NATIONAAL LUCHT- EN RUIMTEVAARTLABORATORIUM

NATIONAL AEROSPACE LABORATORY NLR

THE NETHERLANDS

NLR TR 88108 U

ESTIMATING THE PROBABILITY OF VERTICAL OVERLAP FROM THE PAIRED AIRCRAFT DATA OBTAINED IN THE EUROPEAN VERTICAL DATA COLLECTION USING THE PROGRAM DGLDiF

by

G. Moek

NLR

581-01

DOCUMENT CONTROL SHEET

	ORIGINATOR'S REF. NLR TR 88108 U		SECURITY CLASS. Unclassified

ORIGINATOR
National Aerospace Laboratory NLR, Amsterdam, The Netherlands

TITLE
Estimating the probability of vertical overlap from the paired aircraft data obtained in the european vertical data collection using the program DGLDiF

PRESENTED AT

AUTHORS G. Moek	DATE 880708	pp ref 179 31

DESCRIPTORS

Air traffic control	Midair collisions
Risk	Vertical orientation
Flight altitude	Proximity
Threat evaluation	Probability theory
Mathematic models	Maximum likelihood estimates

ABSTRACT
The problem of estimating the probability of vertical overlap in possible future airway systems with a vertical separation standard of 1000 ft above Flight Level 290 is addressed. It forms part of an overall safety assessment of such potential systems. The problem is solved by developing a mathematical probability model of the vertical distance between aircraft in a pair. A priori unknown model parameters are estimated by means of the maximum likelihood method from presently available data on the height keeping performance of aircraft in Europe. Special attention is given to the effect on the modelling process of the limited amount of data on large height keeping errors. The point estimate obtained is 6.6×10^{-6}, while an associated 95 per cent interval estimate is 2.3×10^{-6} - 14.8×10^{-6}.

NLR TR 88108 U

ESTIMATING THE PROBABILITY OF VERTICAL OVERLAP FROM THE
PAIRED AIRCRAFT DATA OBTAINED IN THE EUROPEAN VERTICAL
DATA COLLECTION USING THE PROGRAM DGLDiF

by

G. Moek

Division: Informatics

Prepared: G. Moek/WK

Approved: W. Loeve/

Completed   : 080788

Ordernumber: 106.758/554.505

Typ.   : PTP NE

SUMMARY

The problem of estimating the probability of vertical overlap in possible future airway systems with a vertical separation standard of 1000 ft above Flight Level 290 is addressed. It forms part of an overall safety assessment of such potential systems. The problem is solved by developing a mathematical probability model of the vertical distance between aircraft in a pair. A priori unknown model parameters are estimated by means of the maximum likelihood method from presently available data on the height keeping performance of aircraft in Europe. Special attention is given to the effect on the modelling process of the limited amount of data on large height keeping errors. The point estimate obtained is 6.6×10^{-6}, while an associated 95 per cent interval estimate is 2.3×10^{-6} - 14.8×10^{-6}.

LIST OF SYMBOLS AND ABBREVIATIONS

a	scale parameter of a Generalized Laplace probability distribution
a_1	scale parameter of a Generalized Laplace core distribution
a_{1_o}	initial estimate of a_1
a_{1_j}, $j=1,2,\ldots$	sequence of estimates of a_1
a_{1_∞}	limit of the sequence a_{1_j}, $j=1,2,\ldots$
a_2	scale parameter of a Generalized Laplace tail distribution
a_{2_o}	initial estimate of a_2
a_{2_j}, $j=1,2,\ldots$	sequence of estimates of a_2
a_{2_∞}	limit of the sequence a_{2_j}, $j=1,2,\ldots$
a_j, $j=1,2,\ldots$	sequence of estimates of a
A	quantity occurring in the log likelihood function for grouped data and defined by eq.(52)
b	shape parameter of a Generalized Laplace probability distribution
b_1	shape parameter of a Generalized Laplace core distribution
b_{1_o}	initial estimate of b_1
b_2	shape parameter of a Generalized Laplace tail distribution
b_{2_o}	initial estimate of b_2
$cov\{\underline{x},\underline{x}^T\}$	covariance matrix of the random vector $\underline{x}$
C_i, $i=0,1,\ldots$	quantity used in the proof of the proposition of Section 4.4.4, eqs.(140),(141)
CFL^{**}_i	estimate of the height of the assigned flight level of aircraft i
CP	Central Processor
d	vertical distance between aircraft in a pair
$det(.)$	determinant of matrix between brackets
d.f.	degrees of freedom
DDE	Double Double Exponential
DE	Double Exponential
$DELTA_i$	height keeping error of aircraft i
DGL	Double Generalized Laplace

E_1 $\frac{\partial}{\partial \alpha} \mathcal{L}(\alpha, a_1, a_2, b_1, b_2)$

E_2 $\frac{\partial}{\partial a_1} \mathcal{L}(\alpha, a_1, a_2, b_1, b_2)$

E_3 $\frac{\partial}{\partial a_2} \mathcal{L}(\alpha, a_1, a_2, b_1, b_2)$

E_4 $\frac{\partial}{\partial b_1} \mathcal{L}(\alpha, a_1, a_2, b_1, b_2)$

E_5 $\frac{\partial}{\partial b_2} \mathcal{L}(\alpha, a_1, a_2, b_1, b_2)$

$E\{\underline{x}\}$	expected value of random variable $\underline{x}$
$f(b)$	function defining the maximum likelihood estimate of the shape parameter of a Generalized Laplace distribution for non-grouped data, eq.(75)
$f_{\underline{d}}(d), f(d)$	probability density function of the random variable $\underline{d}$
$f_{DDE}(d)$	Double Double Exponential probability density function
$f_{DE}(d)$	Double Exponential probability density function
$f_{DGL}(d)$	Double Generalized Laplace probability density function
$f_{GDE}(d)$	Gaussian Double Exponential probability density function
$f_{GLDE}(d)$	Generalized Laplace Double Exponential probability density function
ft	foot, feet
$F(d)$	probability distribution function of the random variable $\underline{d}$
$F_{DE}(d)$	Double Exponential probability distribution function
$F_{DGL}(d)$	Double Generalized Laplace probability distribution function
FLx	Flight Level x
$FOI(\alpha, a_1, a_2, b_1, b_2)$	Fisher Observed Information matrix as a function of the parameters $\alpha, a_1, a_2, b_1, b_2$
F04ATF	Fortran subroutine from the NAG library for solving a system of linear equations by Crout's factorization method
$g(b)$	function defining the maximum likelihood estimate of the shape parameter of a Generalized Laplace distribution for grouped data, eq.(91)
G	Gaussian
G	case, indicating that all the five parameters of a Double Generalized Laplace distribution are estimated from grouped data
GDE	Gaussian Double Exponential
GLDE	Generalized Laplace Double Exponential

G05CAF	Fortran subroutine from the NAG library for computing a pseudo-random number from a uniform distribution between 0 and 1
G_2	case, indicating that only the two parameters α and a_1 of a Double Generalized Laplace distribution are estimted from grouped data
G_3	case, indicating that only the parameters α, a_1 and a_2 of a Double Generalized Laplace distribution are estimated from grouped data
G_4	case, indicating that only the parameters α, a_1, a_2 and b_1 of a Double Generalized Laplace distribution are estimated from grouped data
h(b)	function of the maximum likelihood estimate of the shape parameter of a Generalized Laplace distribution, using non-grouped data, defined by eq.(78)
h_i, i=1,2,...,m	height of the histogram on class interval $\pm$ i, i=1,2,...,m
$H(\alpha,a_1,a_2)$	Hessian matrix of the second partial derivates of the log likelihood function for the Double Double Exponential distribution and grouped data
$H_1(\alpha,a,a)$	submatrix of the matrix $H(\alpha,a,a)$ defined by eq.(128)
$\bar{H}_i$	average, corrected height of aircraft i as derived from precision radar measurements
k(b)	function of the maximum likelihood estimate of the shape parameter of a Generalized Laplace distribution, using grouped data, defined by eq.(94)
log	logarithm with base e
L	likelihood function
L	minus the right-hand side vector of the linearized likelihood equations of the Double Generalized Laplace distribution, eq.(31)
$L(\alpha,a_1,a_2,b_1,b_2)$	likelihood function, depending on α,a_1,a_2,b_1,b_2
ℓ	log likelihood function
$\ell(\alpha,a_1,a_2,b_1,b_2)$	log likelihood function, depending on α,a_1,a_2,b_1,b_2
$L_2(\alpha,a_1,a_2,b_1,b_2)$	likelihood function, depending on α,a_1,a_2,b_1,b_2, given that the mean of each observation is equal to the observed value
$\ell_2(\alpha,a_1,a_2,b_1,b_2)$	log likelihood function, depending on α,a_1,a_2,b_1,b_2, given that the mean of each observation is equal to the observed value

$(\mathcal{L}_2-\mathcal{L})(\alpha,a_1,a_2,b_1,b_2)$	standardized log likelihood function, depending on α,a_1,a_2,b_1,b_2
L_i, $i=1,2,\ldots,5$	elements of the vector L
m	sample mean
m	number of class intervals for grouped data
M	maximum of $|E_1|,|E_2|,\ldots,|E_i|$, $i=2,3,4,5$
ML	Maximum Likelihood
MNO	left-hand side matrix of the linearized likelihood equations of the Double Generalized Laplace distribution, eq.(31)
M_i, $i=1,2,\ldots,5$	elements of the first column of the matrix MNO
n	sample size
NAG	Numerical Algorithms Group
N_i, $i=1,2,\ldots,5$	elements of the second column of the matrix MNO
O_i, $i=1,2,\ldots,5$	elements of the third column of the matrix MNO
P_i, $i=1,2,\ldots,m$	probability of an outcome in class interval i, $i=1,2,\ldots,m$ in a single experiment
p_i, $i=\pm1,2,\ldots,m$	value of the Double Generalized Laplace probability density in the midpoint of class interval i, $i=\pm1,2,\ldots,m$
Prob{A}	probability of the event A
$P_z(1000)$	probability of vertical overlap in a 1000 ft vertical separation standard environment
Q_i, $i=1,2,\ldots,5$	fourth column of the matrix MNO
r_i, $i=\pm1,2,\ldots,m$	linear residual on class interval i, $i=\pm1,2,\ldots,m$, defined by eq.(175)
$R(\alpha,a_1,a_2,b_1,b_2)$	relative likelihood function with arguments α,a_1,a_2,b_1,b_2
R_i, $i=\pm1,2,\ldots,m$	logarithmic residual on class interval i, $i=\pm1,2,\ldots,m$, defined by eq.(176)
R_i, $i=1,2,\ldots,5$	fifth column of the matrix MNO
s	sample standard deviation
S	case, indicating that all the five parameters of a Double Generalized Laplace distribution are estimated from non-grouped data
S_z	vertical separation standard
S_2	case, indicating that only the parameters α and a_1 of a Double Generalized Laplace distribution are estimated from non-grouped data
S_3	case, indicating that only the parameters α,a_1 and a_2 of a Double Generalized Laplace distribution are estimated from non-grouped data

S_4	case, indicating that only the parameters α, a_1, a_2 and b_1 of a Double Generalized Laplace distribution are estimated from non-grouped data
$u_i, \ i=1,2,\ldots,n$	i-th realization of a random variable which is uniformly distributed on the interval $[0,1]$
w	class interval of grouped data
$X_{DE}(0.9)$	0.9 quantile of a random variable with a Double Exponential distribution
$X_{DGL}(0.9)$	0.9 quantile of a random variable with a Double Generalized Laplace distribution
$\underline{x}_i, \ i=1,2,\ldots,m$	random variable denoting the number of observations in class interval i, in the case of grouped data
$x_i, \ i=1,2,\ldots,m$	realization of the random variable $\underline{x}_i, \ i=1,2,\ldots,m$
$x_i, \ i=1,2,\ldots$	sequence used in the proposition of Section 4.4.4, eqs.(140),(141)
x_∞	limit of the sequence $x_i, \ i=1,2,\ldots$
X	Matrix defined by eq.(137)
$y_i, \ i=1,2,\ldots,m$	$(i-\tfrac{1}{2})w$
Y	vector defined by eq.(138)
α	weighting factor of the tail distribution within a Double Generalized Laplace distribution
α_o	initial estimate of α
β_1	(sample) skewness
β_2	(sample) kurtosis
$\beta_2(b)$	kurtosis of a Generalized Laplace probability distribution
$\beta_2(\alpha,a_1,a_2,b_1,b_2)$	kurtosis of a Double Generalized Laplace probability distribution
$\gamma(b,x)$	incomplete Gamma function defined by eq.(147)
$\Gamma(b,x)$	incomplete Gamma function defined by eq.(145)
$\Gamma(b)$	Gamma function defined by eq.(6)
$\Delta\alpha$	correction on α
Δa_1	correction on a_1
$\Delta a_{1_j}, \ j=0,1,\ldots$	correction on $a_{1_j}, \ j=0,1,\ldots$
Δa_2	correction on a_2
$\Delta a_{2_j}, \ j=0,1,\ldots$	correction on $a_{2_j}, \ j=0,1,\ldots$
Δb_1	correction on b_1
Δb_2	correction on b_2

Δx_i, $i=0,1,\ldots$	correction on x_i, $i=0,1,\ldots$
ε	tolerance on the iterative solution process of the likelihood equations
η	second coordinate of confidence ellipses based on the asymptotic maximum likelihood estimator
λ_z	height of an aircraft
$\Lambda(\alpha,a_1,a_2,b_1,b_2)$	likelihood ratio with arguments α,a_1,a_2,b_1 and b_2
μ	population mean
ξ	first coordinate of confidence ellipses based on the asymptotic maximum likelihood estimator
$\rho\{\underline{x},\underline{y}\}$	correlation coefficient between the random variables $\underline{x}$ and $\underline{y}$
σ	population standard deviation
$\sigma(a)$	standard deviation as a function of the scale parameter of a Gaussian or a Double Exponential probability distribution
$\sigma(a,b)$	standard deviation of Generalized Laplace probability distribution as a function of its scale and shape parameter
$\sigma(\alpha,a_1,a_2,b_1,b_2)$	standard deviation of a Double Generalized Laplace probability distribution as a function of its five parameters
σ_1	standard deviation of the core distribution
$\sigma_1(a_1,b_1)$	standard deviation of a Generalized Laplace core distribution
σ_2	standard deviation of the tail distribution
$\sigma_2(a_2,b_2)$	standard deviation of a Generalized Laplace tail distribution
χ^2	Chi-squared
$\chi^2_{.p,(d.f.)}$	p-th percentile of a χ^2-distributed random variable with d.f. degrees of freedom
$\Psi(b)$	derivative of the logarithm of the Gamma function $\Gamma(b)$

If necessary, underlined symbols are used to distinguish random variables from their realizations, and symbols with a "^" or "~" are used to distinguish estimates from their true values.

