commission. Department of Surveying Engineering, University of New Brunswick.
- Chrzanowski, A. and Secord, J. (1990). "The 1989 Integrated Analysis of Deformation Measurements at The Mactaquac Generating Station." Final Report, Prepared for The New Brunswick Electric Power Commission. Department of Surveying Engineering, University of New Brunswick.
- Chrzanowski, A., Caissy, M., Grodecki, J. and Secord, J. (1994). "Software Development and Training for Geometrical Deformation Analysis." Final Report to the Geodetic Survey Division of the Department of Natural Resources. Engineering Surveys Research Group, Department of Geodesy and Geomatics Engineering, University of New Brunswick.
- Chrzanowski, A. (1995). Personal Communication.
- Cook, R.D. and Weisberg, S. (1980). "Characterization of an Empirical Influence Function for Detecting Influential Cases in Regression." *Technometrics*, Vol. 22, 495-508.

- Förstner, W. (1979a). "Konvergenzbeschleunigung bei der a posteriori Varianzschätzung." *Zeitschrift für Vermessungswesen*, No 4/1979, 149-156
- Förstner, W. (1979b). "Ein Verfahren zur Schätzung von Varianz - und Kovarianzkomponenten." *Allgemeine Vermessungsnachrichten*, 11-12/1979, 446-453
- Fox, J. (1991). *Regression Diagnostics: An Introduction*. Sage University Paper Series on Quantitative Applications in the Social Sciences, 07--79. Newbury Park, CA: Sage.
- Grafarend, E. (1978). "Schätzung von Varianz und Kovarianz der Beobachtungen in geodätischen Ausgleichungsmodellen." *Allgemeine Vermessungsnachrichten*, 2, 41-49.
- Grafarend, E., Kleusberg, A. and Schaffrin, B. (1980). "An Introduction to the Variance-Covariance Components Estimation of Helmert Type." *Zeitschrift für Vermessungswesen*, 4, 161-180.
- Grafarend, E. and Kleusberg, A. (1980). "Expectation and Variance Component Estimation of Multivariate Gyrotheodolite Observations I." *Allgemeine Vermessungsnachrichten*, 87, 129-137.
- Grafarend, E. (1984). "Variance-Covariance-Component Estimation of Helmert Type in the Gauß-Helmert Model." *Zeitschrift für Vermessungswesen*, 1, 34-44.
- Gupta, R.D. (1995). Personal Communication.
- Hampel, F.R. (1968). *Contributions to the Theory of Robust Estimation*. Ph.D. Thesis, University of California, Berkeley.
- Hampel, F.R. (1974). "The Influence Curve and Its Role in Robust Estimation." *Journal of the American Statistical Association*, 69, 383-393.
- Hampel, F.R., Ronchetti, E.M., Rousseuw, P.J. and Stakel, W.A. (1986). *Robust Statistics The Approach Based on Influence Functions*. John Wiley & Sons, New York.
- Hartley, H.O. and Rao, J.N.K. (1967). "Maximum Likelihood Estimation for the Mixed Analysis of Variance Model." *Biometrika*, 54, 93-108.
- Harville, D.A. (1977). "Maximum Likelihood Approaches to Variance Component Estimation and to Related Problems." *Journal of the American Statistical Association*, 72, 320-340.

- Helmert, F.R. (1924). *Die Ausgleichsrechnung nach der Methode der kleinsten Quadrate*. Leipzig, 1924.
- Horn, S.D., Horn, R.A. and Duncan, D.B. (1975). "Estimating Heteroscedastic Variances in Linear Models." *Journal of the American Statistical Association*, 70, 380-385.
- Horn, S.D. and Horn, R.A. (1975). "Comparison of Estimators of Heteroscedastic Variances in Linear Models." *Journal of the American Statistical Association*, 70, 872-879.
- Huber, P.J. (1981). *Robust Statistics*. John Wiley & Sons, New York.
- Hultquist, R. (1985). "Variance Components." in: *Encyclopedia of Statistical Sciences*, Vol.9, S. Kotz and N.L. Johnson, Editors-in-Chief, John Wiley & Sons, New York.
- Jennrich, R.I. and Sampson, P.F. (1976). "Newton-Raphson and Related Algorithms for Maximum Likelihood Variance Component Estimation." *Technometrics*, Vol. 18, No.1, 11-17.
- Kackar, R.N. and Harville, D.A. (1981). "Unbiasedness of Two-stage Estimation and Prediction Procedures for Mixed Linear Models." *Communications in Statistics - Theory and Methods*, 10, 1249-1261.
- Kelly, K.M. (1991). *Weight Estimation in Geodetic Levelling Using Variance Components Derived from Analysis of Variance*. Master of Applied Science Thesis, Department of Civil Engineering, University of Toronto.
- Kelm, R. (1978). "Ist die Varianzschätzung nach Helemert MINQUE." *Allgemeine Vermessungsnachrichten*, 2, 49-54.
- Khuri, A.I., and Sahai, H. (1985). "Variance Components Analysis: A Selective Survey." *International Statistical Reviews*, 53, 279-300.
- Kleffe, J. (1977). "Optimal Estimation of Variance Components - a Survey." *Sankhyâ: The Indian Journal of Statistics*, vol. 39, Series B, 211-244.
- Knight, W. (1995). Personal Communication.
- Knight, W. (1997). Personal Communication.
- Koch, K.R. (1981). "Varianz und Kovarianzkomponentenschätzung für Streckenmessungen auf Eichlinien." *Allgemeine Vermessungsnachrichten*, 4/1981, 125-132.

- Koch, K.R. (1986). "Maximum Likelihood Estimate of Variance Components." *Bulletin Géodésique*, 60, 329-338.
- Koch, K.R. (1987). "Bayesian Inference for Variance Components." *Manuscripta Geodaetica*, 12, 309-313.
- Koch, K.R. (1988). "Bayesian Statistics for Variance Components with Informative and Noninformative Priors." *Manuscripta Geodaetica*, 13, 370-373.
- Koch, K.R. (1990). *Bayesian Inference with Geodetic Applications*. Lecture Notes in Earth Sciences, Vol. 31, Springer, Berlin Heidelberg New York Tokyo.
- Kubik, K. (1967). "Schätzung der Gewichte der Fehlergleichungen beim Ausgleichsproblem." *Allgemeine Vermessungsnachrichten*, 92(1967)5, 173-178.
- Kubik, K. (1970). "The Estimation of the Weights of Measured Quantities Within the Method of Least Squares." *Bulletin Géodésique*, 95, 21-40.
- LaMotte, L.R. (1973). "Quadratic Estimation of Variance Components." *Biometrics*, Vol.29, 311-330.
- Lindley, D.V. and Smith, A.F.M. (1972). "Bayes Estimates for the Linear Model." *Journal of the Royal Statistical Society*, Series B, 1-18.
- Lucas, J.R., Bengston, J.M. and Zilkoski, D.B. (1985). "Estimation of Variance Components in Leveling Using Iterated Almost Unbiased Estimation." *Proceedings of the Third International Symposium on the North American Vertical Datum*, held in Rockville, Maryland, USA, on 21-26 April 1985.
- Magnus, J.R. and Neudecker, H. (1988). *Matrix Differential Calculus with Applications in Statistics and Econometrics*. John Wiley & Sons, New York.
- Merminod, B. and Rizos, C. (1988). "A Surveyor's Perspective of Kalman Filtering." *Australian Journal of Geodesy Photogrammetry and Surveying*. No. 49, December 1988, 1-38.
- Miller, J.J. (1977). "Asymptotic Properties of Maximum Likelihood Estimates in the Mixed Model of the Analysis of Variance." *The Annals of Statistics*, 5, 746-762
- Neuwirth, E. (1984). "Sensitivity of Linear Models with Respect to the Covariance Matrix." "Lecture Notes in Statistics, Springer-Verlag, *Linear Statistical Inference* (Ed. T. Calinski and W. Klonecki).



- Oswald, W. (1992). *Zur kombinierten Ausgleichung heterogener Beobachtungen in hybriden Netzen*. Schriftenreihe Wissenschaftlicher Studiengang Vermessungswesen, Universität der Bundeswehr München, Heft 44, Neubiberg.
- Ou, Z. (1989). "Estimation of Variance and Covariance Components." *Bulletin Géodésique*, 63, 139-148.
- Ou, Z. (1991). "Approximative Bayes Estimation for Variance Components." *Manuscripta Geodaetica*, 16, 168-172.
- Patterson, H.D. and Thompson, R. (1971). "Recovery of Inter-block Information when Block Sizes Are Unequal." *Biometrika*, 58, 545-554.
- Pukelsheim, F. (1974). *Schätzen von Mittelwert und Streuungsmatrix in Gauss-Markov Modellen*. Diplomarbeit, Freiburg im Breisgau, Germany.
- Rao, C.R. (1971a). "Estimation of Variance and Covariance Components - MINQUE Theory." *Journal of Multivariate Analysis*, 1, 257-275.
- Rao, C.R. (1971b). "Minimum Variance Quadratic Unbiased Estimation of Variance Components." *Journal of Multivariate Analysis*, 1, 445-456.
- Rao, C.R. (1979). "MINQE Theory and its Relation to ML and MML Estimation of Variance Components." *Sankhyâ: The Indian Journal of Statistics*, vol. 41, Series B, 138-153.
- Rao, C.R., and Kleffe, J. (1988). *Estimation of Variance Components and Applications*. North-Holland, Amsterdam, New York.
- Read, C.B. (1985). "Mean Squared Error." in: *Encyclopedia of Statistical Sciences*, Vol.5, S. Kotz and N.L. Johnson, Editors-in-Chief, John Wiley & Sons, New York.
- Sahai, H. (1979). "A Bibliography on Variance Components." *International Statistical Reviews*, 47, 177-222.
- Sahai, H., Khuri, A.I. and Kapadia, C.H. (1985). "A Second Bibliography on Variance Components." *Communications in Statistics A: Theory and Methods*, 14, 63-115.
- Sahin, M., Cross, P.A. and Sellers, P.C. (1992). "Variance Component Estimation Applied to Satellite Laser Ranging." *Bulletin Géodésique*, 66, 284-285.
- Schaffrin, B. (1981). "Best Invariant Covariance Component Estimator and its Application to the Generalized Multivariate Adjustment of Heterogeneous Deformation Observations." *Bulletin Géodésique*, 55, 73-85.

- Schaffrin, B. (1983). *Varianz-Kovarianz-Komponenten-Schätzung bei der Ausgleichung heterogener Wiederholungsmessungen*. Deutsche Geodätische Kommission Reihe C, Heft 282, München.
- Scholz, F.W. (1985) "Maximum Likelihood Estimation." in: *Encyclopedia of Statistical Sciences*, Vol.5, S. Kotz and N.L. Johnson, Editors-in-Chief, John Wiley & Sons, New York.
- Searle, S.R. (1979). *Notes on Variance Component Estimation: A Detailed Account of Maximum Likelihood and Kindred Methodology*. Technical Report BU-673-M, Biometrics Unit, Cornell University, Ithaca, New York.
- Searle, S.R. and Henderson, H.V. (1979). "Dispersion Matrices for Variance Components Models." *Journal of the American Statistical Association*, 74, 465-470.
- Searle, S.R. (1982). *Matrix Algebra Useful for Statistics*. John Wiley & Sons, New York.
- Searle, S.R. (1987). *Linear Models for Unbalanced Data*. John Wiley & Sons, New York.
- Searle, S.R., Casella, G. and McCulloch, C.E. (1992). *Variance Components*. John Wiley & Sons, New York.
- Silvey, S.D. (1975). *Statistical Inference*. Lowe & Brydone (Printers) Ltd, Thetford, Norfolk.
- Sjöberg, L.E. (1980). "Maximum Likelihood-Estimation of Covariances and Weighted Mean with Application to Repeated EDM-Observations." *Zeitschrift für Vermessungswesen*, No 10/1980, 284-295.
- Sjöberg, L.E. (1983a). "Unbiased Estimation of Variance-Covariance Components in Condition Adjustment." In the Commemorative volume on the occasion of Erik Tengström's 70th birthday. Department of Geodesy, Report No. 19, Uppsala 1983 (Ed. A.J. Anderson and L. Aardom).
- Sjöberg, L.E. (1983b). "Unbiased Estimation of Variance-Covariance Components in Condition Adjustment with Unknowns." *Zeitschrift für Vermessungswesen*, No 9/1983, 382-387.
- Sjöberg, L.E. (1985). "Adjustment and Variance-Covariance Components Estimation with a Singular Covariance Matrix." *Zeitschrift für Vermessungswesen*, No 4/1985, 145-151.