CONTENTS

33 Tables
79 Figures

(179 pages in total)

1   INTRODUCTION

Presently, studies are underway aimed at investigating the possibility of reducing the vertical separation standard at and above Flight Level 290 (FL 290) from 2000 ft to 1000 ft, without exceeding a specified risk level (Refs. 1,2,3). An important element of these studies is a collision risk assessment, i.e. an assessment of the risk of collisions due to the loss of vertical separation between aircraft assigned to adjacent flight levels of the same track. The risk assessment consists of two parts, namely a collision risk estimation process and a process of determining a so-called Target Level of Safety against which the estimated risk is to be compared.

The collision risk estimation process is based on the use of the Reich collision risk model (Refs. 4,5,6). This model relates the collision risk to various characteristics of the population of aircraft as well as of the airway system under consideration. With regard to the latter, the separation standard is particularly important. With regard to the former, the probability of vertical overlap for aircraft assigned to adjacent flight levels is of prime importance. The model is characterized by its structure and by its model parameters. Having agreed on the model structure, numerical values for each of the model parameters are needed. On the assumption that the structure of the model is the same for the various systems of interest, it offers the possibility of estimating the risk of each of those systems, by substituting the appropriate parameter values. As a consequence, estimates of the collision risk associated with e.g. a vertical separation standard of 1000 ft, can be obtained beforehand without actually having to apply such a standard in practice. This is done by extrapolating parameter values valid for the present situation of 2000 ft vertical separation to that of 1000 ft. Of course, some assumptions have to be made for such an extrapolation, and it may be necessary to verify these assumptions in practice a posteriori, after a change in an airway system has taken place (Ref. 7).

In this report, the process of estimating the probability of vertical overlap $P_z(1000)$ is considered. This probability can be computed from the probability distribution of the vertical distance between two aircraft. Several problems have to be solved during this process. Firstly, there is the problem of the data itself. In practice, it is not feasible to measure

this probability distribution directly. It can, however, be obtained from data on the height keeping performance of individually observed aircraft. In fact, it can even be obtained in two ways from those data. In the context of the European Vertical Data Collection, these are called the Singles Approach and the Pairing Approach (Ref. 8). In the Singles Approach, the probability distribution of the deviations from the assigned flight level of a typical aircraft is constructed first from the data. This distribution then is used in a convolution process to arrive at the probability distribution of the vertical distance between aircraft. In the Pairing Approach, the latter distribution is formed directly by combining the data of pairs of individually observed aircraft. Because it was expected that the accuracy of the pairs data would be better than that of the singles data, the Pairing Approach was adopted as the primary approach to be used in the European Vertical Data Collection study (Ref. 9). The aircraft pairs data are extensively described in Ref. 8 and will be recalled in Section 2.

The second problem, which applies to both the approaches, is that although the pertinent observed probability distributions contain many data points in the core, i.e. within say 500 ft from their mean values, there are virtually no data points in the tails of these distributions, the latter parts just being the most important for the collision risk. As a result, it is necessary to describe the observed distributions by means of a mathematical model and to use this model for extrapolation to the tail regions. Because of the inherent uncertainty of the extrapolation, it is desirable to build in some cautiousness into the modelling process.

The process of modelling the observed distribution of aircraft pairs data consists of three steps, namely (i) selecting an appropriate family of analytical probability distributions depending on one or more parameters, (ii) selecting a unique member from the family by fixing the parameter values, and (iii) evaluating the quality of the resulting fit. In practice, one may need to proceed iteratively when it turns out that some assumptions made earlier are not appropriate. The process will be detailed in Section 3 with emphasis on step (i). It will be argued that the family of the so-called Double Generalized Laplace probability distributions is very well suited to describe the aircraft pairs data of Section 2. Step (ii) will be elaborated in Section 4 for the maximum likelihood parameter estimation technique. As the resulting parameter estimation problem is

fairly complicated, some a priori analysis is of great use for the numerical solution process that will have to be used. Before considering the results of the application to the real data of Section 2, the parameter estimation algorithm will be applied to simulated data based on different Double Generalized Laplace probability distributions. In fact, the application to simulated data serves two purposes, namely, firstly, validation of the estimation algorithms when the true parameter values are known and, secondly, assessment of the achievable accuracy and properties of the estimation problem under controlled conditions.

2   AIRCRAFT PAIRS DATA

A first estimate of the probability of vertical overlap due to the loss of vertical separation in a 1000 ft environment can be derived from data on the vertical distance between aircraft in the present 2000 ft situation. The natural (but expensive) way to collect accurate data on the height keeping performance of aircraft is by means of precision lock follow radars. Two complicating factors play a part. The first one is that it is practically not feasible to directly measure the vertical distance between two aircraft because the frequency of the event that two aircraft are at the same pre-defined (horizontal) location (i.e. the radar site) is very low. Moreover, it would require two radars. Thus, it is necessary to use radar measurements of individually observed aircraft. The second problem is that the vertical distance between two aircraft consists of two parts, viz. a part describing the aircraft height keeping errors and a part describing the distance between the assigned flight levels (see Figure 1). Hence, some information on the heights of the flight levels is needed. For the European Vertical Data Collection it was decided, on the basis of both economical, operational, and technical considerations, to derive the required flight level height information from measurements of the atmospheric situation by meteorological balloons. As a compromise between costs and (temporal) resolution, a frequency of (in principle) four balloon ascents per day was adopted.

From the measurements of the atmosphere as made by the meteorological balloons, the heights of the flight levels during the balloon ascents can easily be computed. By means of time interpolation, the flight level heights at arbitrary points of time can be derived. Thus, it is possible to compute for each aircraft measured by the precision lock follow radar its deviation from the assigned flight level, at the time of the measurement (see Figure 2). In the European Vertical Data Collection study, this is called the Singles Approach. The great disadvantage of this approach is that the _absolute_ accuracy of the flight level heights is rather limited when using balloon derived data, due to the required integration process. This disadvantage may be overcome by considering height _differences_ between flight level pairs. Similarly, height differences between aircraft can be considered. This has been called the Pairing Approach (see Figure 3), and has resulted in a sample of aircraft pairs data, representative of a 1000 ft environment.

Reference 8 gives a full description of the processing, analysis, and results of the European Vertical Data Collection. Following Reference 8, the vertical distance d between two aircraft in a pair in a 1000 ft vertical separation standard environment is defined by

$$d = DELTA_2 - DELTA_1 + S_z(\text{measured}) \tag{1}$$

where

$$DELTA_i = \bar{H}_i - CFL^{**}_i \qquad , i=1,2 \tag{2}$$

with $\bar{H}_i$ being the average corrected radar measured height of aircraft i in the pair and CFL^{**}_i being the best estimate (based on the meteorological balloon data) of the height of the assigned flight level of aircraft i in the pair at the time and location of its measurement by the radar. It is remarked that each radar measurement took place over a time interval of 59 seconds, with a frequency of 1 Hz and that the resulting 60 samples were corrected for refractivity, elevation, and upward bending and next averaged (see Ref. 8). The quantity $S_z(\text{measured})$ is defined by

$$S_z(\text{measured}) = \text{best estimate of one thousand feet nominal vertical separation under prevailing meteorological measurement conditions} \tag{3}$$

Figure 4 shows a histogram of the pairs data and Figure 5 shows the logarithm of this histogram to emphasize the data in the tail area. The sample actually consists of 4960 pairs data and is characterized by the following statistics (see Appendix A):

$$\left.\begin{array}{llll}
\text{sample mean} & : & m & = 997.1 \text{ ft} \\
\text{sample standard deviation} & : & s & = 125.2 \text{ ft} \\
\text{sample skewness} & : & \beta_1 & = -0.00015 \\
\text{sample kurtosis} & : & \beta_2 & = 4.33
\end{array}\right\} \quad (4)$$

Because there are no obvious physical reasons why the sample of pairs data should be asymmetrical, and because of the small value of the sample skewness, it will be assumed in the sequel that the observed distribution is symmetrical about a value of 1000 ft. Tables 1 to 3 inclusive give the corresponding histogram data for class intervals of 40, 20, and 10 ft respectively. Notice the four isolated data points in the classes 62 and 63 of Table 3.

3   APPROACH TO THE ESTIMATION OF THE PROBABILITY OF VERTICAL OVERLAP

3.1   The modelling approach

The probability of vertical overlap $P_z(1000)$ may be considered to be the main parameter of the collision risk model. It is, of course, directly related to the height-keeping performance of the aircraft as well as to the vertical separation standard. The histograms of pairs data described in Section two show no data points beyond class 63, i.e. beyond 630 ft away from the separation standard of one thousand feet. From this observation, however, it is not allowable to conclude that such data points have probability of occurrence equal to zero. What has to be done is to extrapolate the observed histogram to the region of critical interest for the collision risk assessment. As in previous assessments of the collision risk due to the loss of separation, the tool to be used is a mathematical model of the probability distribution of the vertical distance between aircraft in a pair. For the range of data points observed, the modelled distribution and the observed distribution are as close as possible in a certain statistical sense. Beyond the range of data points observed, the

modelled probability distribution is to represent the rare events that could not be observed during the limited period of time during which the data collection took place. Clearly, the choice of the modelled probability distribution is critical for the collision risk assessment, and a lot of technical/operational as well as statistical knowledge and experience is necessary to guide this choice.

To be somewhat more specific, the mathematical model consists of a family of analytical probability density functions f(d) depending on one or more parameters. In developing the model, the following three choices have to be made (compare e.g. Ref. 10):

i)   Which family of analytical probability density functions should be considered?
ii)   Which criterion should be used to uniquely select one member from the family to represent the data?
iii) Which criterion should be used to evaluate the quality of the resulting fit?

These questions will be discussed in the following three subsections. It should be noticed that two criteria play a part. The criterion under ii) is usually a mathematical criterion, with theoretically appealing properties, leading to a reproducable objective numerical result. It always remains, however, to verify whether these theoretically nice properties are really important and representative for the practical application at hand (step iii)). As a result, the process may turn out to become an iterative process, in which some of the initial assumptions have to be refined or changed in a later stage.

3.2   Selection of a family of probability distributions

The question of which family of probability densities to fit to the data is not an easy one to answer. A great many choices are possible. Some factors that may influence the choice are: ultimate use of the fitted density, number of parameters, complexity of the parameter estimation problem, prior knowledge about the data (i.e. the underlying physical phenomena), the data itself. On the one hand the family should be sufficiently rich to describe the data, whereas on the other hand it should be avoided to overfit the data.

In Ref. 11, a total of seven analytical probability densities are considered. They all are based on the Gaussian (G) and the Double Exponential (DE) probability density. The Gaussian density is very well known, and is for example applicable when a great number of independent error sources add up. The Double Exponential density is known for its relatively thick tail and is, therefore, particularly useful for getting cautious extrapolations for collision risk assessment purposes. Some of the densities considered in Ref. 11 are mixtures of two Double Exponential and/or Gaussian densities, on the assumption that the overall population consists of two sub-populations. Convoluted versions of (combinations of) the Gaussian and Double Exponential densities are also included. It then was concluded that the histogram of the pairs data looked like something in between of a Gaussian and a Double Exponential density. A class of probability densities which includes these two densities is formed by the so-called Generalized Laplace densities. The Double Exponential density is also known as the first Laplacian density. Some promising preliminary results for the Generalized Laplace probability density, using some ad hoc parameter values are already described in Ref. 11.

The family of Generalized Laplace (GL) probability densities $f_{GL}(d)$ is defined by

$$f_{GL}(d) = \frac{1}{2ab\Gamma(b)} \, e^{-\left|\frac{d-\mu}{a}\right|^{1/b}} \tag{5}$$

(a>0) where the Gamma function $\Gamma(b)$ is defined by

$$\Gamma(b) = \int_0^\infty e^{-t} \, t^{b-1} \, dt \tag{6}$$

This function is extensively tabulated, see e.g. Ref. 12. See also Figure 6. The b-parameter is called the shape parameter wheras the a-parameter is called the scale parameter. The Generalized Laplace density reduces to the Gaussian and Double Exponential densities for b≈0.5 and b=1.0 respectively. The Generalized Laplace density is also known as the power exponential density.

The central moments of the Generalized Laplace density are defined by

$$E\left\{(\underline{d} - \mu)^k\right\} = \int_{-\infty}^{\infty} (x - \mu)^k f_{GL}(x)dx \tag{7}$$

Due to the symmetry about the mean value μ, the moments of the Generalized Laplace density are zero for odd values of k. For even values of k holds

$$E\left\{(\underline{d} - \mu)^k\right\} = \frac{a^k}{\Gamma(b)} \Gamma((k+1)b) \tag{8}$$

The skewness β_1 is zero and the kurtosis β_2 is found to be given by

$$\beta_2(b) = \frac{\Gamma(5b)}{\Gamma(3b)} \frac{\Gamma(b)}{\Gamma(3b)} \tag{9}$$

The variance of a variable having a Generalized Laplace probability density is given by

$$\sigma^2(a,b) = a^2 \frac{\Gamma(3b)}{\Gamma(b)} \qquad . \tag{10}$$

For a Gaussian density, i.e. a shape parameter of b=0.5, eqs. (9) and (10) simplify to β_2=3 and $\sigma=\frac{1}{2}a\sqrt{2}$. For a Double Exponential density, eqs. (9) and (10) yield β_2=6 and $\sigma=a\sqrt{2}$.

Although the family of Generalized Laplace densities is already fairly flexible, it seemed useful to extend it somewhat by considering mixtures of different Generalized Laplace densities. Mixture densities have found wide applications in other fields as well (Ref. 13). Mixture densities are particularly useful when different subpopulations can be distinguished within the overall population. In the case of aircraft pairs data, a distinction between relatively good performance in the core and relatively bad performance in the tail would seem to be an appropriate choice.

Thus, an appropriate model for the pairs data consists of the family of the so-called Double Generalized Laplace probability densities, given in formula by

$$f_{DGL}(d) = (1-\alpha) \frac{1}{2a_1 b_1 \, \Gamma(b_1)} \, e^{-\left|\frac{d-\mu}{a_1}\right|^{1/b_1}} + \alpha \frac{1}{2a_2 b_2 \, \Gamma(b_2)} \, e^{-\left|\frac{d-\mu}{a_2}\right|^{1/b_2}} \quad (11)$$

The family of Double Generalized Laplace probability densities is charac-
terized by five parameters, in addition to the mean value μ, i.e. a
weighting factor α, two scale parameters and two shape parameters. It
includes the (single) Generalized Laplace density ($\alpha=0$ or $a_1=a_2$ and $b_1=b_2$)
as well as the Gaussian Double Exponential ($b_1=0.5$ and $b_2=1.0$) and the
Double Double Exponential ($b_1=b_2=1.0$) densities. Two more special cases
that are included are $b_1=b_2$, $a_1\neq a_2$, i.e. a mixture of two Generalized
Laplace densities with different scale parameters only and $a_1=a_2$, $b_1\neq b_2$,
i.e. a mixture of two differently shaped densities with the same scale
parameter. In principle, either of the two component parts in eq.(11) may
have the larger weighting factor. The descriptions core and tail density
will be used for the first and second density in eq.(11) respectively.

The variance of a random variable having a Double Generalized
Laplace probability density is given by

$$\sigma^2 = \sigma^2(\alpha, a_1, b_1, a_2, b_2) = (1-\alpha)\sigma_1^2 + \alpha\sigma_2^2 \quad (12)$$

where

$$\sigma_1^2 = \sigma^2(a_1, b_1) = a_1^2 \frac{\Gamma(3b_1)}{\Gamma(b_1)} \quad (13)$$

and

$$\sigma_2^2 = \sigma^2(a_2, b_2) = a_2^2 \frac{\Gamma(3b_2)}{\Gamma(b_2)} \quad (14)$$

The skewness of a Double Generalized Laplace distributed random variable
is easily shown to be zero, wheras its kurtosis is found to be given by

$$\beta_2(\alpha, a_1, b_1, a_2, b_2) = (1-\alpha) \frac{\sigma_1^4}{\sigma^4(\alpha, a_1, b_1, a_2, b_2)} \beta_2(b_1) + \alpha \frac{\sigma_2^4}{\sigma^4(\alpha, a_1, b_1, a_2, b_2)} \beta_2(b_2) \quad (15)$$

Notice that additional weighting factors $(\sigma_1/\sigma)^4$ and $(\sigma_2/\sigma)^4$ occur in
eq.(15). This is due to the normalization of the overall kurtosis by the

With a view to a common approach to the selection of an appropriate family of probability densities by the groups involved in the various data collection studies to investigate the possibility of a 1000 ft vertical separation standard, it is necessary to consider the families of probability densities being used elsewhere, notably in the United States and in Japan (Refs. 14-17). As is shown in Appendix B, there exists a large commonality between the families used by the various groups. Further studies in Europe on the basis of singles data also utilize the family of Double Generalized Laplace densities or some particular sub families (Ref. 18).

It is finally remarked that it may not be necessary to estimate all of the five parameters of the Double Generalized Laplace probability densities simultaneously from the aircraft pairs data. The shape parameter of the core and/or tail density, for example, might be fixed beforehand when this is deemed useful. In that way, particular sub families of the family of Double Generalized Laplace probability densities can be examined.

3.3   Selection of a parameter estimation technique
        Once a family of analytical probability density functions has been adopted, the next question to be answered is (recall Section 3.1):

ii)   Which criterion should be used to uniquely select one member from the family to represent the data?

Because different members of the same family are characterized by different parameter values, this question is essentially referring to a useful parameter estimation technique. Various (basic) methods are available within the statistical literature for estimating parameters of probability densities from experimental data, for example (Refs. 19,20):
    -   method of moments
    -   maximum likelihood estimation
    -   maximum a posteriori estimation
    -   minimum Chi-square estimation
    -   least squares estimation
Any of these methods may be extended or modified in order to obtain an estimation method which is even more useful for the application at hand.

In general, the choice of any particular estimation method depends on two factors, namely the statistical properties of the pertinent estimator and the computational complexity of the method. With respect to computational complexity holds that in many cases an analytical solution is not possible and that numerical solution techniques have to be used. Moreover, the computational complexity will generally increase with the number of parameters to be estimated. With respect to the statistical properties of the estimation methods mentioned above, it is first remarked that the least squares estimation method does not possess any general optimum properties, although it does have certain optimum properties for the "linear model" situation. The two most important estimation methods then are the method of moments and the maximum likelihood method, where the former is the older of the two. In some special cases the two methods give the same results. In many cases the maximum likelihood method is superior. Some advantages and disadvantages of the maximum likelihood method are mentioned in Reference 14. The first advantage mentioned there is that the maximum likelihood method provides probability statements about parameter estimates and that confidence regions may be computed (based on the asymptotic distribution of the ML estimator). As a second advantage the use of the likelihood ratio test for comparing the fit of different models is mentioned. According to Reference 14, the major disadvantage of the maximum likelihood method is that it appears to be less sensitive to the data structure within the tail portion of the data than to the structure of the core. In principle, however, this disadvantage may be overcome by weighting the data in the tail more heavily than those in the core. Moreover, the preliminary results presented in Reference 21 show a fairly close correspondence between the sample moments and the maximum likelihood estimates of the moments.

Taking the various observations ito account, the maximum likelihood method will be used to fit the family of Double Generalized Laplace probability densities to the pairs data as obtained in the European Vertical Data Collection. It is worthwhile to remark that the (preliminary) results of References 14 to 18 inclusive and Ref. 21 have also been obtained by means of the maximum likelihood method.

3.4   Evaluation criteria

As described in Section 3.1, the modelling approach to the estimation of the probability of vertical overlap requires a third step to evaluate the quality or fit of the resulting probability density model. This is to verify whether the theoretically well-defined results of the previous two steps meet the practical objectives of the study. In the case of the European Vertical Data Collection, the objective is to obtain the "best possible" estimate of the probability of vertical overlap, given the limitations of the data sample. The main limitations of the data sample described in Section two are that there are no data available in the region of real interest (from the collision risk estimation point of view), i.e. the region of vertical distances between aircraft in a pair of approximately zero, and only few data points in the connecting region of distances of the order of half the vertical separation standard. Due to these limitations, it may be necessary to build in into the modelling process some measures of cautiousness, in order to avoid underestimating the probability of vertical overlap. One such measure could be the use of a Double Exponential tail density within the mixture probability density models. As it is very difficult to predict in advance how the combination of the chosen family and the parameter estimation technique will perform with regard to the limitations mentioned above, an independent evaluation step is necessary.

The basic statistical tool for evaluating the fit of a proposed model is the statistical test, for example the well known χ^2-test and the Kolmogorov-Smirnov test. Although these may be useful for the overall evaluation of the fit, they may not be fully satisfactory for evaluating the fit in the extreme tail regions. Graphical means will, therefore, extensively be used to describe and analyze the residuals. Conclusions drawn from that may even overrule the conclusions drawn from the formal statistical test. Some judgement will probably be necessary for this evaluation.

Further insight into the properties of the best fitting probability density model may be obtained by considering the likelihood function about the estimated parameter values. For example, contour plots of the likelihood ratio or the second derivatives of the likelihood function may be used for establishing confidence bounds on the parameter estimates.

4   MAXIMUM LIKELIHOOD PARAMETER ESTIMATION OF DOUBLE GENERALIZED LAPLACE
     PROBABILITY DENSITIES

4.1   Introduction

In this section, an algorithm for the maximum likelihood method
applied to the family of Double Generalized Laplace probability densities
is derived. It consists of Newton's method for finding a maximum of the
logarithm of the likelihood function. Three aspects require particular
attention, viz. convergence of the algorithm, existence of more than one
local maximum and computation time.

The likelihood function and its logarithm, i.e. the log likelihood
function, are introduced in subsection 4.2. The necessary conditions for
an extremum are derived, together with the sufficient conditions for a
maximum of the log likelihood function. Because of the large number of
individual data points (n=4960, recall Section 2), the number of terms
involved in these equations is very large, i.e. of the order of some
multiples of five thousand, where almost any term involves powers, loga-
rithms or exponentials. As a result, the amount of computation time needed
is relatively large, certainly when several initial guesses of the para-
meter values have to be investigated in order to obtain convergence and/or
all the local maxima. Therefore, a likelihood function based on grouped
data is introduced in subsection 4.3. Grouping the data into class inter-
vals of e.g. 40 to 10 ft reduces the number of terms from a multiple of
five thousand to a multiple of 15 to 60 and the amount of computation time
accordingly. Grouping of the data introduces, of course, some error into
the resulting parameter estimates. However, when the class interval is
sufficiently small, this error should be sufficiently small as well.
Moreover, the resulting estimates based on grouped data may be used as
already fairly accurate initial guesses for the maximum likelihood para-
meter estimates based on the individual, non-grouped data points.

The convergence of the Newton iteration process and the existence of
more than a single local maximum of the log likelihood function for Double
Generalized Laplace densities are difficult to analyze theoretically. A
few special cases, however, are considered in subsection 4.4. Firstly,
three single densities, all being special cases belonging to the Double
Generalized Laplace family are analyzed. These are the Gaussian, the
Double Exponential and the Generalized Laplace densities. The Gaussian and

the Double Exponential case can be dealt with fully analytically, but Newton's method is already required for the case of a single Generalized Laplace density. Next, the Double Double Exponential sub family of the Double Generalized Laplace family will be considered. As a result of the lack of a full analysis for the general case, it may turn out that several initial guesses for the parameter values have to be investigated, in order to obtain convergence of the Newton process or to find all the solutions of the maximum likelihood equations.

4.2 The likelihood function

Let a sample of pairs data of size n be given on the assumption that the underlying probability distribution is Double Generalized Laplace. The likelihood function $L = L(\alpha, a_1, a_2, b_1, b_2)$ is defined by

$$L(\alpha, a_1, a_2, b_1, b_2) = \prod_{i=1}^{n} f(d_i) \tag{16}$$

where d_i, $i = 1, 2, \ldots, n$ denote the vertical distances between the aircraft in a pair and $f(d_i)$ is given by

$$f(d_i) = (1-\alpha) \frac{1}{2a_1 b_1 \Gamma(b_1)} e^{-\left|\frac{d_i - \mu}{a_1}\right|^{1/b_1}} + \alpha \frac{1}{2a_2 b_2 \Gamma(b_2)} e^{-\left|\frac{d_i - \mu}{a_2}\right|^{1/b_2}} \tag{17}$$

The log likelihood function $\mathcal{L} = \mathcal{L}(\alpha, a_1, a_2, b_1, b_2)$ is obtained by taking the (natural) logarithm of the likelihood function $L = L(\alpha, a_1, a_2, b_1, b_2)$. Hence,

$$\mathcal{L} = \mathcal{L}(\alpha, a_1, a_2, b_1, b_2) = \log L(\alpha, a_1, a_2, b_1, b_2) = \sum_{i=1}^{n} \log f(d_i) \tag{18}$$

or

$$\mathcal{L}(\alpha, a_1, a_2, b_1, b_2) = \sum_{i=1}^{n} \log \left\{ (1-\alpha) \frac{1}{2a_1 b_1 \Gamma(b_1)} e^{-\left|\frac{d_i - \mu}{a_1}\right|^{1/b_1}} + \alpha \frac{1}{2a_2 b_2 \Gamma(b_2)} e^{-\left|\frac{d_i - \mu}{a_2}\right|^{1/b_2}} \right\} \tag{19}$$

The main advantage of taking the logarithm of the likelihood function based on mixture densities is that the product as in eq.(16) is replaced by a summation as in eq.(19). When no confusion exists about the arguments of the (log) likelihood function, these will be suppressed.

The log likelihood function $\mathcal{L}$ depends on six parameters in total, i.e. on α, a_1, b_1, a_2 and b_2 on the one hand and the parameter μ on the other. The usual way to formulate the necessary conditions for an extremum of the function $\mathcal{L}$ is to take the partial derivatives with respect to the parameters and to equate these to zero. However, the variable $|d_i - \mu|$, $i=1,2,\ldots,n$ as a function of μ is not differentiable at the point $\mu = d_i$. Thus, obtaining the maximum of the log likelihood function over the parameter μ requires a different approach as can be used for the other parameters. To avoid this complication, μ will be treated as a known parameter. In fact, it will be taken to be $\mu = 1000$ ft, in conformity with the discussion of the aircraft pairs data in Section 2.