- Swallow, W.H. and Searle, S.R. (1978). "Minimum Variance Quadratic Unbiased Estimation (MIVQUE) of Variance Components." *Technometrics*, Vol. 20, No.3, 265-272.
- Swallow, W.H. and Monahan, J.F. (1984). "Monte Carlo Comparison of ANOVA, MIVQUE, REML, and ML Estimators of Variance Components." *Technometrics*, Vol. 26, No.1, 47-57.
- Townsend, E.C. and Searle, S.R. (1971). "Best Quadratic Unbiased Estimation of Variance Components from Unbalanced Data in the 1-Way Classification." *Biometrics*, Vol.27, 643-657.
- Vaníček, P. and Krakiwsky, E. (1986). *Geodesy: The Concepts*. Second Edition, Elsevier Science, Amsterdam.
- Wells, D.E (1990). *Functional Analysis and Adjustments*. Unpublished Lecture Notes, Department of Surveying Engineering, University of New Brunswick, Fredericton.
- Welsch, W. (1978). "A posteriori Varianzschtzung nach HELMERT." *Allgemeine Vermessungsnachrichten*, 2, 55-63.
- Yu, Z. (1992). "A Generalization Theory of Estimation of Variance-Covariance Components." *Manuscripta Geodaetica*, 17, 295-301.

## APPENDIX A

### NOTATION

<b>A</b>	first order design matrix
<b><math>C_{\theta}</math></b>	covariance matrix of observations (function of $\theta$ )
<b>D</b>	datum equations matrix
<b>d</b>	vector of displacements components
<b>E(.)</b>	expectation
<b>H</b>	Helmert transformation matrix
<b>IF(.)</b>	influence function
<b><math>J_M</math></b>	ML Fisher information matrix
<b><math>J_R</math></b>	REML Fisher information matrix
<b>l</b>	vector of observations
<b><math>L_M</math></b>	ML likelihood function
<b><math>l_M</math></b>	ML log-likelihood function
<b><math>L_R</math></b>	REML likelihood function
<b><math>l_R</math></b>	REML log-likelihood function
<b>MSE(.)</b>	mean squared error
<b>p</b>	probability density function

<b>Q</b>	cofactor matrix
<b>S</b>	similarity transformation matrix
<b>tr(.)</b>	trace
<b>v</b>	vector of residuals
<b>var(.)</b>	variance
<b>vec</b>	matrix operator stacking the columns of a matrix one under the other to form a single column
<b>x</b>	vector of unknown location parameters
<b><math>\theta = (\theta_1, \dots, \theta_r)</math></b>	vector of variance-covariance (variance) components
<b><math>\otimes</math></b>	Kronecker product operator
<b><math>\left\{ \begin{matrix} m &amp; a_{ij} \end{matrix} \right\}_{i=1, j=1}^{p \quad q}</math></b>	matrix <b>A</b> of order $p \times q$ composed of the elements $a_{ij}$ , where $a_{ij}$ is the element that is in the $i$ -th row and $j$ -th column of <b>A</b>
<b><math>\left\{ \begin{matrix} m &amp; a_{ij} \end{matrix} \right\}_{i,j=1}^p</math></b>	square matrix of order $p \times p$ composed of the elements $a_{ij}$
<b><math>\left\{ \begin{matrix} d &amp; a_i \end{matrix} \right\}_{i=1}^p</math></b>	diagonal matrix of order $p \times p$ , where $a_i$ denotes the $i$ -th element on the diagonal
<b><math>\left\{ \begin{matrix} c &amp; u_i \end{matrix} \right\}_{i=1}^t</math></b>	column vector <b>u</b> of $t$ elements, where $u_i$ denotes the $i$ -th element of <b>u</b>
<b><math>\left\{ \begin{matrix} r &amp; u_i \end{matrix} \right\}_{i=1}^t</math></b>	row vector <b>u<sup>T</sup></b> of $t$ elements, where $u_i$ denotes the $i$ -th element of <b>u<sup>T</sup></b>

## APPENDIX B

### ASYMPTOTIC PROPERTIES OF GML ESTIMATORS

In Chapter 2 (Subsection 2.2.1) the Hartley-Rao conditions [Hartley and Rao, 1967] for establishing asymptotic properties of ML estimators of the variance components are given. Under these assumptions Hartley and Rao [1967] prove that if the MLE of the vector of variance components provides the global maximum of the likelihood, it is weakly consistent, asymptotically efficient, and asymptotically normally distributed with the mean  $\theta_0$  (the true parameter) and the covariance matrix equal to the inverse of the information matrix computed at  $\theta_0$ .

In the GML approach prior knowledge about the unknown variance components, expressed by the prior probability density function, modifies the likelihood function, resulting in the posterior probability density function. The prior probability density function does not depend on the data while the likelihood function does. Intuitively, as the amount of data increases (e.g. , as in the Hartley-Rao model, cf. eqn. (2.3),  $m \rightarrow \infty$  and  $m_1 \rightarrow \infty$ ) prior information becomes less relevant and the GML estimator approaches the ML estimator.

Consider an experiment fulfilling all three Hartley-Rao conditions. The posterior probability density function is, in general case, given by the Bayes' Theorem (eqn. (2.65)) as:

$$p(\boldsymbol{\theta}|\mathbf{l}) = c p(\boldsymbol{\theta}) p(\mathbf{l}|\boldsymbol{\theta}). \quad (\text{B.1})$$

Taking natural logarithms of both sides we get:

$$\ln[p(\boldsymbol{\theta}|\mathbf{l})] = c_1 + \ln[p(\boldsymbol{\theta})] + \ln[p(\mathbf{l}|\boldsymbol{\theta})]. \quad (\text{B.2})$$

As mentioned before, in the GML approach we may consider a posterior probability density function to be a modified likelihood function. The maximum likelihood method is applied then to a posterior probability density function as if it were a likelihood function.

To prove consistency of GMLE one has to prove that, analogously to the Lemma 1 of Hartley and Rao [1967], for any fixed  $\boldsymbol{\theta}$  different from  $\boldsymbol{\theta}_0$  (true parameter)

$$\text{var}_o \left\{ m^{-1} \ln[p(\boldsymbol{\theta}|\mathbf{l})] \right\} = O(m^{-1}), \quad (\text{B.3})$$

where  $\text{var}_o$  denotes the variance when the true parameter  $\boldsymbol{\theta}_0$  holds.

Since the prior probability density function does not depend on the data, we get:

$$\text{var}_o \left\{ m^{-1} \ln[p(\boldsymbol{\theta}|\mathbf{l})] \right\} = \text{var}_o \left\{ m^{-1} \ln[p(\boldsymbol{\theta})] + m^{-1} \ln[p(\mathbf{l}|\boldsymbol{\theta})] \right\} = \text{var}_o \left\{ m^{-1} \ln[p(\mathbf{l}|\boldsymbol{\theta})] \right\}, \quad (\text{B.4})$$

which by Lemma 1 of Hartley and Rao [1967] is  $O(m^{-1})$ . The remainder of the proof of consistency of GMLE follows then the argument of Hartley and Rao [1967].

To prove asymptotic efficiency of GMLE one has to prove that the second derivatives of the natural logarithm of the posterior probability density function taken at the true value of the vector of parameters ( $\theta_0$ ) converge in probability to their expectations. To prove this hypothesis one only has to prove, as in [Hartley and Rao, 1967], that the variances of the second derivatives are  $O(m)$ . It follows from eqn. (B.2) that the second derivative of the natural logarithm of the posterior probability density function is equal to the sum of the second derivative of the natural logarithm of the prior and the second derivative of the log-likelihood function. Since the prior does not depend on the data, the variance of the second derivative of the natural logarithm of the posterior is the same as the variance of the log-likelihood function, which in turn is proven by Hartley and Rao [1967] to be  $O(m)$ . Applying the arguments of Hartley and Rao one can subsequently prove that the GMLE of the vector of variance components is asymptotically normally distributed with mean  $\theta_0$  and the covariance matrix equal to the inverse of the information matrix computed at  $\theta_0$ .

Instead of analyzing the asymptotic properties of the GML estimators of variance components one may try to formulate the asymptotic properties of the posterior probability density function.

Bernardo and Smith [1993] give the following development for establishing properties of the posterior probability density function as the number of observations,  $m$ , increases to



infinity. They expand the natural logarithms of the prior and the likelihood function about their respective maxima,  $\mu_P$  and  $\hat{\theta}_m$ , obtaining:

$$\ln[p(\boldsymbol{\theta})] = \ln[p(\mu_P)] + \frac{1}{2}(\boldsymbol{\theta} - \mu_P)^T \mathbf{H}_P(\boldsymbol{\theta} - \mu_P) + R_P, \text{ and} \quad (\text{B.5})$$

$$\ln[p(\mathbf{l}|\boldsymbol{\theta})] = \ln[p(\mathbf{l}|\hat{\theta}_m)] + \frac{1}{2}(\boldsymbol{\theta} - \hat{\theta}_m)^T \mathbf{H}(\hat{\theta}_m)(\boldsymbol{\theta} - \hat{\theta}_m) + R_m, \quad (\text{B.6})$$

where  $R_P$  and  $R_m$  are the remainder terms and

$$\mathbf{H}_P = \left\{ m \frac{\partial^2 \ln[p(\boldsymbol{\theta})]}{\partial \theta_i \partial \theta_j} \Big|_{\boldsymbol{\theta}=\mu_P} \right\}_{i,j=1}^r, \text{ and} \quad (\text{B.7})$$

$$\mathbf{H}(\hat{\theta}_m) = \left\{ m \frac{\partial^2 \ln[p(\mathbf{l}|\boldsymbol{\theta})]}{\partial \theta_i \partial \theta_j} \Big|_{\boldsymbol{\theta}=\hat{\theta}_m} \right\}_{i,j=1}^r, \quad (\text{B.8})$$

are the Hessian matrices of the natural logarithm of the prior and the log-likelihood function computed at their respective maxima. Assuming regularity conditions, ensuring that  $R_P$  and  $R_m$  remain small as the number of observations increases to infinity, Bernardo and Smith [1993] arrive at the following asymptotic form of the posterior probability density function:

$$\begin{aligned} p(\boldsymbol{\theta}|\mathbf{l}) &\propto \exp\left[\frac{1}{2}(\boldsymbol{\theta} - \mu_P)^T \mathbf{H}_P(\boldsymbol{\theta} - \mu_P) + \frac{1}{2}(\boldsymbol{\theta} - \hat{\theta}_m)^T \mathbf{H}(\hat{\theta}_m)(\boldsymbol{\theta} - \hat{\theta}_m)\right] \\ &\propto \exp\left[\frac{1}{2}(\boldsymbol{\theta} - \mu_m)^T \mathbf{H}_m(\boldsymbol{\theta} - \mu_m)\right], \end{aligned} \quad (\text{B.9})$$

with

$$\mathbf{H}_m = \mathbf{H}_P + \mathbf{H}(\hat{\theta}_m), \text{ and} \quad (\text{B.10})$$

$$\boldsymbol{\mu}_m = \mathbf{H}_m^{-1}[\mathbf{H}_P \boldsymbol{\mu}_P + \mathbf{H}(\hat{\boldsymbol{\theta}}_m) \hat{\boldsymbol{\theta}}_m]. \quad (\text{B.11})$$

Thus, the posterior probability density function will, subject to regularity conditions, converge to a multivariate normal distribution whose mean is a matrix weighted average of the prior modal value and the maximum likelihood estimate, and whose covariance matrix is a negative of an inverse of a sum of the Hessian matrices of the natural logarithms of the prior and the likelihood function computed at their respective maxima:

$$\text{var}(\boldsymbol{\mu}_m) = (-\mathbf{H}_m)^{-1} = -[\mathbf{H}_P + \mathbf{H}(\hat{\boldsymbol{\theta}}_m)]^{-1}. \quad (\text{B.12})$$

The regularity conditions required for the above development to hold are given by Bernardo and Smith [1993] as follows. Let  $\{p_m(\boldsymbol{\theta}), m = 1, 2, \dots\}$  be a sequence of posterior densities for  $\boldsymbol{\theta}$  of the form  $p_m(\boldsymbol{\theta}) = p(\boldsymbol{\theta} | \mathbf{l}_1, \dots, \mathbf{l}_m)$ . Let  $\boldsymbol{\mu}_m$  be a strict local maximum of  $p_m$  satisfying for all  $m$ :

$$\left\{ \frac{\partial \ln[p_m(\boldsymbol{\theta})]}{\partial \theta_i} \Big|_{\boldsymbol{\theta}=\boldsymbol{\mu}_m} \right\}_{i=1}^r = \mathbf{0}, \quad (\text{B.13})$$

with

$$\boldsymbol{\Sigma}_m = [-\mathbf{H}_m(\boldsymbol{\mu}_m)]^{-1} \quad (\text{B.14})$$

being a positive definite matrix. Under these assumptions, Bernardo and Smith [1993] show that the following conditions are sufficient to ensure that the posterior distribution function is asymptotically equivalent to a multivariate normal distribution function  $N_r(\boldsymbol{\theta} | \boldsymbol{\mu}_m, \boldsymbol{\Sigma}_m^{-1})$ .

- (1) “*Steepness*”.  $\lambda_{\max_m} \rightarrow 0$  as  $m \rightarrow \infty$ , where  $\lambda_{\max_m}$  is the largest eigenvalue of  $\Sigma_m$ .
- (2) “*Smoothness*”. For any  $\varepsilon > 0$ , there exists  $N$  and  $\delta > 0$  such that, for any  $m > N$  and  $\boldsymbol{\theta} \in B_\delta(\boldsymbol{\mu}_m)$ ,  $\mathbf{H}_m(\boldsymbol{\theta})$  exists and satisfies

$$\mathbf{I} - \mathbf{A}(\varepsilon) \leq \mathbf{H}_m(\boldsymbol{\theta}) [-\mathbf{H}_m(\boldsymbol{\mu}_m)]^{-1} \leq \mathbf{I} + \mathbf{A}(\varepsilon), \quad (\text{B.15})$$

where  $\mathbf{I}$  is the  $r \times r$  identity matrix,  $\mathbf{A}(\varepsilon)$  is a  $r \times r$  symmetric positive semidefinite matrix whose largest eigenvalue tends to zero as  $\varepsilon \rightarrow 0$ , and the neighbourhood  $B_\delta$  is defined as

$$B_\delta(\boldsymbol{\theta}^*) = \{\boldsymbol{\theta} \in \Theta; |\boldsymbol{\theta} - \boldsymbol{\theta}^*| < \delta\}. \quad (\text{B.16})$$

- (3) “*Concentration*”. For any  $\delta > 0$ ,

$$\int_{B_\delta(\boldsymbol{\mu}_m)} P_m(\boldsymbol{\theta}) d\boldsymbol{\theta} \rightarrow 1 \text{ as } m \rightarrow \infty. \quad (\text{B.17})$$

An alternative form of condition (3) is given by Bernardo and Smith [1993] as:

- (4) For any  $\delta > 0$ , there exists an integer  $N$  and  $c, d \in \mathfrak{R}^+$  such that, for any  $m > N$  and  $\boldsymbol{\theta} \notin B_\delta(\boldsymbol{\mu}_m)$ ,

$$\ln[p_n(\boldsymbol{\theta})] - \ln[p_n(\mathbf{m}_n)] < -c [(\boldsymbol{\theta} - \mathbf{m}_n)^T \Sigma_n^{-1} (\boldsymbol{\theta} - \mathbf{m}_n)]^d. \quad (\text{B.18})$$

In Chapter 3, the principle of GML estimation is applied to the variance components model, eqn. (3.1), using the inverted gamma probability density function as an informative prior and the Jeffrey's prior as a noninformative prior. The asymptotic properties of the posterior probability density function, for these two choices of priors, will be discussed below. Posterior normality will be proven below only for the case of the single unknown

variance factor in the Gauss-Markov model (cf. eqn. (2.97)). Formulation of a proof of posterior normality in more general cases is not attempted here.

The Gauss-Markov model with the unknown variance factor  $\theta = \sigma_0^2$  reads:

$$\mathbf{I} = \mathbf{A}\mathbf{x} + \mathbf{v}, \quad \mathbf{v} \sim (\mathbf{0}, \mathbf{C}_\theta = \theta\mathbf{I}). \quad (\text{B.19})$$

The Hessian of the marginal log-likelihood function is a scalar of the following form:

$$\begin{aligned} H_{\mathbf{R}} &= \frac{1}{2} \text{tr}(\mathbf{R}\mathbf{I}\mathbf{R}\mathbf{I}) - \mathbf{I}^T \mathbf{R}\mathbf{I}\mathbf{R}\mathbf{I} = \frac{1}{2\sigma_0^2} \text{tr}(\mathbf{R}) - \frac{1}{\sigma_0^6} \hat{\mathbf{v}}^T \hat{\mathbf{v}} = \frac{m-u}{2\sigma_0^4} - \frac{m-u}{\sigma_0^4} \frac{\hat{\mathbf{v}}^T \hat{\mathbf{v}}}{(m-u)\sigma_0^2} \\ &= \frac{m-u}{\sigma_0^4} \left( \frac{1}{2} - \frac{\hat{\sigma}_0^2}{\sigma_0^2} \right) \rightarrow -\frac{m-u}{2\sigma_0^4}. \end{aligned} \quad (\text{B.20})$$

The information matrix (a scalar in this case) is in turn found as:

$$\mathbf{J}_{\mathbf{R}} = \frac{1}{2} \text{tr}(\mathbf{R}\mathbf{I}\mathbf{R}\mathbf{I}) = \frac{1}{2\sigma_0^2} \text{tr}(\mathbf{R}) = \frac{m-u}{2\sigma_0^4}. \quad (\text{B.21})$$

The Hessian of the inverted gamma prior, given by eqn. (3.18), is computed as

$$H_{\mathbf{P}0} = \frac{p+1}{\sigma_0^4} - \frac{2b}{\sigma_0^6}, \quad (\text{B.22})$$

where  $p$  and  $b$  are positive constants (cf. eqn. (2.74)). The inverse of the posterior Hessian, being a sum of  $H_{\mathbf{R}}$  (eqn. (B.20)) and  $H_{\mathbf{P}0}$ , eqn. (B.22)) approaches thus zero as the number of observations,  $m$ , increases to infinity, which verifies condition (1).

Condition (2) can be verified by observing that the posterior Hessian is a continuous function of  $\sigma_0^2$ .

Condition (4), and thus condition (3), is verified as follows. The natural logarithm of the posterior probability density function, given by eqn. (3.5), is computed as:

$$\begin{aligned} l_p &= c - (p+1)\ln(\sigma_0^2) - \frac{b}{\sigma_0^2} - \frac{1}{2}\ln(\sigma_0^{2m}) - \frac{1}{2}\ln\left(\frac{1}{\sigma_0^{2u}}|\mathbf{A}^T\mathbf{A}|\right) - \frac{\mathbf{1}^T\mathbf{R}\mathbf{1}}{2} \\ &= c - (p+1)\ln(\sigma_0^2) - \frac{b}{\sigma_0^2} - \frac{1}{2}\ln(\sigma_0^{2(m-u)}) - \frac{1}{2}\ln|\mathbf{A}^T\mathbf{A}| - \frac{\hat{\mathbf{v}}^T\hat{\mathbf{v}}}{2\sigma_0^2}. \end{aligned} \quad (\text{B.23})$$

The second derivative of  $l_p$  with respect to  $\theta = \sigma_0^2$ , being a sum of  $H_R$  and  $H_{p0}$ , is determined in turn as:

$$l_p''(\theta) = \frac{p+1}{\theta^2} - \frac{2b}{\theta^3} + \frac{m-u}{2\theta^2} - \frac{\hat{\mathbf{v}}^T\hat{\mathbf{v}}}{\theta^3} \rightarrow -\frac{m-u}{2\theta^2} + \frac{p+1}{\theta^2} - \frac{2b}{\theta^3}, \quad (\text{B.24})$$

and, for sufficiently large  $m$ , is strictly negative, which in turn makes  $l_p$  strictly concave, in the neighbourhood of the GMLE of  $\theta$  ( $\mu_m$ ). The remainder of the proof follows the approach presented in Bernardo and Smith [1993, pp. 293-294]. By the strict concavity of  $l_p$ , for any  $\delta > 0$  and  $\theta \notin B_\delta(\mu_m)$ , we have, for some  $\theta^+$  between  $\theta$  and  $\mu_m$ , with angle  $\phi$  between  $\theta - \mu_m$  and  $\nabla l_p(\theta^+)$ ,

$$\begin{aligned} l_p(\theta) - l_p(\mu_m) &= (\theta - \mu_m)\nabla l_p(\theta^+) = |\theta - \mu_m|\|\nabla l_p(\theta^+)\|\cos\phi < -c|\theta - \mu_m| \\ &< -c_1[(\theta - \mu_m)\Sigma_m^{-1}(\theta - \mu_m)]^{1/2}, \end{aligned} \quad (\text{B.25})$$

where:

$$c_1 = c\Sigma_m^{1/2}, \quad c = \inf\{|\nabla l_p(\theta^+)|; \theta \notin B_\delta(\mu_m)\} > 0. \quad (\text{B.26})$$

This concludes the proof of posterior asymptotic normality for the case when the inverted gamma function is used as a prior.

The proof of asymptotic posterior normality with the noninformative prior is analogous to the proof presented above. The Hessian of the noninformative prior is computed from eqn. (3.114) and (3.115) as:

$$\begin{aligned}
H_{N0} &= \frac{1}{2} \text{tr} \left[ J_R^{-1} \left( \frac{\partial^2 J_R}{\partial \theta \partial \theta} - \frac{\partial J_R}{\partial \theta} J_R^{-1} \frac{\partial J_R}{\partial \theta} \right) \right] \\
&= \frac{1}{2} \text{tr} \left\{ \frac{2\sigma_0^4}{m-u} \left[ \text{tr}(3\mathbf{R}\mathbf{R}\mathbf{R}\mathbf{R}) - \text{tr}(\mathbf{R}\mathbf{R}\mathbf{R}) \frac{2\sigma_0^4}{m-u} \text{tr}(\mathbf{R}\mathbf{R}\mathbf{R}) \right] \right\}. \quad (\text{B.27}) \\
&= \frac{1}{2} \text{tr} \left\{ \frac{2\sigma_0^4}{m-u} \left[ \frac{3(m-u)}{\sigma_0^8} - \frac{2(m-u)}{\sigma_0^8} \right] \right\} = \frac{1}{\sigma_0^4}.
\end{aligned}$$

It follows then from eqn. (B.21) and eqn. (B.27) that as the number of observations,  $m$ , approaches infinity the inverse of the posterior Hessian approaches zero, which proves condition (1).

As before, condition (2) follows from the fact that the posterior Hessian, being a sum of  $H_R$  and  $H_{N0}$ , is a continuous function of  $\sigma_0^2$ .

Condition (4) is proven as before by observing that the second derivative of the natural logarithm of the posterior probability density function, given by eqn. (3.38) as

$$l_N = c + \frac{1}{2} \ln |\mathbf{J}_R| - \frac{1}{2} \ln |\mathbf{C}_\theta| - \frac{1}{2} \ln |\mathbf{A}^T \mathbf{C}_\theta^{-1} \mathbf{A}| - \frac{\mathbf{l}^T \mathbf{R} \mathbf{l}}{2}, \quad (\text{B.28})$$

is, for sufficiently large  $m$ , strictly negative in the neighbourhood of  $\mu_m$ . Using the result of eqn. (B.21) we find:

$$\begin{aligned} l_N &= c + \frac{1}{2} \ln \left( \frac{m-u}{\sigma_0^4} \right) - \frac{1}{2} \ln (\sigma_0^{2m}) - \frac{1}{2} \ln \left( \frac{1}{\sigma_0^{2u}} |\mathbf{A}^T \mathbf{A}| \right) - \frac{\mathbf{l}^T \mathbf{R} \mathbf{l}}{2} \\ &= c + \frac{1}{2} \ln \left( \frac{m-u}{\sigma_0^4} \right) - \frac{1}{2} \ln (\sigma_0^{2(m-u)}) - \frac{1}{2} \ln |\mathbf{A}^T \mathbf{A}| - \frac{\hat{\mathbf{v}}^T \hat{\mathbf{v}}}{2\sigma_0^2}. \end{aligned} \quad (\text{B.29})$$

The second derivative of  $l_N$  with respect to  $\theta = \sigma_0^2$  is computed as:

$$l_N''(\theta) = \frac{1}{\theta^2} + \frac{m-u}{2\theta^2} - \frac{\hat{\mathbf{v}}^T \hat{\mathbf{v}}}{\theta^3} \rightarrow \frac{1}{\theta^2} - \frac{m-u}{2\theta^2}, \quad (\text{B.30})$$

which verifies the assumption.

## APPENDIX C

### RESULTS OF THE ESTIMATION OF VARIANCE COMPONENTS

The results of the estimation of variance components for all four epochs, using the REML, GML with the inverted gamma prior, GML with the noninformative prior and the dispersion-mean model with weighted constraints estimation methods, are listed below.

#### C.1 REML Estimation

The REML method was applied to campaigns of 1986, 1989, 1991, and 1993. The iterative process was always started with initial values of 1.0 being assigned to the unknown variance components. The convergence criterion was set to 0.001%.

Table C.1 REML estimation: 1986 campaign, 2 parameter model

Method of Estimation	Estimated Component	Standard Deviation	Speed of Convergence
REML	$a_s^2 = 1.689 \text{ mm}^2$	0.311	7 iterations
	$b_s^2 = 1.660 \text{ ppm}^2$	1.030	

Table C.2 REML estimation: 1989 campaign, 2 parameter model

Method of Estimation	Estimated Component	Standard Deviation	Speed of Convergence
REML	$a_s^2 = \text{--- mm}^2$	---	Diverges
	$b_s^2 = \text{--- ppm}^2$	---	



Table C.3 REML estimation: 1989 campaign, 1 parameter models

Method of Estimation	Estimated Component	Standard Deviation	Speed of Convergence
REML	$a_S^2 = 4.417 \text{ mm}^2$	0.441	1 iteration
	$b_S^2 = 46.42 \text{ ppm}^2$	4.63	1 iteration

Table C.4 REML estimation: 1991 campaign, 3 parameter model

Method of Estimation	Estimated Component	Standard Deviation	Speed of Convergence
REML	$a_S^2 = 0.299 \text{ mm}^2$	0.102	8 iterations
	$b_S^2 = 3.644 \text{ ppm}^2$	0.802	
	$a_D^2 = 1.053 \text{ sec}^2$	0.190	

Table C.5 REML estimation: 1993 campaign, 3 parameter model

Method of Estimation	Estimated Component	Standard Deviation	Speed of Convergence
REML	$a_S^2 = 0.948 \text{ mm}^2$	0.202	6 iterations
	$b_S^2 = 0.596 \text{ ppm}^2$	0.692	
	$a_D^2 = 0.631 \text{ sec}^2$	0.119	

Table C.6 REML estimation: 1993 campaign, 2 parameter models

Method of Estimation	Estimated Component	Standard Deviation	Speed of Convergence
REML	$a_S^2 = 1.107 \text{ mm}^2$	0.140	10 iterations
	$a_D^2 = 0.626 \text{ sec}^2$	0.118	
	$b_S^2 = 13.82 \text{ ppm}^2$	1.73	6 iterations
	$a_D^2 = 0.591 \text{ sec}^2$	0.111	

## C.2 GML Estimation with Inverted Gamma Prior

The GML estimation method with inverted gamma prior was applied to campaigns of 1986, 1989, 1991, and 1993. The values of prior variance components were chosen to be identical to the empirical values given in Chapter 7 (Table 7.2). The iterative process was always started at the set of prior values. The convergence criterion was set to 0.001%.