The necessary conditions for an extremum of the log likelihood function $\mathcal{L}$ now become

$$E_1 = \frac{\partial \mathcal{L}}{\partial \alpha} = \sum_{i=1}^{n} \frac{1}{f(d_i)} \frac{\partial}{\partial \alpha} f(d_i) = 0 \tag{20}$$

$$E_2 = \frac{\partial \mathcal{L}}{\partial a_1} = \sum_{i=1}^{n} \frac{1}{f(d_i)} \frac{\partial}{\partial a_1} f(d_i) = 0 \tag{21}$$

$$E_3 = \frac{\partial \mathcal{L}}{\partial a_2} = \sum_{i=1}^{n} \frac{1}{f(d_i)} \frac{\partial}{\partial a_2} f(d_i) = 0 \tag{22}$$

$$E_4 = \frac{\partial \mathcal{L}}{\partial b_1} = \sum_{i=1}^{n} \frac{1}{f(d_i)} \frac{\partial}{\partial b_1} f(d_i) = 0 \tag{23}$$

$$E_5 = \frac{\partial \mathcal{L}}{\partial b_2} = \sum_{i=1}^{n} \frac{1}{f(d_i)} \frac{\partial}{\partial b_2} f(d_i) = 0 \tag{24}$$

These equations are usually called the likelihood equations. The system of equations (20)-(24) is highly non-linear in the unknown parameters $\alpha, a_1, a_2, b_1, b_2$ and, therefore, needs to be solved numerically. This is done here by means of Newton iteration, i.e. by linearization of the system (20)-(24). The linearized version of the system reads:

$$\sum_{i=1}^{n} \frac{1}{f(d_i)} \left[\frac{\partial}{\partial \alpha} f(d_i) + \left(\frac{\partial^2}{\partial \alpha^2} f(d_i) - \frac{1}{f(d_i)} \frac{\partial}{\partial \alpha} f(d_i) \frac{\partial}{\partial \alpha} f(d_i)\right) \Delta \alpha + \right.$$

$$+ \left(\frac{\partial^2}{\partial \alpha \partial a_1} f(d_i) - \frac{1}{f(d_i)} \frac{\partial}{\partial a_1} f(d_i) \frac{\partial}{\partial \alpha} f(d_i)\right) \Delta a_1 +$$

$$+ \left(\frac{\partial^2}{\partial \alpha \partial a_2} f(d_i) - \frac{1}{f(d_i)} \frac{\partial}{\partial a_2} f(d_i) \frac{\partial}{\partial \alpha} f(d_i)\right) \Delta a_2 +$$

$$+ \left(\frac{\partial^2}{\partial \alpha \partial b_1} f(d_i) - \frac{1}{f(d_i)} \frac{\partial}{\partial b_1} f(d_i) \frac{\partial}{\partial \alpha} f(d_i)\right) \Delta b_1 +$$

$$\left. + \left(\frac{\partial^2}{\partial \alpha \partial b_2} f(d_i) - \frac{1}{f(d_i)} \frac{\partial}{\partial b_2} f(d_i)\frac{\partial}{\partial \alpha} f(d_i)\right) \Delta b_2 \right] = 0 \qquad (25)$$

$$\sum_{i=1}^{n} \frac{1}{f(d_i)} \left[\frac{\partial}{\partial a_1} f(d_i) + \left(\frac{\partial^2}{\partial \alpha \partial a_1} f(d_i) - \frac{1}{f(d_i)} \frac{\partial}{\partial \alpha} f(d_i) \frac{\partial}{\partial a_1} f(d_i)\right) \Delta \alpha + \right.$$

$$+ \left(\frac{\partial^2}{\partial a_1^2} f(d_i) - \frac{1}{f(d_i)} \frac{\partial}{\partial a_1} f(d_i) \frac{\partial}{\partial a_1} f(d_i)\right) \Delta a_1 +$$

$$+ \left(\frac{\partial^2}{\partial a_1 \partial a_2} f(d_i) - \frac{1}{f(d_i)} \frac{\partial}{\partial a_2} f(d_i) \frac{\partial}{\partial a_1} f(d_i)\right) \Delta a_2 +$$

$$+ \left(\frac{\partial^2}{\partial a_1 \partial b_1} f(d_i) - \frac{1}{f(d_i)} \frac{\partial}{\partial b_1} f(d_i) \frac{\partial}{\partial a_1} f(d_i)\right) \Delta b_1 +$$

$$\left. + \left(\frac{\partial^2}{\partial a_1 \partial b_2} f(d_i) - \frac{1}{f(d_i)} \frac{\partial}{\partial b_2} f(d_i)\frac{\partial}{\partial a_1} f(d_i)\right) \Delta b_2 \right] = 0 \qquad (26)$$

$$\sum_{i=1}^{n} \frac{1}{f(d_i)} \left[\frac{\partial}{\partial a_2} f(d_i) + \left(\frac{\partial^2}{\partial \alpha \partial a_2} f(d_i) - \frac{1}{f(d_i)} \frac{\partial}{\partial \alpha} f(d_i) \frac{\partial}{\partial a_2} f(d_i)\right) \Delta \alpha + \right.$$

$$+ \left(\frac{\partial^2}{\partial a_1 a_2} f(d_i) - \frac{1}{f(d_i)} \frac{\partial}{\partial a_1} f(d_i) \frac{\partial}{\partial a_2} f(d_i)\right) \Delta a_1 +$$

$$+ \left(\frac{\partial^2}{\partial a_2^2} f(d_i) - \frac{1}{f(d_i)} \frac{\partial}{\partial a_2} f(d_i) \frac{\partial}{\partial a_2} f(d_i)\right) \Delta a_2 +$$

$$+ \left(\frac{\partial^2}{\partial a_2 \partial b_1} f(d_i) - \frac{1}{f(d_i)} \frac{\partial}{\partial b_1} f(d_i) \frac{\partial}{\partial a_2} f(d_i)\right) \Delta b_1 +$$

$$\left. + \left(\frac{\partial^2}{\partial a_2 \partial b_2} f(d_i) - \frac{1}{f(d_i)} \frac{\partial}{\partial b_2} f(d_i)\frac{\partial}{\partial a_2} f(d_i)\right) \Delta b_2 \right] = 0 \qquad (27)$$

$$\sum_{i=1}^{n} \frac{1}{f(d_i)} \left[\frac{\partial}{\partial b_1} f(d_i) + \left(\frac{\partial^2}{\partial \alpha \partial b_1} f(d_i) - \frac{1}{f(d_i)} \frac{\partial}{\partial \alpha} f(d_i) \frac{\partial}{\partial b_1} f(d_i)\right) \Delta \alpha + \right.$$

$$+ \left(\frac{\partial^2}{\partial a_1 \partial b_1} f(d_i) - \frac{1}{f(d_i)} \frac{\partial}{\partial a_1} f(d_i) \frac{\partial}{\partial b_1} f(d_i) \right) \Delta a_1 +$$

$$+ \left(\frac{\partial^2}{\partial a_2 \partial b_1} f(d_i) - \frac{1}{f(d_i)} \frac{\partial}{\partial a_2} f(d_i) \frac{\partial}{\partial b_1} f(d_i) \right) \Delta a_2 +$$

$$+ \left(\frac{\partial^2}{\partial b_1^2} f(d_i) - \frac{1}{f(d_i)} \frac{\partial}{\partial b_1} f(d_i) \frac{\partial}{\partial b_1} f(d_i) \right) \Delta b_1 +$$

$$+ \left(\frac{\partial^2}{\partial b_1 \partial b_2} f(d_i) - \frac{1}{f(d_i)} \frac{\partial}{\partial b_2} f(d_i) \frac{\partial}{\partial b_1} f(d_i) \right) \Delta b_2 \Bigg] = 0 \tag{28}$$

$$\sum_{i=1}^{n} \frac{1}{f(d_i)} \Bigg[\frac{\partial}{\partial b_2} f(d_i) + \left(\frac{\partial^2}{\partial \alpha \partial b_2} f(d_i) - \frac{1}{f(d_i)} \frac{\partial}{\partial \alpha} f(d_i) \frac{\partial}{\partial b_2} f(d_i) \right) \Delta \alpha +$$

$$+ \left(\frac{\partial^2}{\partial a_1 \partial b_2} f(d_i) - \frac{1}{f(d_i)} \frac{\partial}{\partial a_1} f(d_i) \frac{\partial}{\partial b_2} f(d_i) \right) \Delta a_1 +$$

$$+ \left(\frac{\partial^2}{\partial a_2 \partial b_2} f(d_i) - \frac{1}{f(d_i)} \frac{\partial}{\partial a_2} f(d_i) \frac{\partial}{\partial b_2} f(d_i) \right) \Delta a_2 +$$

$$+ \left(\frac{\partial^2}{\partial b_1 \partial b_2} f(d_i) - \frac{1}{f(d_i)} \frac{\partial}{\partial b_1} f(d_i) \frac{\partial}{\partial b_2} f(d_i) \right) \Delta b_1 +$$

$$+ \left(\frac{\partial^2}{\partial b_2^2} f(d_i) - \frac{1}{f(d_i)} \frac{\partial}{\partial b_2} f(d_i) \frac{\partial}{\partial b_2} f(d_i) \right) \Delta b_2 \Bigg] = 0 \tag{29}$$

The various partial derivatives are given in detail in Appendix C. Symbolically, the linearized system (25)-(29) may be written as

$$\begin{pmatrix} M_1 & N_1 & O_1 & Q_1 & R_1 \\ M_2 & N_2 & O_2 & Q_2 & R_2 \\ M_3 & N_3 & O_3 & Q_3 & R_3 \\ M_4 & N_4 & O_4 & Q_4 & R_4 \\ M_5 & N_5 & O_5 & Q_5 & R_5 \end{pmatrix} \begin{pmatrix} \Delta \alpha \\ \Delta a_1 \\ \Delta a_2 \\ \Delta b_1 \\ \Delta b_2 \end{pmatrix} = - \begin{pmatrix} L_1 \\ L_2 \\ L_3 \\ L_4 \\ L_5 \end{pmatrix} \tag{30}$$

or as

$$(MNO) \begin{pmatrix} \Delta \alpha \\ \Delta a_1 \\ \Delta a_2 \\ \Delta b_1 \\ \Delta b_2 \end{pmatrix} = - \begin{pmatrix} L \end{pmatrix} \tag{31}$$

Each of the elements of the coefficient matrix and of the right-hand side vector in eq.(30) is to be evaluated for the current values α_i, a_{1_i}, a_{2_i}, b_{1_i}, b_{2_i} of the parameters of the Double Generalized Laplace probability densities. The system (30) may be solved with any routine for solving systems of linear equations, for instance with subroutine F04ATF from the NAG subroutine library which is based on Crout's factorisation method. Once the system of eqs.(30) has been solved, new parameter values are computed by adding the corrections $\Delta\alpha$, Δa_1, Δa_2, Δb_1, Δb_2 to the current values. The solution process is stopped when either

$$\max \{|E_1|,|E_2|,|E_3|,|E_4|,|E_5|\} < \varepsilon \qquad (32)$$

, with $\varepsilon = 1.0 \times 10^{-10}$, or the number of iterations exceeds a specified limit, usually 25.

Any solution of the likelihood equations does not necessarily define a maximum of the log likelihood function. In order that a solution does define a maximum, it is sufficient that the matrix of second partial derivatives of the log likelihood function be negative definite. It turns out that this matrix is just the coefficient matrix MNO of the system of eqs.(30).

So far, it has been assumed that the complete set of parameters $\alpha, a_1, a_2, b_1, b_2$ had to be determined. In fact, any subset of parameters may be determined, on the assumption that suitable values of the remaining parameters are known, by taking the appropriate rows and columns from eqs.(30). The following selection was made for the purpose of the present investigations:

case	parameters to be estimated	parameters known
S	$\alpha, a_1, a_2, b_1, b_2$	none
S_4	α, a_1, a_2, b_1	b_2
S_3	α, a_1, a_2	b_1, b_2
S_2	α, a_1	a_2, b_1, b_2

(33)

Once the maximum likelihood parameter (point) estimates have been determined, it is of interest to have a closer look at the likelihood function or the log likelihood function at and about the point obtained in the parameter space. Firstly, the value of the likelihood function obtained for the maximum likelihood estimates of the parameters is of interest. This value may be expressed as

$$L(\alpha,\hat{a}_1,\hat{a}_2,\hat{b}_1,\hat{b}_2) = \prod_{i=1}^{n} \left[\text{Prob} \{\underline{d}_i \in [d_i, d_i + \Delta d_i) \mid E\{\underline{d}_i\}=\mu\} \,/\, \Delta d_i \right]$$

$$= \prod_{i=1}^{n} (1-\alpha) \frac{1}{2\hat{a}_1\hat{b}_1\Gamma(\hat{b}_1)} e^{-\left|\frac{d_i-\mu}{\hat{a}_1}\right|^{1/\hat{b}_1}} + \alpha \frac{1}{2\hat{a}_2\hat{b}_2\Gamma(\hat{b}_2)} e^{-\left|\frac{d_i-\mu}{\hat{a}_2}\right|^{1/\hat{b}_2}} \qquad (34)$$

It is a relative maximum value in the sense that it is always smaller than the value that would be obtained for the same parameter values, but given that the mean value of each observation $\underline{d}_i$ (i=1,2,...,n) had been exactly the observed value d_i. Hence, the absolute maximum value of the likelihood function may be expressed as

$$L_2(\alpha,\hat{a}_1,\hat{a}_2,\hat{b}_1,\hat{b}_2) = \prod_{i=1}^{n} \left[\text{Prob} \{\underline{d}_i \in [d_i, d_i + \Delta d_i) \mid E\{\underline{d}_i\}=d_i\} \,/\, \Delta d_i \right]$$

$$= \prod_{i=1}^{n} \left[(1-\alpha) \frac{1}{2\hat{a}_1\hat{b}_1\Gamma(\hat{b}_1)} + \alpha \frac{1}{2\hat{a}_2\hat{b}_2\Gamma(\hat{b}_2)} \right] \qquad (35)$$

The standardized log likelihood value is defined by

$$(\ell_2 - \ell)(\alpha,\hat{a}_1,\hat{a}_2,\hat{b}_1,\hat{b}_2) = \log L_2(\alpha,\hat{a}_1,\hat{a}_2,\hat{b}_1,\hat{b}_2) - \log L(\alpha,\hat{a}_1,\hat{a}_2,\hat{b}_1,\hat{b}_2)$$

$$= \sum_{i=1}^{n} \log \frac{(1-\alpha)\frac{1}{2\hat{a}_1\hat{b}_1\Gamma(\hat{b}_1)} + \alpha \frac{1}{2\hat{a}_2\hat{b}_2\Gamma(\hat{b}_2)}}{(1-\alpha)\frac{1}{2\hat{a}_1\hat{b}_1\Gamma(\hat{b}_1)} e^{-\left|\frac{d_i-\mu}{\hat{a}_1}\right|^{1/\hat{b}_1}} + \alpha \frac{1}{2\hat{a}_2\hat{b}_2\Gamma(\hat{b}_2)} e^{-\left|\frac{d_i-\mu}{\hat{a}_2}\right|^{1/\hat{b}_2}}} \qquad (36)$$

Secondly, the likelihood function in a neighbourhood of the point in the parameter space defined by the maximum likelihood estimates may be considered. More specifically, the relative likelihood function $R(\alpha, a_1, a_2, b_1, b_2)$ is defined by

$$R(\alpha, a_1, a_2, b_1, b_2) = \frac{L(\alpha, a_1, a_2, b_1, b_2)}{L(\hat{\alpha}, \hat{a}_1, \hat{a}_2, \hat{b}_1, \hat{b}_2)} \tag{37}$$

and the likelihood ratio statistic by

$$\Lambda(\alpha, a_1, a_2, b_1, b_2) = -2 \log R(\alpha, a_1, a_2, b_1, b_2) \tag{38}$$

Asymptotically, the likelihood ratio statistic has a χ^2-distribution with 5 degrees of freedom (Refs. 22,23). It is difficult to visualize the relative likelihood function and the likelihood ratio in a five-dimensional space. As a solution, contours in the planes defined by any two out of the five parameters are usually drawn. For the remaining parameters, the maximum likelihood point estimates are substituted. The likelihood ratio statistic can be used for constructing confidence bounds on the parameters, because its asymptotic distribution is χ^2. A joint 95% confidence region for all of the five parameters of the Double Generalized Laplace probability density is defined by the relation

$$R(\alpha, a_1, a_2, b_1, b_2) = e^{-\frac{1}{2}\chi^2_{.95,(5)}} \tag{39}$$

where $\chi^2_{.95,(5)}$ denotes the 95 percentile of a variable having a χ^2-distribution with five degrees of freedom. Similarly, confidence regions for any subset of two parameters, e.g. α and a_1 may be defined by the relation

$$R(\alpha, a_1, \hat{a}_2, \hat{b}_1, \hat{b}_2) = e^{-\frac{1}{2}\chi^2_{.95,(2)}} \tag{40}$$

where $\hat{a}_2$, $\hat{b}_1$, and $\hat{b}_2$ again denote the maximum likelihood point estimates of the pertinent parameters and where $\chi^2_{.95,(2)}$ denotes the 95 percentile of a variable having a χ^2-distribution with two degrees of freedom.