Table C.7 GML estimation with inverted gamma prior: 1986 campaign, 2 parameter model

Method of Estimation	Standard Deviation of the Prior	Estimated Component	Standard Deviation	Speed of Convergence
GML with Inverted Gamma Prior	$\sigma_{a_{S_0}^2} = a_{S_0}^2$	$a_S^2 = 1.259 \text{ mm}^2$	0.262	10 iterations
	$\sigma_{b_{S_0}^2} = b_{S_0}^2$	$b_S^2 = 3.348 \text{ ppm}^2$	0.885	
Prior	$\sigma_{a_{S_0}^2} = 2 a_{S_0}^2$	$a_S^2 = 1.380 \text{ mm}^2$	0.266	9 iterations
	$\sigma_{b_{S_0}^2} = 2 b_{S_0}^2$	$b_S^2 = 2.778 \text{ ppm}^2$	0.854	
Prior components:	$\sigma_{a_{S_0}^2} = 4 a_{S_0}^2$	$a_S^2 = 1.416 \text{ mm}^2$	0.268	9 iterations
	$\sigma_{b_{S_0}^2} = 4 b_{S_0}^2$	$b_S^2 = 2.607 \text{ ppm}^2$	0.847	
$a_{S_0}^2 = 0.25 \text{ mm}^2$ $b_{S_0}^2 = 9 \text{ ppm}^2$	$\sigma_{a_{S_0}^2} = 8 a_{S_0}^2$	$a_S^2 = 1.426 \text{ mm}^2$	0.268	9 iterations
	$\sigma_{b_{S_0}^2} = 8 b_{S_0}^2$	$b_S^2 = 2.562 \text{ ppm}^2$	0.845	
	$\sigma_{a_{S_0}^2} = 16 a_{S_0}^2$	$a_S^2 = 1.429 \text{ mm}^2$	0.268	9 iterations
	$\sigma_{b_{S_0}^2} = 16 b_{S_0}^2$	$b_S^2 = 2.550 \text{ ppm}^2$	0.844	

Table C.8 GML estimation with inverted gamma prior: 1989 campaign, 2 parameter model

Method of Estimation	Standard Deviation of the Prior	Estimated Component	Standard Deviation	Speed of Convergence
GML with Inverted Gamma Prior	$\sigma_{a_{S_0}^2} = a_{S_0}^2$	$a_S^2 = 3.382 \text{ mm}^2$	0.512	38 iterations (*)
	$\sigma_{b_{S_0}^2} = b_{S_0}^2$	$b_S^2 = 4.077 \text{ ppm}^2$	1.059	
Prior components:	$\sigma_{a_{S_0}^2} = 2 a_{S_0}^2$	$a_S^2 = 3.527 \text{ mm}^2$	0.505	36 iterations (*)
	$\sigma_{b_{S_0}^2} = 2 b_{S_0}^2$	$b_S^2 = 3.317 \text{ ppm}^2$	0.995	
Prior components:	$\sigma_{a_{S_0}^2} = 4 a_{S_0}^2$	$a_S^2 = 3.573 \text{ mm}^2$	0.503	(**)
	$\sigma_{b_{S_0}^2} = 4 b_{S_0}^2$	$b_S^2 = 3.082 \text{ ppm}^2$	0.974	
$a_{S_0}^2 = 0.09 \text{ mm}^2$ $b_{S_0}^2 = 16 \text{ ppm}^2$	$\sigma_{a_{S_0}^2} = 8 a_{S_0}^2$	$a_S^2 = 3.585 \text{ mm}^2$	0.502	(**)
	$\sigma_{b_{S_0}^2} = 8 b_{S_0}^2$	$b_S^2 = 3.020 \text{ ppm}^2$	0.968	
$a_{S_0}^2 = 0.09 \text{ mm}^2$ $b_{S_0}^2 = 16 \text{ ppm}^2$	$\sigma_{a_{S_0}^2} = 16 a_{S_0}^2$	$a_S^2 = 3.588 \text{ mm}^2$	0.502	(**)
	$\sigma_{b_{S_0}^2} = 16 b_{S_0}^2$	$b_S^2 = 3.004 \text{ ppm}^2$	0.967	

(\*) Started at the set of prior values initially diverges very rapidly, then finally converges after a lengthy process;

(\*\*) Started at the set of prior values diverges and has to be restarted with closer approximated starting values.

Table C.9 GML estimation with inverted gamma prior: 1991 campaign, 3 parameter model

Method of Estimation	Standard Deviation of the Prior	Estimated Component	Standard Deviation	Speed of Convergence
GML with Inverted Gamma Prior	$\sigma_{a_{S_0}^2} = a_{S_0}^2$	$a_S^2 = \text{--- mm}^2$	---	diverges
	$\sigma_{b_{S_0}^2} = b_{S_0}^2$	$b_S^2 = \text{--- ppm}^2$	---	
	$\sigma_{a_{D_0}^2} = a_{D_0}^2$	$a_D^2 = \text{--- sec}^2$	---	
Prior components: $a_{S_0}^2 = 0.09 \text{ mm}^2$ $b_{S_0}^2 = 4 \text{ ppm}^2$ $a_{D_0}^2 = 0.81 \text{ sec}^2$	$\sigma_{a_{S_0}^2} = 2 a_{S_0}^2$	$a_S^2 = 0.234 \text{ mm}^2$	0.107	converges to within 6% after 40 iterations
	$\sigma_{b_{S_0}^2} = 2 b_{S_0}^2$	$b_S^2 = 3.749 \text{ ppm}^2$	0.822	
	$\sigma_{a_{D_0}^2} = 2 a_{D_0}^2$	$a_D^2 = 0.984 \text{ sec}^2$	0.181	
	$\sigma_{a_{S_0}^2} = 4 a_{S_0}^2$	$a_S^2 = 0.257 \text{ mm}^2$	0.102	converges to within 10% after 20 iterations
	$\sigma_{b_{S_0}^2} = 4 b_{S_0}^2$	$b_S^2 = 3.626 \text{ ppm}^2$	0.822	
	$\sigma_{a_{D_0}^2} = 4 a_{D_0}^2$	$a_D^2 = 0.985 \text{ sec}^2$	0.182	
	$\sigma_{a_{S_0}^2} = 8 a_{S_0}^2$	$a_S^2 = 0.244 \text{ mm}^2$	0.106	converges to within 1% after 20 iterations
	$\sigma_{b_{S_0}^2} = 8 b_{S_0}^2$	$b_S^2 = 3.698 \text{ ppm}^2$	0.825	
	$\sigma_{a_{D_0}^2} = 8 a_{D_0}^2$	$a_D^2 = 0.985 \text{ sec}^2$	0.182	
	$\sigma_{a_{S_0}^2} = 16 a_{S_0}^2$	$a_S^2 = 0.255 \text{ mm}^2$	0.103	converges to within 8% after 20 iterations
$\sigma_{b_{S_0}^2} = 16 b_{S_0}^2$	$b_S^2 = 3.637 \text{ ppm}^2$	0.823		
$\sigma_{a_{D_0}^2} = 16 a_{D_0}^2$	$a_D^2 = 0.985 \text{ sec}^2$	0.182		

Table C.10 GML estimation with inverted gamma prior: 1993 campaign, 3 parameter model

Method of Estimation	Standard Deviation of the Prior	Estimated Component	Standard Deviation	Speed of Convergence
GML with Inverted Gamma Prior	$\sigma_{a_{So}^2} = a_{So}^2$	$a_S^2 = 0.658 \text{ mm}^2$	0.166	8 iterations
	$\sigma_{b_{So}^2} = b_{So}^2$	$b_S^2 = 1.760 \text{ ppm}^2$	0.576	
	$\sigma_{a_{Do}^2} = a_{Do}^2$	$a_D^2 = 0.595 \text{ sec}^2$	0.112	
Prior components:	$\sigma_{a_{So}^2} = 2 a_{So}^2$	$a_S^2 = 0.728 \text{ mm}^2$	0.167	8 iterations
	$\sigma_{b_{So}^2} = 2 b_{So}^2$	$b_S^2 = 1.421 \text{ ppm}^2$	0.543	
	$\sigma_{a_{Do}^2} = 2 a_{Do}^2$	$a_D^2 = 0.593 \text{ sec}^2$	0.113	
$a_{So}^2 = 0.09 \text{ mm}^2$ $b_{So}^2 = 4 \text{ ppm}^2$ $a_{Do}^2 = 0.64 \text{ sec}^2$	$\sigma_{a_{So}^2} = 4 a_{So}^2$	$a_S^2 = 0.750 \text{ mm}^2$	0.167	10 iterations
	$\sigma_{b_{So}^2} = 4 b_{So}^2$	$b_S^2 = 1.313 \text{ ppm}^2$	0.530	
	$\sigma_{a_{Do}^2} = 4 a_{Do}^2$	$a_D^2 = 0.592 \text{ sec}^2$	0.113	
	$\sigma_{a_{So}^2} = 8 a_{So}^2$	$a_S^2 = 0.756 \text{ mm}^2$	0.167	10 iterations
	$\sigma_{b_{So}^2} = 8 b_{So}^2$	$b_S^2 = 1.284 \text{ ppm}^2$	0.527	
	$\sigma_{a_{Do}^2} = 8 a_{Do}^2$	$a_D^2 = 0.592 \text{ sec}^2$	0.114	
	$\sigma_{a_{So}^2} = 16 a_{So}^2$	$a_S^2 = 0.758 \text{ mm}^2$	0.167	10 iterations
	$\sigma_{b_{So}^2} = 16 b_{So}^2$	$b_S^2 = 1.277 \text{ ppm}^2$	0.526	
	$\sigma_{a_{Do}^2} = 16 a_{Do}^2$	$a_D^2 = 0.592 \text{ sec}^2$	0.114	

Table C.11 GML estimation with inverted gamma prior: 1993 campaign, 2 parameter model

Method of Estimation	Standard Deviation of the Prior	Estimated Component	Standard Deviation	Speed of Convergence
GML with Inverted Gamma Prior	$\sigma_{b_{So}^2} = b_{So}^2$	$b_S^2 = 13.15 \text{ ppm}^2$	1.68	5 iterations
	$\sigma_{a_{Do}^2} = a_{Do}^2$	$a_D^2 = 0.558 \text{ sec}^2$	0.105	
Prior components:	$\sigma_{b_{So}^2} = 2 b_{So}^2$	$b_S^2 = 13.26 \text{ ppm}^2$	1.69	6 iterations
	$\sigma_{a_{Do}^2} = 2 a_{Do}^2$	$a_D^2 = 0.555 \text{ sec}^2$	0.106	
$b_{So}^2 = 4 \text{ ppm}^2$ $a_{Do}^2 = 0.64 \text{ sec}^2$	$\sigma_{b_{So}^2} = 4 b_{So}^2$	$b_S^2 = 13.29 \text{ ppm}^2$	1.70	6 iterations
	$\sigma_{a_{Do}^2} = 4 a_{Do}^2$	$a_D^2 = 0.555 \text{ sec}^2$	0.106	
	$\sigma_{b_{So}^2} = 8 b_{So}^2$	$b_S^2 = 13.29 \text{ ppm}^2$	1.70	6 iterations
	$\sigma_{a_{Do}^2} = 8 a_{Do}^2$	$a_D^2 = 0.554 \text{ sec}^2$	0.106	
	$\sigma_{b_{So}^2} = 16 b_{So}^2$	$b_S^2 = 13.29 \text{ ppm}^2$	1.70	6 iterations
	$\sigma_{a_{Do}^2} = 16 a_{Do}^2$	$a_D^2 = 0.554 \text{ sec}^2$	0.106	

### C.3 GML Estimation with Noninformative Prior

The GML method with noninformative prior was applied to campaigns of 1986, 1989, 1991, and 1993. The iterative process was always started with initial values of 1.0 being assigned to the unknown variance components. The convergence criterion was set to 0.001%.