Finally, the asymptotic distribution of the maximum likelihood estimator itself should be mentioned. It is well known that, under certain regularity conditions, the maximum likelihood estimator is jointly Gaussian distributed with covariance matrix given by the inverse of Fisher's information matrix. More specifically,

$$
\text{cov}\left\{ \begin{pmatrix} \hat{\underline{\alpha}} \\ \hat{\underline{a}}_1 \\ \hat{\underline{a}}_2 \\ \hat{\underline{b}}_1 \\ \hat{\underline{b}}_2 \end{pmatrix} \begin{pmatrix} \hat{\underline{\alpha}} \\ \hat{\underline{a}}_1 \\ \hat{\underline{a}}_2 \\ \hat{\underline{b}}_1 \\ \hat{\underline{b}}_2 \end{pmatrix}^T \right\} =
$$

$$
- \begin{pmatrix}
E\left\{\frac{\partial^2}{\partial\alpha^2}\underline{\mathcal{L}}\right\} & E\left\{\frac{\partial^2}{\partial\alpha\partial a_1}\underline{\mathcal{L}}\right\} & E\left\{\frac{\partial^2}{\partial\alpha\partial a_2}\underline{\mathcal{L}}\right\} & E\left\{\frac{\partial^2}{\partial\alpha\partial b_1}\underline{\mathcal{L}}\right\} & E\left\{\frac{\partial^2}{\partial\alpha\partial b_2}\underline{\mathcal{L}}\right\} \\
E\left\{\frac{\partial^2}{\partial\alpha\partial a_1}\underline{\mathcal{L}}\right\} & E\left\{\frac{\partial^2}{\partial a_1^2}\underline{\mathcal{L}}\right\} & E\left\{\frac{\partial^2}{\partial a_1\partial a_2}\underline{\mathcal{L}}\right\} & E\left\{\frac{\partial^2}{\partial a_1\partial b_1}\underline{\mathcal{L}}\right\} & E\left\{\frac{\partial^2}{\partial a_1\partial b_2}\underline{\mathcal{L}}\right\} \\
E\left\{\frac{\partial^2}{\partial\alpha\partial a_2}\underline{\mathcal{L}}\right\} & E\left\{\frac{\partial^2}{\partial a_1\partial a_2}\underline{\mathcal{L}}\right\} & E\left\{\frac{\partial^2}{\partial a_2^2}\underline{\mathcal{L}}\right\} & E\left\{\frac{\partial^2}{\partial a_2\partial b_1}\underline{\mathcal{L}}\right\} & E\left\{\frac{\partial^2}{\partial a_2\partial b_2}\underline{\mathcal{L}}\right\} \\
E\left\{\frac{\partial^2}{\partial\alpha\partial b_1}\underline{\mathcal{L}}\right\} & E\left\{\frac{\partial^2}{\partial a_1\partial b_1}\underline{\mathcal{L}}\right\} & E\left\{\frac{\partial^2}{\partial a_2\partial b_1}\underline{\mathcal{L}}\right\} & E\left\{\frac{\partial^2}{\partial b_1^2}\underline{\mathcal{L}}\right\} & E\left\{\frac{\partial^2}{\partial b_1\partial b_2}\underline{\mathcal{L}}\right\} \\
E\left\{\frac{\partial^2}{\partial\alpha\partial b_2}\underline{\mathcal{L}}\right\} & E\left\{\frac{\partial^2}{\partial a_1\partial b_2}\underline{\mathcal{L}}\right\} & E\left\{\frac{\partial^2}{\partial a_2\partial b_2}\underline{\mathcal{L}}\right\} & E\left\{\frac{\partial^2}{\partial b_1\partial b_2}\underline{\mathcal{L}}\right\} & E\left\{\frac{\partial^2}{\partial b_2^2}\underline{\mathcal{L}}\right\}
\end{pmatrix}^{-1} \quad (41)
$$

In this equation, the vector $(\underline{\alpha},\hat{\underline{a}}_1,\hat{\underline{a}}_2,\hat{\underline{b}}_1,\hat{\underline{b}}_2)^T$ denotes the maximum likelihood estimator of the parameter vector $(\alpha,a_1,a_2,b_1,b_2)^T$ and $\underline{\mathcal{L}}$ denotes the log likelihood function defined by eq.(18) with the maximum likelihood estimator as argument vector. Because of the occurrence of the expectation operator $E\{.\}$ in the matrix to be inverted in the right-hand side of eq.(41), that matrix is called the matrix of _expected_ information. As may be seen from Appendix C, the mean values are not easily expressed in closed analytical form. A useful approximation then is the matrix of _observed_ information, i.e.

$$
\text{cov}\left\{ \begin{pmatrix} \hat{\underline{\alpha}} \\ \hat{\underline{a}}_1 \\ \hat{\underline{a}}_2 \\ \hat{\underline{b}}_1 \\ \hat{\underline{b}}_2 \end{pmatrix} \begin{pmatrix} \hat{\underline{\alpha}} \\ \hat{\underline{a}}_1 \\ \hat{\underline{a}}_2 \\ \hat{\underline{b}}_1 \\ \hat{\underline{b}}_2 \end{pmatrix}^T \right\} = - \left(FOI(\hat{\underline{\alpha}},\hat{\underline{a}}_1,\hat{\underline{a}}_2,\hat{\underline{b}}_1,\hat{\underline{b}}_2) \right)^{-1} \quad (42)
$$

where the matrix $FOI(\alpha, a_1, a_2, b_1, b_2)$ is defined by

$$
\begin{pmatrix}
\dfrac{\partial^2 \ell}{\partial \alpha^2} & \dfrac{\partial^2 \ell}{\partial \alpha \partial a_1} & \dfrac{\partial^2 \ell}{\partial \alpha \partial a_2} & \dfrac{\partial^2 \ell}{\partial \alpha \partial b_1} & \dfrac{\partial^2 \ell}{\partial \alpha \partial b_2} \\[2ex]
\dfrac{\partial^2 \ell}{\partial \alpha \partial a_1} & \dfrac{\partial^2 \ell}{\partial a_1^2} & \dfrac{\partial^2 \ell}{\partial a_1 \partial a_2} & \dfrac{\partial^2 \ell}{\partial a_1 \partial b_1} & \dfrac{\partial^2 \ell}{\partial a_1 \partial b_2} \\[2ex]
\dfrac{\partial^2 \ell}{\partial \alpha \partial a_2} & \dfrac{\partial^2 \ell}{\partial a_1 \partial a_2} & \dfrac{\partial^2 \ell}{\partial a_2^2} & \dfrac{\partial^2 \ell}{\partial a_2 \partial b_1} & \dfrac{\partial^2 \ell}{\partial a_2 \partial b_2} \\[2ex]
\dfrac{\partial^2 \ell}{\partial \alpha \partial b_1} & \dfrac{\partial^2 \ell}{\partial a_1 \partial b_1} & \dfrac{\partial^2 \ell}{\partial a_2 \partial b_1} & \dfrac{\partial^2 \ell}{\partial b_1^2} & \dfrac{\partial^2 \ell}{\partial b_1 \partial b_2} \\[2ex]
\dfrac{\partial^2 \ell}{\partial \alpha \partial b_2} & \dfrac{\partial^2 \ell}{\partial a_1 \partial b_2} & \dfrac{\partial^2 \ell}{\partial a_2 \partial b_2} & \dfrac{\partial^2 \ell}{\partial b_1 \partial b_2} & \dfrac{\partial^2 \ell}{\partial b_2^2}
\end{pmatrix} \tag{43}
$$

with $\ell = \ell(\alpha, a_1, a_2, b_1, b_2)$ defined by eqs.(18) and (19). It should be noticed that the matrix $FOI(\alpha, a_1, a_2, b_1, b_2)$ is exactly equal to the coefficient matrix MNO defined by eqs. (30) and (31) and resulting from the linearization of the likelihood equations (20)-(24). Recall that the matrix MNO should be negative definite for a solution of the likelihood equations to define a maximum of the log likelihood function. Hence, taking the minus sign in equation (42) into account, a positive definite asymptotic covariance matrix results.

In the foregoing paragraphs, the standardized log likelihood value, the relative likelihood function, the likelihood ratio statistic and the asymptotic distribution all have been discussed for case S of eq.(33) i.e. for the case of five unknown parameters. The discussion can be adapted straightforwardly to the cases S_4, S_3, and S_2 by deleting the pertinent known parameters from eqs.(34) to (43).

4.3   The likelihood function based on grouped data

Because the number of data points d_i to be used in the maximum likelihood method as described in Section 4.2 is so large, it is desirable to have a method with which the results of the former method can be approximated fairly accurately, and which is much more efficient from the point of view of computation time needed. The method to be described here is based on grouping the individual data points into classes as a realization of a multinomial probability distribution. As a result, the

summation over the number of individual data points reduces to a summation over the number of classes considered.

Consider a subdivision of the vertical distance d-axis as shown in Figure 7. The class interval is denoted as w and the number of classes is m. Let the random variable $\underline{d}$ denote the vertical distance between the aircraft in a pair. On the assumption that the probability density of $\underline{d}$ is Double Generalized Laplace, it follows that

$$p_i = \text{Prob}\{\underline{d} \in [\mu+(i-1)w, \mu+iw) \cup (\mu-iw, \mu-(i-1)w]\} =$$

$$= 2\ \text{Prob}\{\underline{d} \in [\mu+(i-1),w, \mu+iw)\}$$

$$= 2 \int_{\mu+(i-1)w}^{\mu+iw} f_{DGL}(x)\,dx \qquad , \ i=1,2,\ldots,m \qquad (44)$$

Let $\underline{x}_i$ denote the number of vertical distances observed in class i in a sample of size n, class i being the conjunction of the sets $[\mu+(i-1)w,$ $\mu+iw)$ and $(\mu-iw, \mu-(i-1)w]$, for $i=1,2,\ldots,m$. The variable $\underline{x}_i$ then is binomially distributed with parameters n and p_i. In order to keep the variables p_i as a function of the parameters $\alpha, a_1, a_2, b_1, b_2$ of the Double Generalized Laplace probability densities as simple as possible, it is useful to use the following approximation:

$$p_i \approx 2w\ f_{DGL}\ (\mu+(i-\tfrac{1}{2})w) \qquad (45)$$

or

$$p_i \approx 2w \left\{ (1-\alpha)\ \frac{1}{2a_1 b_1 \Gamma(b_1)}\ e^{-\left(\frac{(i-\frac{1}{2})w}{a_1}\right)^{1/b_1}} + \alpha\ \frac{1}{2a_2 b_2 \Gamma(b_2)}\ e^{-\left(\frac{(i-\frac{1}{2})w}{a_2}\right)^{1/b_2}} \right\} \qquad (46)$$
$$, \ i=1,2,\ldots,m$$

Thus, the probabilities p_i, $i=1,2,\ldots,m$ are proportional with the values of the Double Generalized Laplace probability density in the centres of the classes.

Next, consider the joint probability distribution of the random variables $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_m$. It is clear that their distribution is multinomial with parameters $p_1, p_2, \ldots, p_m$ and n. Hence,

$$\text{Prob}\{\underline{x}_1 = x_1, \ \underline{x}_2 = x_2, \ \ldots, \ \underline{x}_m = x_m\} = \frac{n!}{x_1! x_2! \ldots x_m!}\ p_1^{x_1} p_2^{x_2} \ldots p_m^{x_m} \qquad (47)$$

with the probabilities p_i, i=1,2,...,m given by eq.(46) and with

$$\sum_{i=1}^{m} x_i = n \tag{48}$$

and

$$\sum_{i=1}^{m} p_i \approx 1 \tag{49}$$

The likelihood function $L = L(\alpha, a_1, a_2, b_1, b_2)$ is defined by

$$L = L(\alpha, a_1, a_2, b_1, b_2) = \frac{n!}{x_1! x_2! \ldots x_m!} p_1^{x_1} p_2^{x_2} \ldots p_m^{x_m} \tag{50}$$

and the log likelihood function $\mathcal{L} = \mathcal{L}(\alpha, a_1, a_2, b_1, b_2) = \log L$ becomes

$$\mathcal{L} = \mathcal{L}(\alpha, a_1, a_2, b_1, b_2) = \log \frac{n!}{x_1! x_2! \ldots x_m!} + \sum_{i=1}^{m} x_i \log p_i \tag{51}$$

Using eq.(46) and defining

$$A = \log \frac{n!}{x_1! x_2! \ldots x_m!} \tag{52}$$

it is found that

$$\mathcal{L} = \mathcal{L}(\alpha, a_1, a_2, b_1, b_2) = A + n \log 2w +$$

$$+ \sum_{i=1}^{m} x_i \log (1-\alpha) \frac{1}{2a_1 b_1 \Gamma(b_1)} e^{-\left(\frac{(i-\frac{1}{2})w}{a_1}\right)^{1/b_1}} + \alpha \frac{1}{2a_2 b_2 \Gamma(b_2)} e^{-\left(\frac{(i-\frac{1}{2})w}{a_2}\right)^{1/b_2}} \tag{53}$$

Eq.(53) defines the log likelihood function for grouped data, with the probability of an outcome in any of the classes being determined by the Double Generalized Laplace probability density.