Table C.12 GML estimation with noninformative prior: 1986 campaign, 2 parameter model

Method of Estimation	Estimated Component	Standard Deviation	Speed of Convergence
GML with Noninf Prior	$a_S^2 = 1.701 \text{ mm}^2$	0.327	5 iterations
	$b_S^2 = 1.785 \text{ ppm}^2$	1.112	

Table C.13 GML estimation with noninformative prior: 1989 campaign, 2 parameter model

Method of Estimation	Estimated Component	Standard Deviation	Speed of Convergence
GML with Noninf. Prior	$a_S^2 = \text{---} \text{ mm}^2$	---	Diverges
	$b_S^2 = \text{---} \text{ ppm}^2$	---	

Table C.14 GML estimation with noninformative prior: 1989 campaign, 1 parameter models

Method of Estimation	Estimated Component	Standard Deviation	Speed of Convergence
GML with Noninf. Prior	$a_S^2 = 4.462 \text{ mm}^2$	0.447	1 iteration
	$b_S^2 = 46.88 \text{ ppm}^2$	4.65	1 iteration

Table C.15 GML estimation with noninformative prior: 1991 campaign, 3 parameter model

Method of Estimation	Estimated Component	Standard Deviation	Speed of Convergence
GML with Noninformative Prior	$a_S^2 = 0.319 \text{ mm}^2$	0.113	10 iterations
	$b_S^2 = 3.678 \text{ ppm}^2$	0.847	
	$a_D^2 = 1.087 \text{ sec}^2$	0.199	

Table C.16 GML estimation with noninformative prior: 1993 campaign, 3 parameter model

Method of Estimation	Estimated Component	Standard Deviation	Speed of Convergence
GML with Noninformative Prior	$a_S^2 = 0.950 \text{ mm}^2$	0.219	7 iterations
	$b_S^2 = 0.734 \text{ ppm}^2$	0.793	
	$a_D^2 = 0.654 \text{ sec}^2$	0.126	

Table C.17 GML estimation with noninformative prior: 1993 campaign, 2 parameter models

Method of Estimation	Estimated Component	Standard Deviation	Speed of Convergence
GML with. Noninformative Prior	$a_S^2 = 1.123 \text{ mm}^2$	0.143	10 iterations
	$a_D^2 = 0.650 \text{ sec}^2$	0.125	
	$b_S^2 = 14.02 \text{ ppm}^2$	1.77	6 iterations
	$a_D^2 = 0.613 \text{ sec}^2$	0.118	

#### C.4 The Dispersion-Mean Model with Weighted Constraints

The dispersion-mean model with weighted constraints was applied to campaigns of 1986, 1989, 1991, and 1993 to estimate the variance components. The values of prior variance components were chosen to be identical to the empirical values listed in Table 7.2. The iterative process was always started at the set of prior values. The convergence criterion was set to 0.001%.



Table C.18 Dispersion-mean model with weighted constraints: 1986 campaign,  
2 parameter model

Method of Estimation	Standard Deviation of the Prior	Estimated Component	Standard Deviation	Speed of Convergence	
Dispersion-Mean Model with Weighted Constraints	$\sigma_{a_{S_0}^2} = a_{S_0}^2$	$a_S^2 = 0.835 \text{ mm}^2$	0.168	15 iterations	
	$\sigma_{b_{S_0}^2} = b_{S_0}^2$	$b_S^2 = 5.135 \text{ ppm}^2$	1.060		
	Prior components:	$\sigma_{a_{S_0}^2} = 2 a_{S_0}^2$	$a_S^2 = 1.287 \text{ mm}^2$	0.242	14 iterations
		$\sigma_{b_{S_0}^2} = 2 b_{S_0}^2$	$b_S^2 = 2.936 \text{ ppm}^2$	1.032	
	Prior components:	$\sigma_{a_{S_0}^2} = 4 a_{S_0}^2$	$a_S^2 = 1.559 \text{ mm}^2$	0.288	11 iterations
		$\sigma_{b_{S_0}^2} = 4 b_{S_0}^2$	$b_S^2 = 2.015 \text{ ppm}^2$	1.027	
	Prior components: $a_{S_0}^2 = 0.25 \text{ mm}^2$	$\sigma_{a_{S_0}^2} = 8 a_{S_0}^2$	$a_S^2 = 1.654 \text{ mm}^2$	0.304	9 iterations
		$\sigma_{b_{S_0}^2} = 8 b_{S_0}^2$	$b_S^2 = 1.752 \text{ ppm}^2$	1.029	
	Prior components: $b_{S_0}^2 = 9 \text{ ppm}^2$	$\sigma_{a_{S_0}^2} = 16 a_{S_0}^2$	$a_S^2 = 1.680 \text{ mm}^2$	0.309	9 iterations
		$\sigma_{b_{S_0}^2} = 16 b_{S_0}^2$	$b_S^2 = 1.683 \text{ ppm}^2$	1.030	

Table C.19 Dispersion-mean model with weighted constraints: 1989 campaign,  
2 parameter model

Method of Estimation	Standard Deviation of the Prior	Estimated Component	Standard Deviation	Speed of Convergence
Dispersion-Mean Model	$\sigma_{a_{S_0}^2} = a_{S_0}^2$	$a_S^2 = 0.412 \text{ mm}^2$	0.085	31 iterations
	$\sigma_{b_{S_0}^2} = b_{S_0}^2$	$b_S^2 = 28.38 \text{ ppm}^2$	3.14	
with Weighted Constraints	$\sigma_{a_{S_0}^2} = 2 a_{S_0}^2$	$a_S^2 = 0.979 \text{ mm}^2$	0.160	0.1% conv. in 20 iterations
	$\sigma_{b_{S_0}^2} = 2 b_{S_0}^2$	$b_S^2 = 16.72 \text{ ppm}^2$	2.30	
Prior components:	$\sigma_{a_{S_0}^2} = 4 a_{S_0}^2$	$a_S^2 = 1.977 \text{ mm}^2$	0.283	0.03% conv. in 40 iterations
	$\sigma_{b_{S_0}^2} = 4 b_{S_0}^2$	$b_S^2 = 7.003 \text{ ppm}^2$	1.688	
$a_{S_0}^2 = 0.09 \text{ mm}^2$	$\sigma_{a_{S_0}^2} = 8 a_{S_0}^2$	$a_S^2 = 3.019 \text{ mm}^2$	0.442	28 iterations
	$\sigma_{b_{S_0}^2} = 8 b_{S_0}^2$	$b_S^2 = 2.894 \text{ ppm}^2$	1.622	
$b_{S_0}^2 = 16 \text{ ppm}^2$	$\sigma_{a_{S_0}^2} = 16 a_{S_0}^2$	$a_S^2 = 3.739 \text{ mm}^2$	0.571	0.01% conv. in 40 iterations
	$\sigma_{b_{S_0}^2} = 16 b_{S_0}^2$	$b_S^2 = 1.388 \text{ ppm}^2$	1.690	

Table C.20 Dispersion-mean model with weighted constraints: 1991 campaign,  
3 parameter model

Method of Estimation	Standard Deviation of the Prior	Estimated Component	Standard Deviation	Speed of Convergence	
Dispersion-Mean Model with Weighted Constraints	$\sigma_{a_{S_0}^2} = a_{S_0}^2$	$a_S^2 = 0.215 \text{ mm}^2$	0.060	18 iterations	
	$\sigma_{b_{S_0}^2} = b_{S_0}^2$	$b_S^2 = 4.034 \text{ ppm}^2$	0.717		
	$\sigma_{a_{D_0}^2} = a_{D_0}^2$	$a_D^2 = 1.044 \text{ sec}^2$	0.183		
	Prior components:	$\sigma_{a_{S_0}^2} = 2 a_{S_0}^2$	$a_S^2 = 0.261 \text{ mm}^2$	0.083	11 iterations
		$\sigma_{b_{S_0}^2} = 2 b_{S_0}^2$	$b_S^2 = 3.810 \text{ ppm}^2$	0.762	
		$\sigma_{a_{D_0}^2} = 2 a_{D_0}^2$	$a_D^2 = 1.051 \text{ sec}^2$	0.188	
	$a_{S_0}^2 = 0.09 \text{ mm}^2$	$\sigma_{a_{S_0}^2} = 4 a_{S_0}^2$	$a_S^2 = 0.287 \text{ mm}^2$	0.096	10 iterations
		$\sigma_{b_{S_0}^2} = 4 b_{S_0}^2$	$b_S^2 = 3.697 \text{ ppm}^2$	0.788	
		$\sigma_{a_{D_0}^2} = 4 a_{D_0}^2$	$a_D^2 = 1.053 \text{ sec}^2$	0.190	
	$b_{S_0}^2 = 4 \text{ ppm}^2$	$\sigma_{a_{S_0}^2} = 8 a_{S_0}^2$	$a_S^2 = 0.296 \text{ mm}^2$	0.100	9 iterations
		$\sigma_{b_{S_0}^2} = 8 b_{S_0}^2$	$b_S^2 = 3.659 \text{ ppm}^2$	0.798	
		$\sigma_{a_{D_0}^2} = 8 a_{D_0}^2$	$a_D^2 = 1.053 \text{ sec}^2$	0.190	
	$a_{D_0}^2 = 0.81 \text{ sec}^2$	$\sigma_{a_{S_0}^2} = 16 a_{S_0}^2$	$a_S^2 = 0.298 \text{ mm}^2$	0.102	9 iterations
		$\sigma_{b_{S_0}^2} = 16 b_{S_0}^2$	$b_S^2 = 3.648 \text{ ppm}^2$	0.801	
		$\sigma_{a_{D_0}^2} = 16 a_{D_0}^2$	$a_D^2 = 1.053 \text{ sec}^2$	0.190	

Table C.21 Dispersion-mean model with weighted constraints: 1993 campaign,  
3 parameter model

Method of Estimation	Standard Deviation of the Prior	Estimated Component	Standard Deviation	Speed of Convergence	
Dispersion-Mean Model with Weighted Constraints	$\sigma_{a_{S_0}^2} = a_{S_0}^2$	$a_S^2 = 0.328 \text{ mm}^2$	0.072	10 iterations	
	$\sigma_{b_{S_0}^2} = b_{S_0}^2$	$b_S^2 = 3.410 \text{ ppm}^2$	0.708		
	$\sigma_{a_{D_0}^2} = a_{D_0}^2$	$a_D^2 = 0.652 \text{ sec}^2$	0.120		
	Prior components:	$\sigma_{a_{S_0}^2} = 2 a_{S_0}^2$	$a_S^2 = 0.510 \text{ mm}^2$	0.116	12 iterations
		$\sigma_{b_{S_0}^2} = 2 b_{S_0}^2$	$b_S^2 = 2.301 \text{ ppm}^2$	0.707	
		$\sigma_{a_{D_0}^2} = 2 a_{D_0}^2$	$a_D^2 = 0.651 \text{ sec}^2$	0.121	
	$a_{S_0}^2 = 0.09 \text{ mm}^2$ $b_{S_0}^2 = 4 \text{ ppm}^2$	$\sigma_{a_{S_0}^2} = 4 a_{S_0}^2$	$a_S^2 = 0.724 \text{ mm}^2$	0.161	12 iterations
		$\sigma_{b_{S_0}^2} = 4 b_{S_0}^2$	$b_S^2 = 1.335 \text{ ppm}^2$	0.706	
		$\sigma_{a_{D_0}^2} = 4 a_{D_0}^2$	$a_D^2 = 0.644 \text{ sec}^2$	0.121	
	$a_{D_0}^2 = 0.64 \text{ sec}^2$	$\sigma_{a_{S_0}^2} = 8 a_{S_0}^2$	$a_S^2 = 0.871 \text{ mm}^2$	0.188	10 iterations
		$\sigma_{b_{S_0}^2} = 8 b_{S_0}^2$	$b_S^2 = 0.822 \text{ ppm}^2$	0.697	
		$\sigma_{a_{D_0}^2} = 8 a_{D_0}^2$	$a_D^2 = 0.636 \text{ sec}^2$	0.120	
		$\sigma_{a_{S_0}^2} = 16 a_{S_0}^2$	$a_S^2 = 0.927 \text{ mm}^2$	0.167	12 iterations
		$\sigma_{b_{S_0}^2} = 16 b_{S_0}^2$	$b_S^2 = 0.655 \text{ ppm}^2$	0.526	
		$\sigma_{a_{D_0}^2} = 16 a_{D_0}^2$	$a_D^2 = 0.632 \text{ sec}^2$	0.119	

Table C.22 Dispersion-mean model with weighted constraints: 1993 campaign,  
2 parameter models

Method of Estimation	Standard Deviation of the Prior	Estimated Component	Standard Deviation	Speed of Convergence
Dispersion-Mean Model	$\sigma_{b_{So}^2} = b_{So}^2$	$b_S^2 = 12.47 \text{ ppm}^2$	1.46	7 iterations
	$\sigma_{a_{Do}^2} = a_{Do}^2$	$a_D^2 = 0.601 \text{ sec}^2$	0.112	
with Weighted Constraints	$\sigma_{b_{So}^2} = 2 b_{So}^2$	$b_S^2 = 13.39 \text{ ppm}^2$	1.64	6 iterations
	$\sigma_{a_{Do}^2} = 2 a_{Do}^2$	$a_D^2 = 0.594 \text{ sec}^2$	0.112	
Prior components:	$\sigma_{b_{So}^2} = 4 b_{So}^2$	$b_S^2 = 13.71 \text{ ppm}^2$	1.70	6 iterations
	$\sigma_{a_{Do}^2} = 4 a_{Do}^2$	$a_D^2 = 0.592 \text{ sec}^2$	0.112	
$b_{So}^2 = 4 \text{ ppm}^2$	$\sigma_{b_{So}^2} = 8 b_{So}^2$	$b_S^2 = 13.79 \text{ ppm}^2$	1.72	6 iterations
	$\sigma_{a_{Do}^2} = 8 a_{Do}^2$	$a_D^2 = 0.591 \text{ sec}^2$	0.112	
$a_{Do}^2 = 0.64 \text{ sec}^2$	$\sigma_{b_{So}^2} = 16 b_{So}^2$	$b_S^2 = 13.81 \text{ ppm}^2$	1.73	6 iterations
	$\sigma_{a_{Do}^2} = 16 a_{Do}^2$	$a_D^2 = 0.591 \text{ sec}^2$	0.112	

### C.5 Comparison of the Estimated Variance Components

The results of the estimation of variance components for all four epochs, using the REML, GML with the inverted gamma prior, GML with the noninformative prior and the dispersion-mean model with weighted constraints estimation methods, are given below in graphical form in Figures C.1 to C.10. The results are given on an epoch by epoch basis, separately for each component. To facilitate interpretation the square roots of the estimated variance components are given. The prior values (means of the prior distribution), which were used in the GML with the inverted gamma prior and the dispersion-mean model with weighted constraints estimation methods, are also indicated in all figures.