It is useful to make a comparison between the likelihood functions for non-grouped data, eq.(19), and for grouped data, eq.(53). In fact, only the parameter dependent part, i.e. the last expression, of eq.(53) needs to be considered. When the class interval w is small, the quantity $(i-\frac{1}{2})w$ is a good approximation for $|d_j - \mu|$ for any individual data point j of class i. When, in addition, the factor x_i is written as a sum over all

the individual data points j of class i, and this summation is further summed over all the classes, effectively a summation over the individual data points as in eq.(19) will result. Thus, it may be expected that parameter estimates obtained by maximizing the log likelihood function for grouped data, eq.(53), are a good approximation of the parameter estimates obtained by maximizing the original log likelihood function, provided that the class interval w is sufficiently small.

The necessary conditions for an extremum of the log likelihood function for grouped data become, similar to eqs.(20)-(24):

$$E_1 = \frac{\partial \mathcal{L}}{\partial \alpha} = \sum_{i=1}^{m} \frac{x_i}{P_i} \frac{\partial p_i}{\partial \alpha} = 0 \qquad (54)$$

$$E_2 = \frac{\partial \mathcal{L}}{\partial a_1} = \sum_{i=1}^{m} \frac{x_i}{P_i} \frac{\partial p_i}{\partial a_1} = 0 \qquad (55)$$

$$E_3 = \frac{\partial \mathcal{L}}{\partial a_2} = \sum_{i=1}^{m} \frac{x_i}{P_i} \frac{\partial p_i}{\partial a_2} = 0 \qquad (56)$$

$$E_4 = \frac{\partial \mathcal{L}}{\partial b_1} = \sum_{i=1}^{m} \frac{x_i}{P_i} \frac{\partial p_i}{\partial b_1} = 0 \qquad (57)$$

$$E_5 = \frac{\partial \mathcal{L}}{\partial b_2} = \sum_{i=1}^{m} \frac{x_i}{P_i} \frac{\partial p_i}{\partial b_2} = 0 \qquad (58)$$

These likelihood equations are again non-linear in the parameters $\alpha, a_1, a_2, b_1, b_2$, and are solved, therefore, numerically by means of Newton's method. Their linearized versions read:

$$\sum_{i=1}^{m} \frac{x_i}{P_i} \left[\frac{\partial p_i}{\partial \alpha} + \left(\frac{\partial^2 p_i}{\partial \alpha^2} - \frac{1}{P_i} \frac{\partial p_i}{\partial \alpha} \frac{\partial p_i}{\partial \alpha} \right) \Delta\alpha + \right.$$

$$+ \left(\frac{\partial^2 p_i}{\partial \alpha \partial a_1} - \frac{1}{P_i} \frac{\partial p_i}{\partial a_1} \frac{\partial p_i}{\partial \alpha} \right) \Delta a_1 + \left(\frac{\partial^2 p_i}{\partial \alpha \partial a_2} - \frac{1}{P_i} \frac{\partial p_i}{\partial a_2} \frac{\partial p_i}{\partial \alpha} \right) \Delta a_2 +$$

$$\left. + \left(\frac{\partial^2 p_i}{\partial \alpha \partial b_1} - \frac{1}{P_i} \frac{\partial p_i}{\partial b_1} \frac{\partial p_i}{\partial \alpha} \right) \Delta b_1 + \left(\frac{\partial^2 p_i}{\partial \alpha \partial b_2} - \frac{1}{P_i} \frac{\partial p_i}{\partial b_2} \frac{\partial p_i}{\partial \alpha} \right) \Delta b_2 \right] = 0 \qquad (59)$$

$$\sum_{i=1}^{m} \frac{x_i}{p_i} \left[\frac{\partial p_i}{\partial a_1} + \left(\frac{\partial^2 p_i}{\partial \alpha \partial a_1} - \frac{1}{p_i} \frac{\partial p_i}{\partial \alpha} \frac{\partial p_i}{\partial a_1} \right) \Delta\alpha + \right.$$

$$+ \left(\frac{\partial^2 p_i}{\partial a_1^2} - \frac{1}{p_i} \frac{\partial p_i}{\partial a_1} \frac{\partial p_i}{\partial a_1} \right) \Delta a_1 + \left(\frac{\partial^2 p_i}{\partial a_1 \partial a_2} - \frac{1}{p_i} \frac{\partial p_i}{\partial a_2} \frac{\partial p_i}{\partial a_1} \right) \Delta a_2 +$$

$$+ \left. \left(\frac{\partial^2 p_i}{\partial a_1 \partial b_1} - \frac{1}{p_i} \frac{\partial p_i}{\partial b_1} \frac{\partial p_i}{\partial a_1} \right) \Delta b_1 + \left(\frac{\partial^2 p_i}{\partial a_1 \partial b_2} - \frac{1}{p_i} \frac{\partial p_i}{\partial b_2} \frac{\partial p_i}{\partial a_1} \right) \Delta b_2 \right] = 0 \quad (60)$$

$$\sum_{i=1}^{m} \frac{x_i}{p_i} \left[\frac{\partial p_i}{\partial a_2} + \left(\frac{\partial^2 p_i}{\partial \alpha \partial a_2} - \frac{1}{p_i} \frac{\partial p_i}{\partial \alpha} \frac{\partial p_i}{\partial a_2} \right) \Delta\alpha + \right.$$

$$+ \left(\frac{\partial^2 p_i}{\partial a_1 \partial a_2} - \frac{1}{p_i} \frac{\partial p_i}{\partial a_1} \frac{\partial p_i}{\partial a_2} \right) \Delta a_1 + \left(\frac{\partial^2 p_i}{\partial a_2^2} - \frac{1}{p_i} \frac{\partial p_i}{\partial a_2} \frac{\partial p_i}{\partial a_2} \right) \Delta a_2 +$$

$$+ \left. \left(\frac{\partial^2 p_i}{\partial a_2 \partial b_1} - \frac{1}{p_i} \frac{\partial p_i}{\partial b_1} \frac{\partial p_i}{\partial a_2} \right) \Delta b_1 + \left(\frac{\partial^2 p_i}{\partial a_2 \partial b_2} - \frac{1}{p_i} \frac{\partial p_i}{\partial b_2} \frac{\partial p_i}{\partial a_2} \right) \Delta b_2 \right] = 0 \quad (61)$$

$$\sum_{i=1}^{m} \frac{x_i}{p_i} \left[\frac{\partial p_i}{\partial b_1} + \left(\frac{\partial^2 p_i}{\partial \alpha \partial b_1} - \frac{1}{p_i} \frac{\partial p_i}{\partial \alpha} \frac{\partial p_i}{\partial b_1} \right) \Delta\alpha + \right.$$

$$+ \left(\frac{\partial^2 p_i}{\partial a_1 \partial b_1} - \frac{1}{p_i} \frac{\partial p_i}{\partial a_1} \frac{\partial p_i}{\partial b_1} \right) \Delta a_1 + \left(\frac{\partial^2 p_i}{\partial a_2 \partial b_1} - \frac{1}{p_i} \frac{\partial p_i}{\partial a_2} \frac{\partial p_i}{\partial b_1} \right) \Delta a_2 +$$

$$+ \left. \left(\frac{\partial^2 p_i}{\partial b_1^2} - \frac{1}{p_i} \frac{\partial p_i}{\partial b_1} \frac{\partial p_i}{\partial b_1} \right) \Delta b_1 + \left(\frac{\partial^2 p_i}{\partial b_1 \partial b_2} - \frac{1}{p_i} \frac{\partial p_i}{\partial b_2} \frac{\partial p_i}{\partial b_1} \right) \Delta b_2 \right] = 0 \quad (62)$$

$$\sum_{i=1}^{m} \frac{x_i}{p_i} \left[\frac{\partial p_i}{\partial b_2} + \left(\frac{\partial^2 p_i}{\partial \alpha \partial b_2} - \frac{1}{p_i} \frac{\partial p_i}{\partial \alpha} \frac{\partial p_i}{\partial b_2} \right) \Delta\alpha + \right.$$

$$+ \left(\frac{\partial^2 p_i}{\partial a_1 \partial b_2} - \frac{1}{p_i} \frac{\partial p_i}{\partial a_1} \frac{\partial p_i}{\partial b_2} \right) \Delta a_1 + \left(\frac{\partial^2 p_i}{\partial a_2 \partial b_2} - \frac{1}{p_i} \frac{\partial p_i}{\partial a_2} \frac{\partial p_i}{\partial b_2} \right) \Delta a_2 +$$

$$+ \left. \left(\frac{\partial^2 p_i}{\partial b_1 \partial b_2} - \frac{1}{p_i} \frac{\partial p_i}{\partial b_1} \frac{\partial p_i}{\partial b_2} \right) \Delta b_1 + \left(\frac{\partial^2 p_i}{\partial b_2^2} - \frac{1}{p_i} \frac{\partial p_i}{\partial b_2} \frac{\partial p_i}{\partial b_2} \right) \Delta b_2 \right] = 0 \quad (63)$$

The partial derivatives in the eqs.(59)-(63) are very similar to those in eqs.(25)-(29). In fact, they can immediately be obtained from the expressions given in Appendix C by the following steps: (i) replace $f(d_i)$ by p_i in the left-hand sides of eqs. (C.1) to (C.20) inclusive, (ii) replace $d_i - \mu$ by $(i-\frac{1}{2})w$, and (iii) add a factor 2w to the right-hand side of each of the eqs.(C.1) - (C.20). The system of eqs.(59)-(63) may again symbolically be written as (compare eqs.(30) and (31)):

$$(\text{MNO}) \begin{pmatrix} \Delta\alpha \\ \Delta a_1 \\ \Delta a_2 \\ \Delta b_1 \\ \Delta b_2 \end{pmatrix} = - \begin{pmatrix} L \\ \\ \end{pmatrix} \qquad (64)$$

and be solved by means of the NAG library subroutine FO4ATF until the desired precision is obtained.

From hereon, the same procedure as for the likelihood function based on the original non-grouped data can be followed. The negative definiteness of the matrix of second partial derivatives of the log likelihood function for grouped data needs to be verified first for any solution of the likelihood equations. Various subsets of parameters may be estimated, keeping the remaining parameters fixed at predetermined values. Similar to (33), this leads to the cases G, G_4, G_3 and G_2 according to the following table:

case	parameters to be estimated	parameters known
G	$\alpha, a_1, a_2, b_1, b_2$	none
G_4	α, a_1, a_2, b_1	b_2
G_3	α, a_1, a_2	b_1, b_2
G_2	α, a_1	a_2, b_1, b_2

(65)

The calculation of the standardized log likelihood value based on grouped data proceeds as follows. Firstly, compute

$$L(\hat{\alpha}, \hat{a}_1, \hat{a}_2, \hat{b}_1, \hat{b}_2) = \text{Prob } \{\underline{x}_1 = x_1, \underline{x}_2 = x_2, \ldots, \underline{x}_m = x_m \mid E\{\underline{x}_i\} = np_i, \ i = 1, 2, \ldots, m\} =$$

$$= \frac{n!}{x_1! x_2! \ldots x_m!} \, p_1^{x_1} \, p_2^{x_2} \, \ldots \, p_m^{x_m} \qquad (66)$$

In this equation, $p_1, p_2, \ldots, p_m$ are computed by means of eq.(46) using the maximum likelihood parameter estimates $\hat{\alpha}, \hat{a}_1, \hat{a}_2, \hat{b}_1, \hat{b}_2$. Next, compute

$$L_2(\hat{\alpha},\hat{a}_1,\hat{a}_2,\hat{b}_1,\hat{b}_2) = \text{Prob}\{\underline{x}_1=x_1,\underline{x}_2=x_2,\ldots,\underline{x}_m=x_m \,|\, E\{\underline{x}_i\} = x_i\} =$$

$$= \frac{n!}{x_1!x_2!\ldots x_m!} \left(\frac{x_1}{n}\right)^{x_1} \left(\frac{x_2}{n}\right)^{x_2} \ldots \left(\frac{x_m}{n}\right)^{x_m} \tag{67}$$

The standardized log likelihood value is now defined by

$$(\mathcal{L}_2-\mathcal{L})(\hat{\alpha},\hat{a}_1,\hat{a}_2,\hat{b}_1,\hat{b}_2) = \log L_2 - \log L = \sum_{i=1}^{m} x_i \log \frac{x_i}{np_i} \tag{68}$$

Finally, eqs.(37) to (43) inclusive dealing with the relative likelihood function, the likelihood ratio statistic and the asymptotic covariance matrix of the maximum likelihood estimator are directly applicable to the (log) likelihood function for grouped data as defined by eqs.(50) and (51).

4.4 Maximum likelihood estimation for some particular sub families of probability densities

As follows from the two foregoing Sections 4.2 and 4.3, the parameter estimation problem resulting from the maximum likelihood method applied to Double Generalized Laplace densities is very complex. This applies to the non-grouped data case as well as the grouped data case. For the latter, less computational effort is needed due to the much smaller number of terms in the equations. Although some subcases (S_4-S_2, G_4-G_2) have been defined for which the parameter estimation problem is somewhat less complex, each of these subcases still involves mixture densities and it is the mixture that highly contributes to the complexity. Hence, it is worthwhile to consider some sub families of single densities as well. This is useful, firstly, because the resulting system of maximum likelihood equations is more likely to be accessible to analysis. Secondly, it offers the possibility to investigate the improvement that can be obtained by using mixture densities rather than single ones. The first sub family to be considered is that of the Generalized Laplace densities, characterized by both a scale parameter and a shape parameter. Next, the sub families of Gaussian and Double Exponential densities, each characterized by a scale parameter only, are investigated. Finally, a particular sub family of mixture densities is considered as well, viz. the family of Double Double Exponential densities.

4.4.1 Generalized Laplace densities

The family of Generalized Laplace densities can be obtained from the family of mixtures of Generalized Laplace densities by putting the weighting factor α equal to zero. Both the non-grouped data and the grouped data case are considered below.

Consider a sample of non-grouped data points d_i, $i=1,2,\ldots,n$. On the assumption that the underlying probability distribution is Generalized Laplace, the likelihood function is given by

$$L = L(a,b) = \prod_{i=1}^{n} \frac{1}{2ab\Gamma(b)} e^{-\left|\frac{d_i-\mu}{a}\right|^{1/b}} \tag{69}$$

and the log likelihood function by

$$\mathcal{L} = \mathcal{L}(a,b) = -n \log 2ab \, \Gamma(b) - \sum_{i=1}^{n} \left|\frac{d_i-\mu}{a}\right|^{1/b} \tag{70}$$

Maximum likelihood estimates of the parameters a and b are to be obtained by maximizing the log likelihood function over a and b. The necessary conditions for an extremum are found to be

$$\frac{\partial \mathcal{L}}{\partial a} = -\frac{n}{a} + \frac{1}{b} \frac{1}{a^{1+1/b}} \sum_{i=1}^{n} |d_i-\mu|^{1/b} = 0 \tag{71}$$

$$\frac{\partial \mathcal{L}}{\partial b} = -\frac{n}{b\Gamma(b)} \left(b \frac{d}{db} \Gamma(b) + \Gamma(b)\right) + \frac{1}{b^2} \sum_{i=1}^{n} \left|\frac{d_i-\mu}{a}\right|^{1/b} \log\left|\frac{d_i-\mu}{a}\right| = 0 \tag{72}$$

Eq.(71) can be used to express the maximum likelihood estimate of the scale parameter as a function of the maximum likelihood estimate of the shape parameter b:

$$a = \left(\frac{1}{b}\right)^b \left(\frac{1}{n} \sum_{i=1}^{n} |d_i-\mu|^{1/b}\right)^b \tag{73}$$

Substituting eq.(73) into the necessary condition (72) and using

$$\psi(b) = \frac{d}{db} \log \Gamma(b) = \frac{1}{\Gamma(b)} \frac{d}{db} \Gamma(b) \tag{74}$$

yields:

$$f(b) \equiv \left(b\psi(b)+1\right) \left(\frac{1}{n} \sum_{i=1}^{n} |d_i-\mu|^{1/b}\right) - \frac{1}{n} \sum_{i=1}^{n} \left(|d_i-\mu|^{1/b} \log|d_i-\mu|\right) +$$

$$- \left(b \log b - b \log \left(\frac{1}{n} \sum_{i=1}^{n} |d_i-\mu|^{1/b}\right)\right) \left(\frac{1}{n} \sum_{i=1}^{n} |d_i-\mu|^{1/b}\right) = 0 \qquad (75)$$

The function $\psi(b)$ is shown in Figure 6.2 and a series expansion of it is given by eq.(C.22) of Appendix C. As can be seen from eqs.(73) and (75), the maximum likelihood parameter estimation problem for a Generalized Laplace density can not be solved fully analytically. The maximum likelihood estimate of the shape parameter has to be solved numerically from eq.(75). This may again be done with Newton's method. Linearizing eq.(75) yields

$$\Delta b = - \frac{f(b)}{f'(b)} \qquad (76)$$

where

$$f'(b) = \frac{1}{b^2} \left[\frac{1}{n} \sum_{i=1}^{n} |d_i-\mu|^{1/b} \log^2|d_i-\mu| + \right.$$

$$- \left(\frac{1}{n} \sum_{i=1}^{n} |d_i-\mu|^{1/b} \log|d_i-\mu|\right) \left(h(b) + 1 + b\right) +$$

$$\left. + b \left(\frac{1}{n} \sum_{i=1}^{n} |d_i-\mu|^{1/b}\right) \left(b^2 \psi'(b) + h(b) - b\right)\right] \qquad (77)$$

where

$$h(b) = b \left(\psi(b) - \log b + \log \frac{1}{n} \sum_{i=1}^{n} |d_i-\mu|^{1/b}\right) \qquad (78)$$

The iteration proces eq.(76) may be stopped whenever $|f(b)| < \varepsilon = 10^{-10}$.

A sufficient condition for a solution of eqs.(71) and (72), or equivalently eqs.(73) and (75), to define a maximum of the log likelihood function is that the matrix of second partial derivatives is negative definite. These partial derivatives are given by

$$\frac{\partial^2 \mathcal{L}}{\partial a^2} = \frac{n}{a^2} \left(1 - \frac{b+1}{b^2} \frac{1}{a^{1/b}} \frac{1}{n} \sum_{i=1}^{n} |d_i - \mu|^{1/b}\right) \tag{79}$$

$$\frac{\partial^2 \mathcal{L}}{\partial a \partial b} = - \frac{n}{b^3 a^{1+1/b}} \left[\left(\log a + b\right) \left(\frac{1}{n} \sum_{i=1}^{n} |d_i - \mu|^{1/b}\right) + \frac{1}{n} \sum_{i=1}^{n} |d_i - \mu|^{1/b} \log |d_i - \mu| \right] \tag{80}$$

$$\frac{\partial^2 \mathcal{L}}{\partial b^2} = - \frac{n}{b^3} \left[b^3 \psi'(b) - b + \frac{1}{b} \frac{1}{n} \sum_{i=1}^{n} \left|\frac{d_i - \mu}{a}\right|^{1/b} \log^2 \left|\frac{d_i - \mu}{a}\right| \right]$$

$$+ 2 \frac{1}{n} \sum_{i=1}^{n} \left|\frac{d_i - \mu}{a}\right|^{1/b} \log \left|\frac{d_i - \mu}{a}\right| \tag{81}$$

The matrix of partial derivatives is negative definite when its determinant is positive and $\partial^2 \mathcal{L}/\partial a^2$ is negative. Denoting any solution of eqs.(71) and (72) by $\hat{a}$ and $\hat{b}$ and using eq.(73) immediately gives

$$\frac{\partial^2 \mathcal{L}}{\partial a^2} (\hat{a}, \hat{b}) = - \frac{n}{\hat{a}^2 \hat{b}} \tag{82}$$

and, after some analysis

$$\det(\hat{a}, \hat{b}) = \frac{n^2}{\hat{a}^2 \hat{b}^4} \left\{ \hat{b}^3 \psi'(\hat{b}) - \hat{b} - \hat{b}^2 - \hat{b}^2 \psi(\hat{b}) - \hat{b}\psi(\hat{b}) - 1 + \right.$$

$$+ \left. \frac{1}{\hat{b}} \frac{1}{n} \sum_{i=1}^{n} \left|\frac{d_i - \mu}{\hat{a}}\right|^{1/\hat{b}} \log^2 \left|\frac{d_i - \mu}{\hat{a}}\right| \right\} \tag{83}$$

Eq.(82) is clearly negative for any positive estimate $\hat{b}$, but eq.(83) is not immediately seen to be positive for $\hat{a}$ and $\hat{b}$ satisfying eqs.(71) and (72). Thus, it remains to verify by computation whether eq.(83) is positive after the likelihood equations have been solved.

The standardized log likelihood value can be derived from eq.(36) by taking $\alpha=0$ in eq.(36). The resulting expression is

$$\mathcal{L}_2 - \mathcal{L} = \sum_{i=1}^{n} \log \frac{1}{e^{-\left|\frac{d_i - \mu}{a}\right|^{1/b}}} = \left(\frac{1}{a}\right)^{1/b} \sum_{i=1}^{n} |d_i - \mu|^{1/b} \tag{84}$$

Using eq.(73) this may be further simplified as

$$\mathcal{L}_2 - \mathcal{L} = n \hat{b} \tag{85}$$

where $\hat{b}$ denotes the maximum likelihood estimate of the shape parameter b of the Generalized Laplace densities.

Consider now the likelihood function for data grouped into m classes with the underlying distribution of the data points being Generalized Laplace. It is given by

$$L = \frac{n!}{x_1! x_2! \ldots x_m!} \cdot \prod_{i=1}^{m} \left(2w \frac{1}{2ab\Gamma(b)} e^{-\left(\frac{(i-\frac{1}{2})w}{a}\right)^{1/b}} \right)^{x_i} \tag{86}$$

whereas the log likelihood function is

$$\mathcal{L} = A + n \log 2w - n \log 2ab\Gamma(b) - \sum_{i=1}^{m} x_i \left(\frac{(i-\frac{1}{2})w}{a}\right)^{1/b} \tag{87}$$

In these equations, x_i denotes the number of observations in class i, $i=1,2,\ldots,m$ and A is given by eq.(52). Notice the similarity between the two log likelihood functions given by eqs.(70) and (87). As a result, the necessary conditions for an extremum of this log likelihood function are also very similar to eqs.(71) and (72), namely

$$\frac{\partial \mathcal{L}}{\partial a} = -\frac{n}{a} + \frac{1}{b} \frac{1}{a^{1+1/b}} \sum_{i=1}^{m} x_i \left((i-\frac{1}{2})w\right)^{1/b} = 0 \tag{88}$$

$$\frac{\partial \mathcal{L}}{\partial b} = -\frac{n}{b\Gamma(b)} \left(b\frac{d}{db}\Gamma(b) + \Gamma(b)\right) + \frac{1}{b^2} \sum_{i=1}^{m} x_i \left(\frac{(i-\frac{1}{2})w}{a}\right)^{1/b} \log\left(\frac{(i-\frac{1}{2})w}{a}\right) = 0 \tag{89}$$

Clearly, the summation now is over the number of intervals, and the individual distances $|d_i - \mu|$ have been replaced by the distances of the centres of the classes to the mean value. The maximum likelihood estimate of the scale parameter a can be solved from eq.(88) as a function of the estimate of the shape parameter b:

$$a = \left(\frac{1}{b}\right)^b \left\{ \frac{1}{n} \sum_{i=1}^{m} x_i \left((i-\frac{1}{2})w\right)^{1/b} \right\}^b \tag{90}$$

Compare this estimate with eq.(73) for the non-grouped data case. Substitution of eq.(90) into eq.(89) gives for the maximum likelihood estimate of the shape parameter b the following non-linear equation:

$$
g(b) \equiv \left(b\psi(b)+1\right)\left\{\frac{1}{n}\sum_{i=1}^{m}x_i\left((i-\tfrac{1}{2})w\right)^{1/b}\right\} - \frac{1}{n}\sum_{i=1}^{m}\left\{x_i\left((i-\tfrac{1}{2})w\right)^{1/b}\log((i-\tfrac{1}{2})w)\right\} +
$$

$$
- \left\{b \log b - b \log \frac{1}{n}\sum_{i=1}^{m}x_i\left((i-\tfrac{1}{2})w\right)^{1/b}\right\}\left\{\frac{1}{n}\sum_{i=1}^{m}x_i\left((i-\tfrac{1}{2})w\right)^{1/b}\right\} = 0 \quad (91)
$$

Compare eqs.(91) and (75) for the grouped data and the non-grouped data case respectively. Like eq.(75), eq.(91) needs to be solved numerically. This may again be done by Newton's method. The correction Δb after linearizing the function $g(b)$ around b is given by

$$
\Delta b = -\frac{g(b)}{g'(b)} \quad (92)
$$

with $g(b)$ defined in eq.(91) and $g'(b)$ given by

$$
g'(b) = \frac{1}{b^2}\left[\frac{1}{n}\sum_{i=1}^{m}x_i\left((i-\tfrac{1}{2})w\right)^{1/b}\log^2((i-\tfrac{1}{2})w) + \right.
$$

$$
- \left\{ \frac{1}{n}\sum_{i=1}^{m}x_i\left((i-\tfrac{1}{2})w\right)^{1/b}\log((i-\tfrac{1}{2})w)\right\} \left(k(b) + 1 + b\right) +
$$

$$
\left. + b\left\{ \frac{1}{n}\sum_{i=1}^{m}x_i\left((i-\tfrac{1}{2})w\right)^{1/b}\right\} \left\{b^2\psi'(b) + k(b) - b\right\} \right] \quad (93)
$$

where
$$
k(b) = b\left\{\psi(b) - \log b + \log\frac{1}{n}\sum_{i=1}^{m}x_i\left((i-\tfrac{1}{2})w\right)^{1/b}\right\} \quad (94)
$$

Eqs.(93) and (94) are straightforward generalizations of eqs.(77) and (78) for the non-grouped data case. In the same way, the eqs.(79) to (83) generalize. The sufficient conditions for a maximum become

$$
\frac{\partial^2 \ell}{\partial a^2}(a,b) = -\frac{n}{a^2 b} < 0 \quad (95)
$$

and

$$\det(a,b) = \frac{n^2}{a^2b^4} \left\{ b^3 \psi'(b) - b - b^2 - b^2\psi(b) - b\psi(b) + \right.$$

$$\left. - 1 + \frac{1}{b} \frac{1}{n} \sum_{i=1}^{m} x_i \left(\frac{(i-\frac{1}{2})w}{a}\right)^{1/b} \log^2\left(\frac{(i-\frac{1}{2})w}{a}\right) \right\} > 0 \qquad (96)$$

where a, b now denote any solution of the likelihood equations (88) and (89) for grouped data (rather than the true parameter values).

The standardized log likelihood value can be derived from eq.(68) by substituting the appropriate expression for p_i:

$$\mathcal{L}_2 - \mathcal{L} = \sum_{i=1}^{m} x_i \log \frac{x_i}{n2w \frac{1}{2ab\Gamma(b)} e^{-\left(\frac{(i-\frac{1}{2})w}{a}\right)^{1/b}}} \qquad (97)$$

Evaluating this gives, with the aid of eq.(90):

$$\mathcal{L}_2 - \mathcal{L} = nb + \sum_{i=1}^{m} x_i \log x_i - n \log 2w + n \log\left(\frac{2ab\Gamma(b)}{n}\right) \qquad (98)$$

where a and b denote the maximum likelikhood estimates. Eq.(98) may also be expressed as a function of b alone by means of eq.(90), but this does not provide any further essential insight. Notice the common term nb in the two expressions (98) and (85) for the standardized log likelihood value for grouped and non-grouped data respectively.