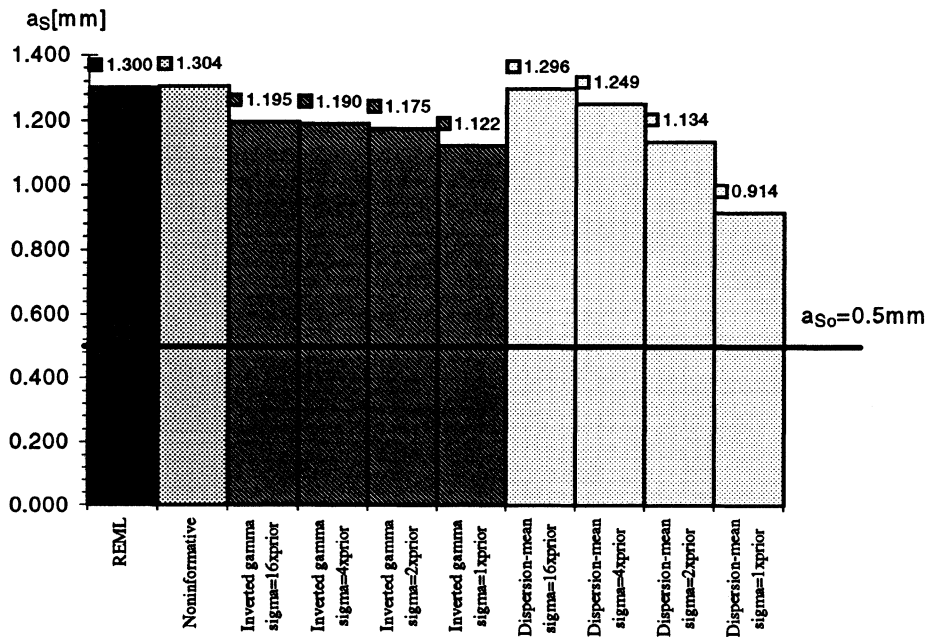


Figure C.1 Campaign of 1986 ( $a_s$  component)

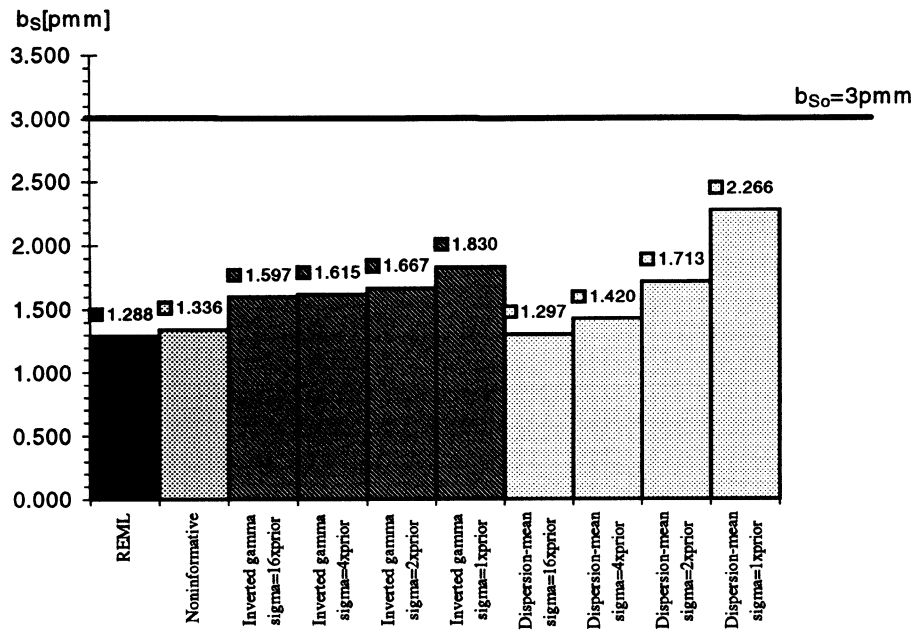


Figure C.2 Campaign of 1986 ( $b_s$  component)

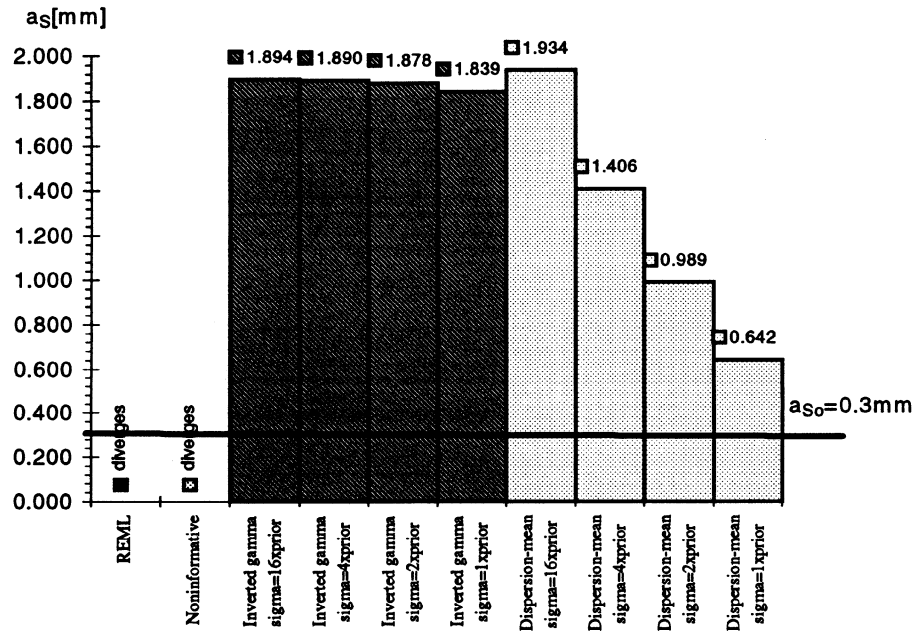


Figure C.3 Campaign of 1989 ( $a_s$  component)

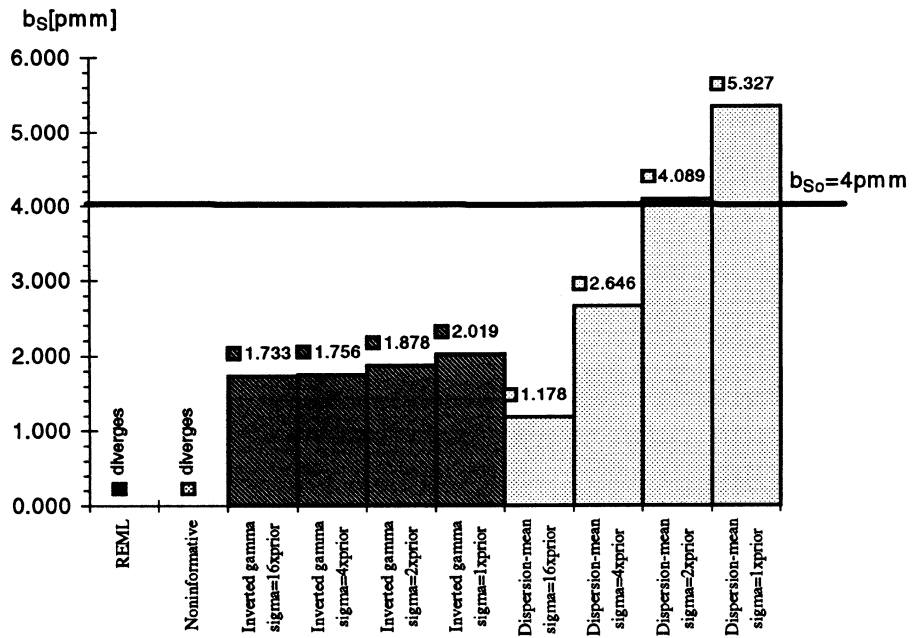


Figure C.4 Campaign of 1989 ( $b_s$  component)

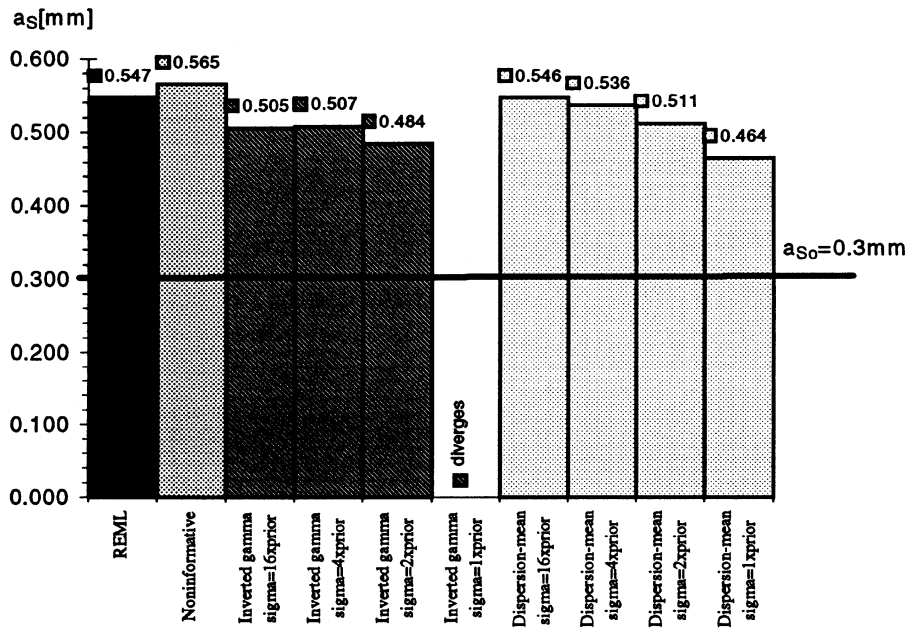


Figure C.5 Campaign of 1991 ( $a_s$  component)



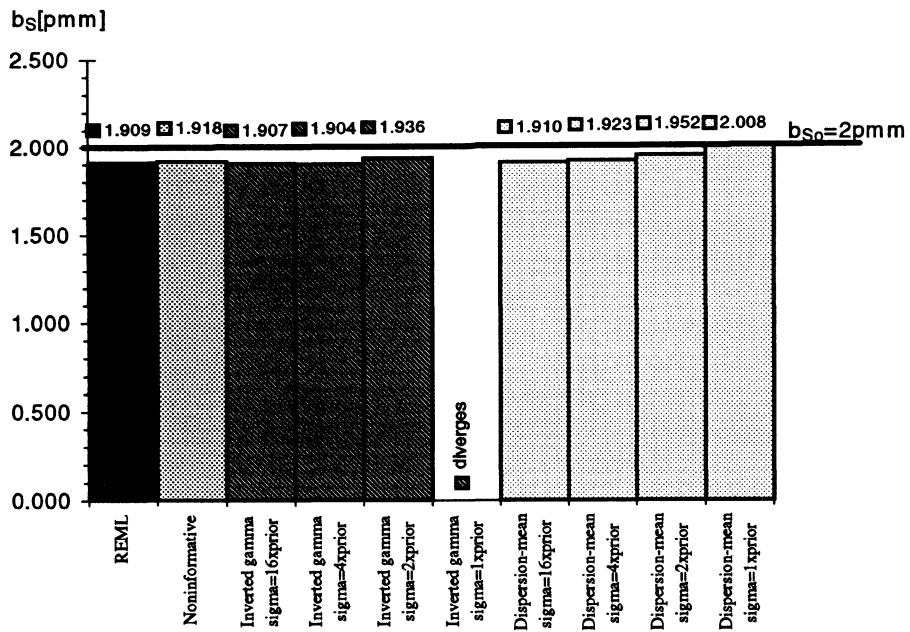


Figure C.6 Campaign of 1991 ( $b_s$  component)

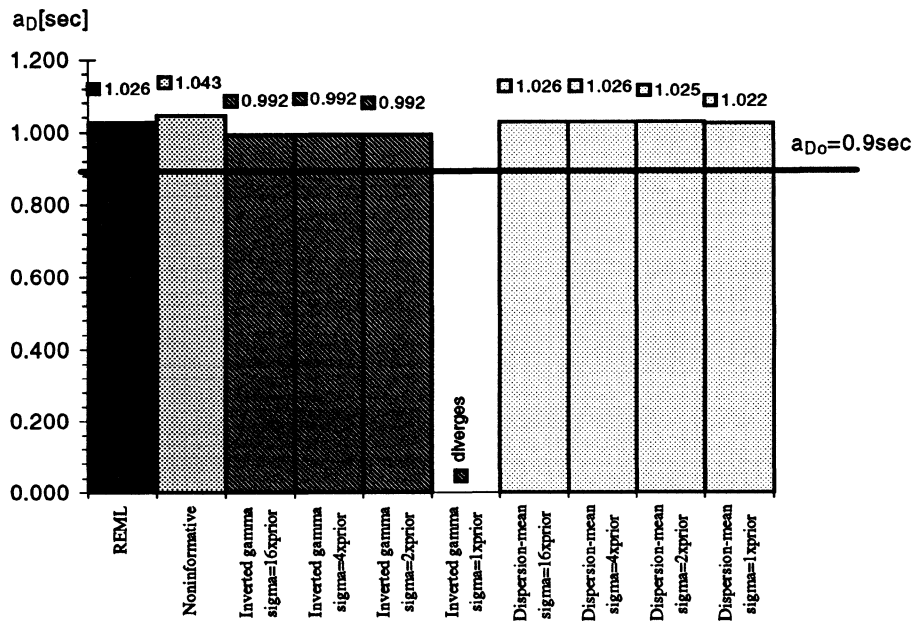


Figure C.7 Campaign of 1991 ( $a_D$  component)

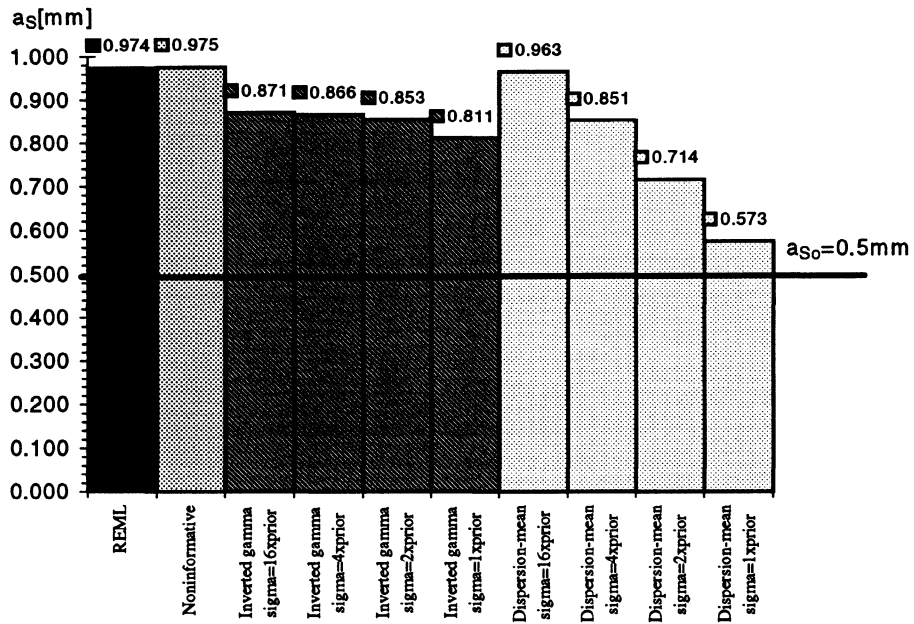


Figure C.8 Campaign of 1993 ( $a_s$  component)

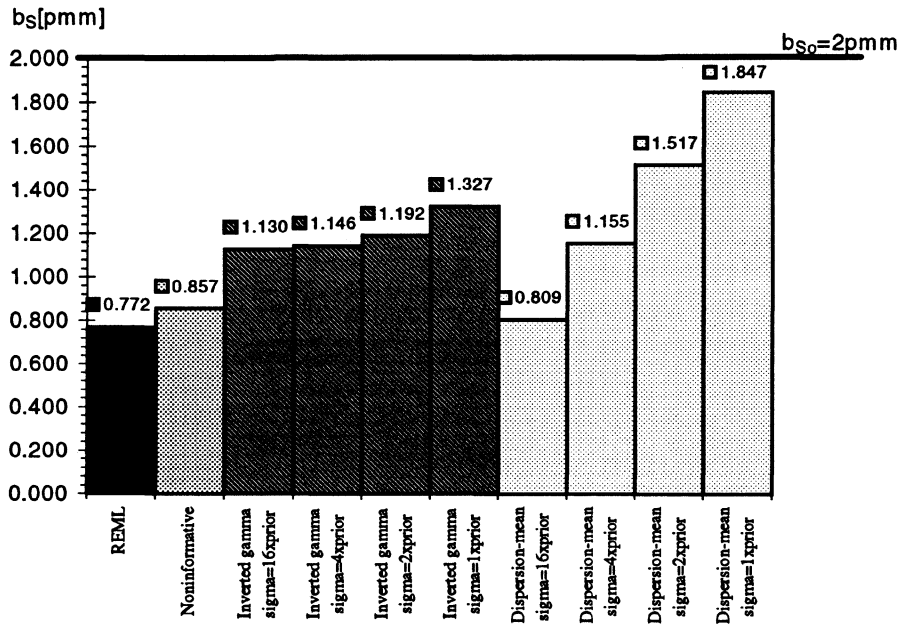


Figure C.9 Campaign of 1993 ( $b_s$  component)

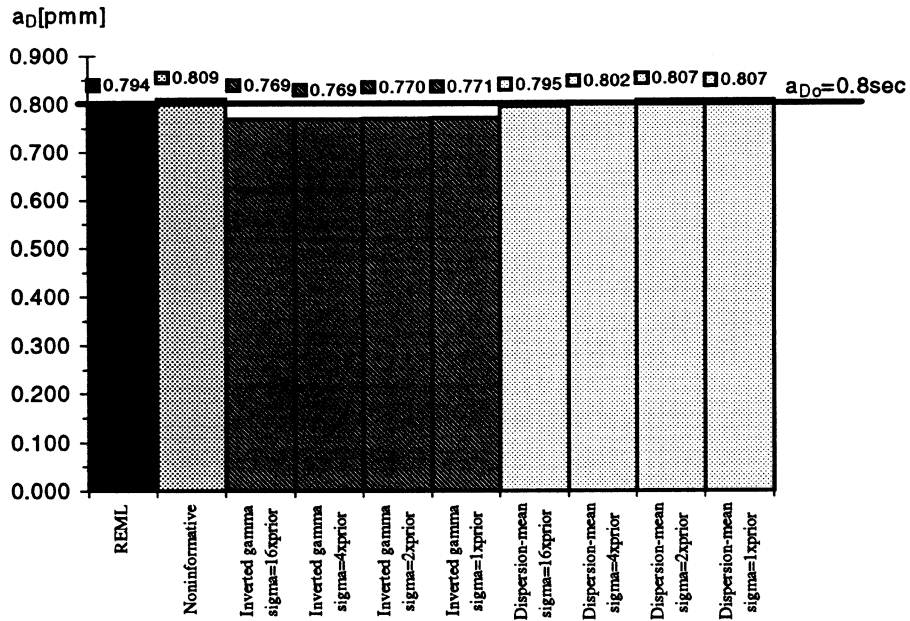


Figure C.10 Campaign of 1993 ( $a_D$  component)

### C.6 Ratios of the Estimated Variance Components to their Asymptotic Standard Deviations

The ratios of the estimated variance components to their asymptotic standard deviations, expressing the tightness of the asymptotic confidence intervals, are given below in Figures C.11 to C.20 - in the same order as the corresponding estimates appearing in Figures C.1 through C.10.

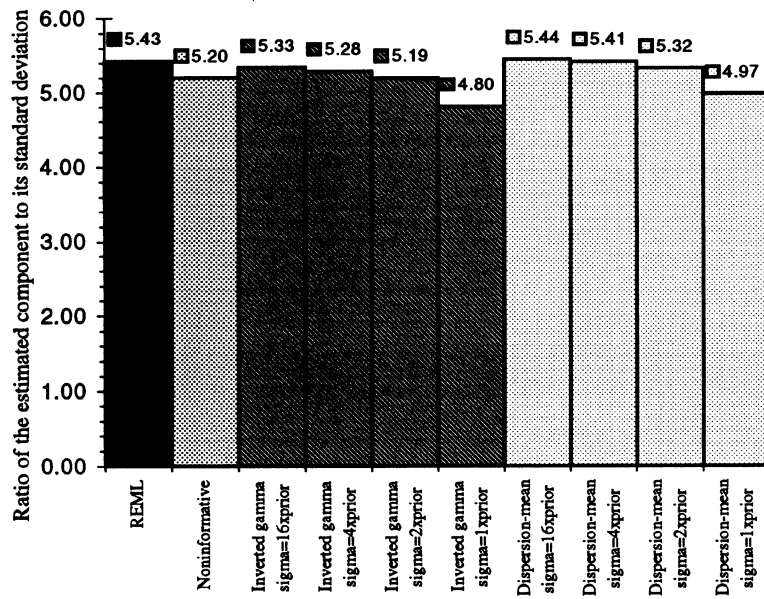


Figure C.11 Campaign of 1986 ( $a_s^2$  component)

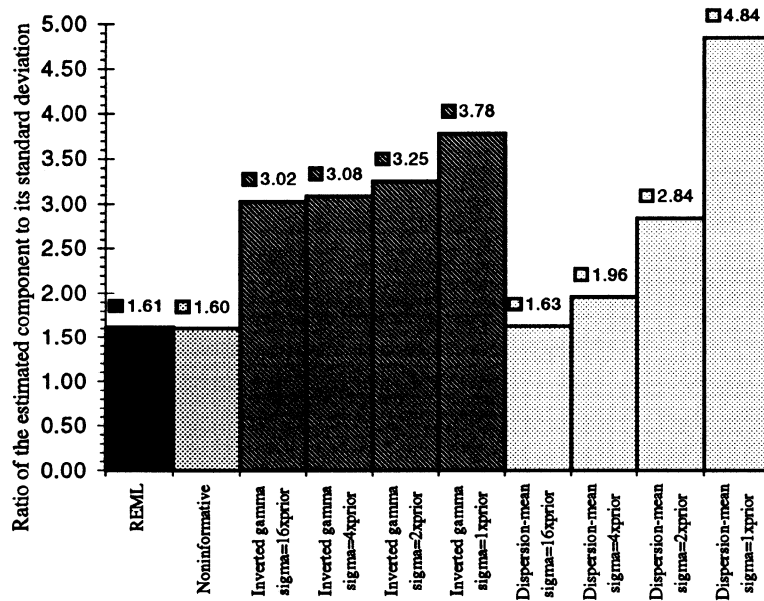


Figure C.12 Campaign of 1986 ( $b_s^2$  component)

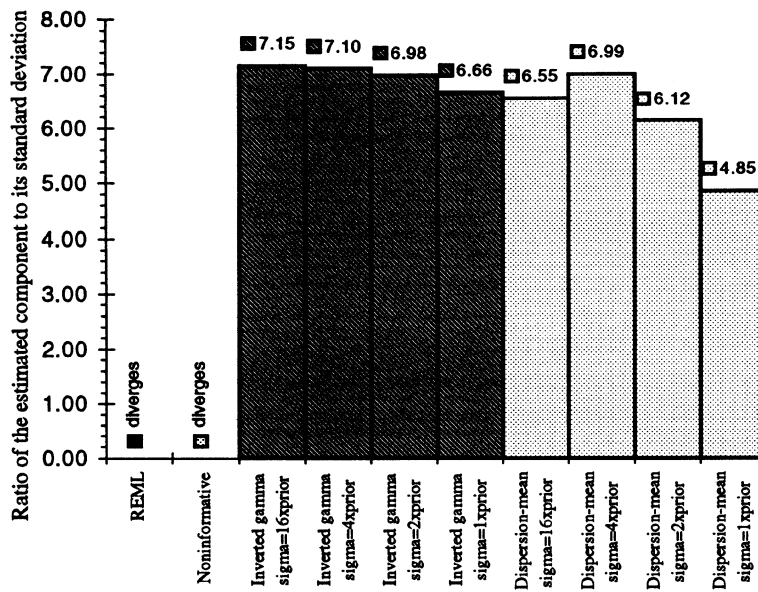


Figure C.13 Campaign of 1989 ( $a_s^2$  component)

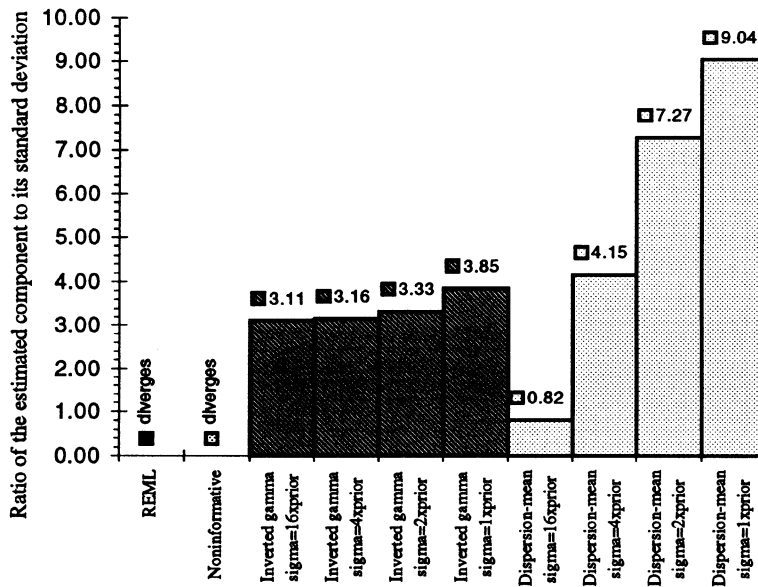


Figure C.14 Campaign of 1989 ( $b_s^2$  component)

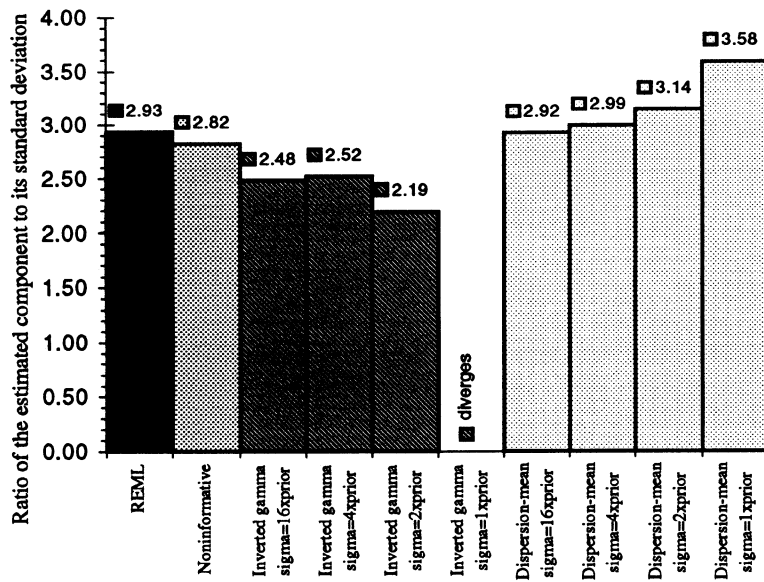


Figure C.15 Campaign of 1991 ( $a_s^2$  component)

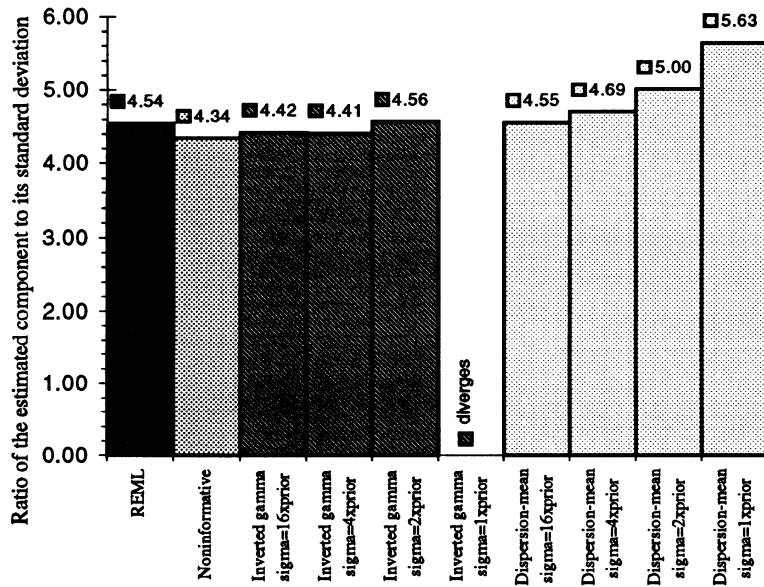


Figure C.16 Campaign of 1991 ( $b_s^2$  component)

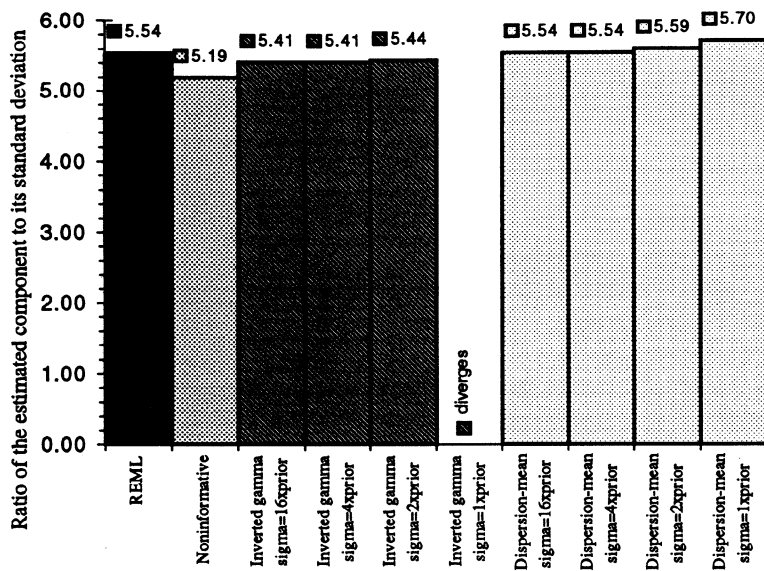


Figure C.17 Campaign of 1991 ( $a_D^2$  component)

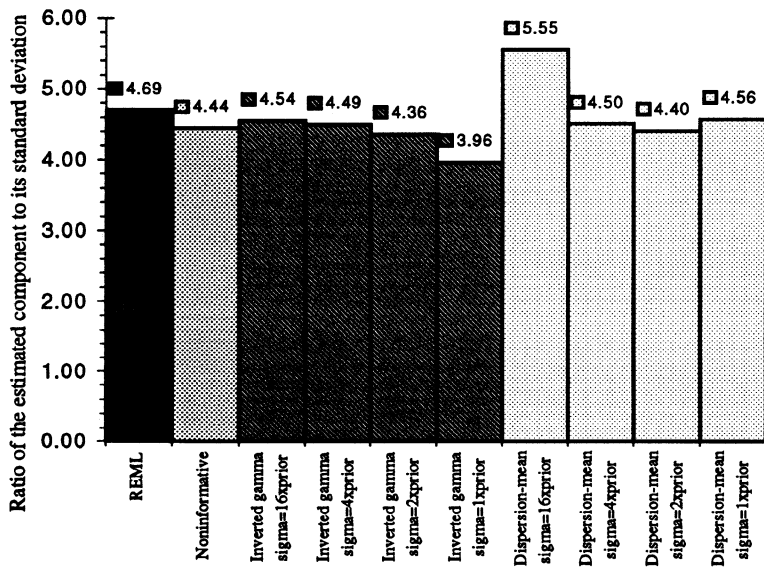


Figure C.18 Campaign of 1993 ( $a_S^2$  component)

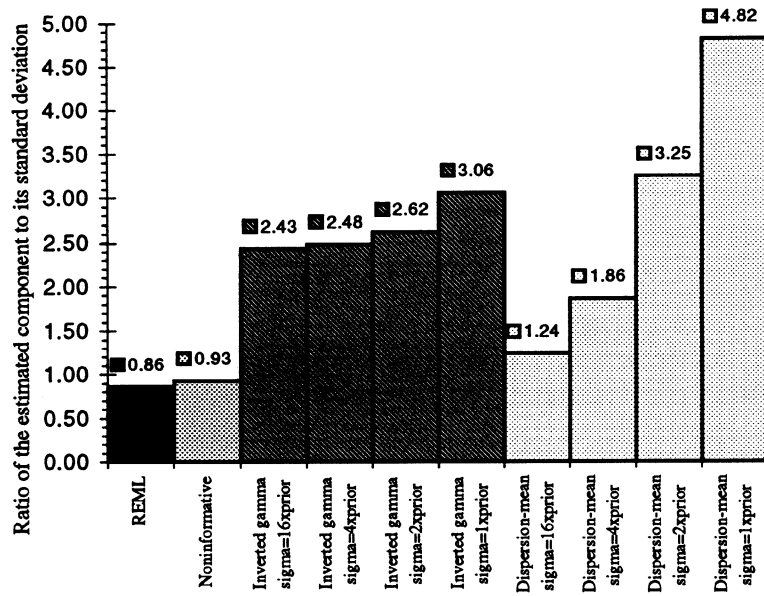


Figure C.19 Campaign of 1993 ( $b_s^2$  component)

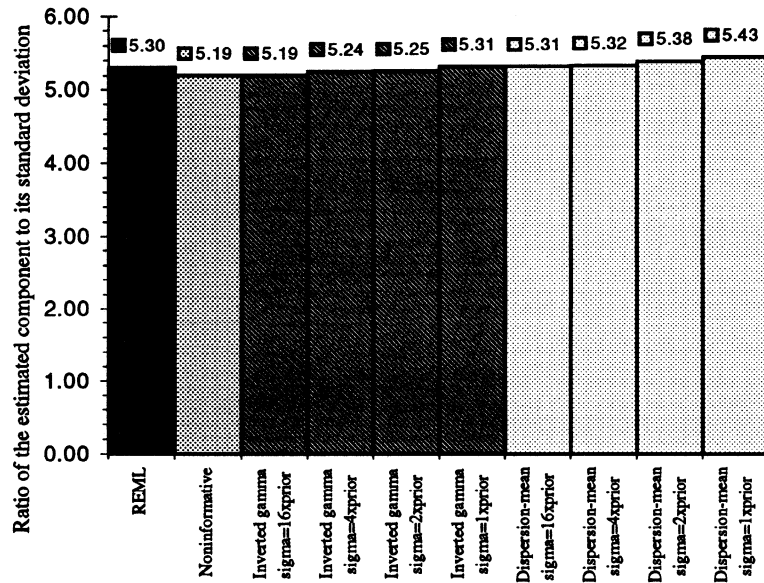


Figure C.20 Campaign of 1993 ( $a_D^2$  component)



## VITA

Candidate's full name: Jacek Grodecki

Place and date of birth: Mikolow, Poland  
May 11, 1964

Permanent address: 2 Sloneczna Street, Mikolow, 43-190, Poland

Universities attended: Technical University of Mining and Metallurgy in Krakow,  
Department of Mining Surveying  
1983-1989 **M.Sc.E.**

Technical University of Mining and Metallurgy in Krakow,  
Interdepartmental School of Foreign Trade Economics for  
Mining Engineers  
1985-1990 **Dipl.**

University of New Brunswick, Department of Geodesy and  
Geomatics Engineering, Fredericton, N.B.  
1990-1997 **Ph.D.** candidate

## PUBLICATIONS

- Chrzanowski, A., M. Caissy, J. Grodecki and J. Secord (1994). "Software Development and Training for Geometrical Deformation Analysis." Final Report to the Department of Natural Resources, Geodetic Survey Division, Ottawa, Ontario.
- Chrzanowski, A., W.J.T. Greening, J. Grodecki and J. Robbins (1993). "Design of Geodetic Control for Tunnelling Projects." *Canadian Tunnelling Canadian*, pp. 1-11.
- Grodecki, J., G.L. Robinson, W.J.T. Greening and J. Robbins (1993). "Orthometric Corrections to Height Differences in Super Collider's Primary Vertical Control Network." Technical Report CPB-100500, The PB/MK Team, Superconducting Super Collider Project, Dallas, Texas.
- Grodecki, J., W.J.T. Greening and J. Robbins (1992). "Analysis of Influence of Accuracy of Gravity Data on Vertical Control." Technical Report CPB-100083, The PB/MK Team, Superconducting Super Collider Project, Dallas, Texas.
- Grodecki, J., G.L. Robinson, W.J.T. Greening and J. Robbins (1992). "Analysis of Influence of Tidal Phenomena on Vertical Control." Technical Report CPB-100014, The PB/MK Team, Superconducting Super Collider Project, Dallas, Texas.

Chrzanowski, A., Grodecki, J., W.J.T. Greening and J. Robbins (1992). "Horizontal Tunnel Control Analysis." Technical Report CPB-001140, The PB/MK Team, Superconducting Super Collider Project, Dallas, Texas.

Grodecki, J. (1989). "Analysis of Ionospheric Correction in GPS and TRANSIT Satellite Systems." Master Thesis. Technical University of Mining and Metallurgy in Krakow, Department of Mining Surveying, 99 pp., (in Polish).

#### CONFERENCE PRESENTATIONS

Grodecki, J. (1994). "The Analysis of Geodetic Aspects for Vertical Control in the Superconducting Super Collider." Presented at the 5th Annual Atlantic Institute Research Conference, May 13-14, Fredericton, New Brunswick.

Chrzanowski, A. , Y.Q. Chen, J. Grodecki and J. Westrop (1991). "Effects of Systematic Errors of GPS Observations on Deformation Modelling." Presented at AGU Chapman Conference on Time Dependent Positioning: Modelling Crustal Deformation, September 23-25, Annapolis, Maryland.