4.4.2 Gaussian densities

The Gaussian density is a member of the family of Generalized Laplace probability densities. Its shape parameter is equal to 0.5. Hence, it may be expressed as

$$f_G(d) = \frac{1}{a\Gamma(0.5)} e^{-\left(\frac{d-\mu}{a}\right)^2} \qquad (99)$$

The scale parameter a is related to the standard deviation by the relation $a = \sigma\sqrt{2}$. Further holds that $\Gamma(0.5) = \sqrt{\pi}$. Eq.(99) may thus also be written in the more familiar form

$$f_G(d) = \frac{1}{\sigma\sqrt{2\pi}} \, e^{-\left(\frac{d-\mu}{\sigma\sqrt{2}}\right)^2} \qquad (100)$$

Consider the non-grouped data case first. The log likelihood function for the Gaussian probability density follows immediately from eq.(70) if b = 0.5 is substituted in it. Thus,

$$\mathcal{L} = \mathcal{L}_{(a)} = -n \log a \, \Gamma(0.5) - \sum_{i=1}^{n} \left(\frac{d_i-\mu}{a}\right)^2 \qquad (101)$$

where d_i, i=1,2,...,n denote the individual data points and n is the sample size. Taking the derivative with respect to the scale parameter a and equating that with zero yields, similar to eq.(71):

$$-\frac{n}{a} + 2 \, \frac{1}{a^3} \sum_{i=1}^{n} (d_i - \mu)^2 = 0 \qquad (102)$$

This gives the following explicit expression for the maximum likelihood estimate of the scale parameter, similar to eq.(73):

$$\hat{a} = \sqrt{2} \left\{ \frac{1}{n} \sum_{i=1}^{n} (d_i-\mu)^2 \right\}^{\frac{1}{2}} \qquad (103)$$

Because the scale parameter is equal to $\sigma\sqrt{2}$, eq.(103) basically expresses that the maximum likelihood estimate of the standard deviation of a Gaussian variable with known mean value μ is equal to the sample standard deviation with known mean value μ. The second derivative of the log likelihood function becomes (see eq.(82)):

$$\frac{d^2\mathcal{L}}{da^2}(\hat{a}) = -\frac{2n}{\hat{a}^2} = -\frac{n}{\frac{1}{n} \sum_{i=1}^{n} (d_i-\mu)^2} \qquad (104)$$

and is negative. The standard deviation $\sigma(a)$ of the maximum likelihood estimate of the scale parameter thus becomes

$$\sigma(\hat{a}) = \frac{\hat{a}}{\sqrt{2n}} \qquad (105)$$

For grouped data, distributed over m classes, the maximum likelihood estimate of the scale parameter becomes from eq.(90) for b = 0.5:

$$\hat{\hat{a}} = \sqrt{2} \left\{ \frac{1}{n} \sum_{i=1}^{m} x_i \left((i-\tfrac{1}{2})w \right)^2 \right\}^{\frac{1}{2}} \tag{106}$$

Clearly, when the number of classes m increases over the same domain, eq.(106) approaches eq.(103) for the non-grouped data case. The standardized log likelihood value can be computed from eq.(98) for b = 0.5.

4.4.3 Double Exponential densities

The Double Exponential density also belongs to the family of Generalized Laplace densities, viz. with shape parameter b = 1.0. Thus, it may be expressed as

$$f_{DE}(d) = \frac{1}{2a} e^{-\left| \frac{d-\mu}{a} \right|} \tag{107}$$

The scale parameter a is related to the standard deviation by $\sigma = a\sqrt{2}$.

The log likelihood function for the Double Exponential density follows again from eq.(70) for non-grouped data. Taking b=1.0 yields

$$\mathcal{L} = - n \log 2a - \sum_{i=1}^{n} \left| \frac{d_i - \mu}{a} \right| \tag{108}$$

Taking b=1.0 in eq.(73) yields the maximum likelihood estimate of the scale parameter of a Double Exponentially distributed variable:

$$\hat{a} = \frac{1}{n} \sum_{i=1}^{n} |d_i - \mu| \tag{109}$$

Thus, the maximum likelihood estimate of the scale parameter is exactly the average distance between the data points d_i and the mean value μ. The second derivative of the log likelihood function with respect to the scale parameter is

$$\frac{d^2\ell}{da^2}(a) = \frac{n}{a^2} - \frac{2}{a^3} \sum_{i=1}^{n} |d_i - \mu| \tag{110}$$

For a defined by eq.(99), this can be expressed as

$$\frac{d^2\ell}{da^2} = -\frac{n}{\hat{a}^2} \tag{111}$$

Hence, the estimate given by eq.(109) does indeed define a maximum of the log likelihood function. The asymptotic standard deviation of the maximum likelihood estimator of the scale parameter is found to be

$$\sigma(\hat{a}) = \frac{\hat{a}}{\sqrt{n}} = \frac{1}{\sqrt{n}} \left(\frac{1}{n} \sum_{i=1}^{n} |d_i - \mu| \right) \tag{112}$$

For grouped data, distributed over m classes again, eq.(90) for b=1.0 gives as an estimate

$$\hat{\hat{a}} = \frac{1}{n} \sum_{i=1}^{m} x_i \left((i-\tfrac{1}{2})w \right) \tag{113}$$

As for the non-grouped data case, this estimate is easily shown to define a maximum of the pertinent log likelihood function. The asymptotic standard deviation of the maximum likelihood estimator for grouped data is also given by $\sigma(a) = a/\sqrt{n}$. Using the estimate eq.(113), this gives

$$\sigma(\hat{\hat{a}}) = \frac{\hat{\hat{a}}}{\sqrt{n}} = \frac{1}{\sqrt{n}} \left(\frac{1}{n} \sum_{i=1}^{m} x_i((i-\tfrac{1}{2})w) \right) \tag{114}$$

4.4.4 Double Double Exponential densities

The Double Double Exponential (DDE) density is a mixture of two Generalized Laplace densities, where the shape parameters of these two densities are both given a value of 1.0 (one). The DDE density has been used extensively in modelling aircraft navigation and height keeping error data. Mathematically, it is defined by the relation

$$f_{DDE}(d) = (1-\alpha) \frac{1}{2a_1} e^{-\left|\frac{d-\mu}{a_1}\right|} + \frac{\alpha}{2a_2} e^{-\left|\frac{d-\mu}{a_2}\right|} \tag{115}$$

The three unknown parameters, in addition to the mean value μ, are the weighting factor α and the two scale parameters a_1 and a_2. These parameters can be estimated as for the general case described in the sections 4.2 and 4.3 for non-grouped data and grouped data respectively, by applying Newton's method to the appropriate system of likelihood equations. Two particular problems will be analyzed below. The first problem concerns the question whether the Double Exponential (DE) density defines a local maximum of the likelihood function for the DDE family of densities. It turns out that this is the case indeed. The second problem concerns the question what will happen to the iterants of the Newton process when the initial guesses a_{1_0} and a_{2_0}, say, for the two scale parameters are the same. This choice of initial values means that the initial density is DE. It turns out that the density remains DE throughout the iteration process, but that the iteration process may either converge or diverge, depending on the initial value $a_{1_0} = a_{2_0}$. Only the case of grouped data will be analyzed explicitly. It will be seen that the case of non-grouped data may be analyzed in exactly the same manner.

The log likelihood function for the family of DDE probability densities and grouped data (with class interval w) is

$$\mathcal{L} = A + n\log 2w + \sum_{i=1}^{m} x_i \log\left\{(1-\alpha) \frac{1}{2a_1} e^{-(i-\frac{1}{2})\frac{w}{a_1}} + \alpha \frac{1}{2a_2} e^{-(i-\frac{1}{2})\frac{w}{a_2}}\right\} \quad (116)$$

with A given by eq.(53), m being the number of classes and x_i, i=1,2,...,m the number of observations in class i. The necessary conditions for an extreme value of the log likelihood function are:

$$\frac{\partial}{\partial \alpha} \mathcal{L} = 2w \sum_{i=1}^{m} \frac{x_i}{p_i} \left(- \frac{1}{2a_1} e^{-(i-\frac{1}{2})\frac{w}{a_1}} + \frac{1}{2a_2} e^{-(i-\frac{1}{2})\frac{w}{a_2}}\right) = 0 \quad (117)$$

$$\frac{\partial \mathcal{L}}{\partial a_1} = 2w (1-\alpha) \sum_{i=1}^{m} \frac{x_i}{p_i} \left\{ \frac{1}{2a_1} e^{-(i-\frac{1}{2})\frac{w}{a_1}} \left(\frac{(i-\frac{1}{2})w}{a_1^2} - \frac{1}{a_1}\right)\right\} = 0 \quad (118)$$

$$\frac{\partial \ell}{\partial a_2} = 2w \; \alpha \; \sum_{i=1}^{m} \frac{x_i}{p_i} \left\{ \frac{1}{2a_2} e^{-(i-\frac{1}{2})\frac{w}{a_2}} \left(\frac{(i-\frac{1}{2})w}{a_2^2} - \frac{1}{a_2} \right) \right\} = 0 \tag{119}$$

with p_i, $i=1,2,\ldots,m$ defined by

$$p_i = 2w \left\{ (1-\alpha) \; \frac{1}{2a_1} e^{-(i-\frac{1}{2})\frac{w}{a_1}} + \alpha \; \frac{1}{2a_2} e^{-(i-\frac{1}{2})\frac{w}{a_2}} \right\} = 0 \tag{120}$$

The factor 2w after the first equal sign in eqs.(117)-(119) will be dropped in the discussion to follow.

The first question with regard to the likelihood equations (117)-(119) is whether a Double Exponential density satisfies this system. As a DE density may be expressed as a DDE density with arbitrary weighting factor but identical scale factors, the question is whether solutions of the form α,a,a exist of eqs.(117)-(119). It follows that such solutions satisfy eq.(117) and that they satisfy eqs.(118) and (119) as well if

$$(1-\alpha) \sum_{i=1}^{m} x_i \left(\frac{(i-\frac{1}{2})w}{a^2} - \frac{1}{a} \right) = 0 \tag{121}$$

and

$$\alpha \sum_{i=1}^{m} x_i \left(\frac{(i-\frac{1}{2})w}{a^2} - \frac{1}{a} \right) = 0 \tag{122}$$

Eq.(121) gives either a solution for α:

$$1 - \alpha = 0 \tag{123}$$

or a solution for a:

$$\sum_{i=1}^{m} x_i \left(\frac{(i-\frac{1}{2})w}{a^2} - \frac{1}{a} \right) = 0 \tag{124}$$

The first solution, i.e. eq.(123) gives $\alpha=1.0$, which substituted into eq.(122) gives again eq.(124) or

$$\widehat{\widehat{a}} = \frac{1}{n} \sum_{i=1}^{m} x_i (i-\frac{1}{2})w \tag{125}$$

Eq.(125) is exactly equal to eq.(113) of Section 4.4.3, defining the maximum likelihood estimate of the scale parameter of a Double Exponential density. The other solution of eq.(121), i.e. that solution for a defined by eq.(124) yields exactly $\hat{\hat{a}}$ defined in eq.(125). Substituting that in eq.(122) gives $\alpha x 0 = 0$, which is satisfied for any value of α. Indeed, when both the densities making up the mixture have the same scale parameter, the value of the weighting factor is undetermined and of no relevance. The same conclusion is arrived at when eq.(122) rather than eq.(121) is taken as the starting point of the analysis of the system of eqs.(121) and (122).

Thus, it has been shown that the DDE density with parameters $\alpha, \hat{\hat{a}}, \hat{\hat{a}}$, which is effectively a DE density with parameter $\hat{\hat{a}}$ satisfies the likelihood equations for the family of DDE densities. Considered as a DE density, it defines a maximum of the log likelihood function within the family of DE densities. However, it remains to see whether, for an arbitrary value of α, it also defines a (local) maximum of the log likelihood function within the larger family of DDE densities. Let $H(\alpha, a_1, a_2)$ denote the Hessian matrix of the second partial derivatives of the log likelihood function for the family of DDE densities. It follows that $H(\alpha, a, a)$ is given by

$$
H(\alpha,a,a) = \begin{pmatrix} 0 & 0 & 0 \\ \\ 0 & (1-\alpha)^2\frac{1}{a^2}\sum_{i=1}^{m}x_i\left\{\left(\frac{y_i}{a}\right)^2 - 4\frac{y_i}{a} +2 -(1-\alpha)\left(-1+\frac{y_i}{a}\right)^2\right\} & -\alpha(1-\alpha)^2\sum_{i=1}^{m}x_i\left(-1+\frac{y_i}{a}\right)^2 \\ \\ 0 & -\alpha^2(1-\alpha)\sum_{i=1}^{m}x_i\left(-1+\frac{y_i}{a}\right)^2 & \alpha^2\frac{1}{a^2}\sum_{i=1}^{m}x_i\left\{\left(\frac{y_i}{a}\right)^2 - 4\frac{y_i}{a} +2 -\alpha\left(-1+\frac{y_i}{a}\right)^2\right\} \end{pmatrix}
$$

$$(126)$$

where $y_i = (i-\frac{1}{2})w$. This Hessian is singular, indicating that the weighting factor α is undetermined. This is correct, because the weighting factor should not play a part in a mixture of two identical probability densities. A Taylor series expansion of the log likelihood function defined by eq.(116) around the point α, a, a gives

$$
\mathcal{L}(\alpha,a_1,a_2) = \mathcal{L}(\alpha,a,a) + (\Delta a_1 \; \Delta a_2) \, H_1(\alpha,a,a) \, (\Delta a_1 \; \Delta a_2)^T \tag{127}
$$

where $H_1(\alpha,a,a)$ denotes the non-trivial submatrix of the Hessian (α,a,a), i.e.

$$H_1(\alpha,a,a) = \begin{pmatrix} (1-\alpha)^2\frac{1}{a^2}\sum_{i=1}^{m} x_i\left\{\left(\frac{y_i}{a}\right)^2 - 4\frac{y_i}{a} +2-(1-\alpha)\left(-1+\frac{y_i}{a}\right)^2\right\} & -\alpha(1-\alpha)^2 \sum_{i=1}^{m} x_i\left(-1+\frac{y_i}{a}\right)^2 \\ \\ -\alpha^2(1-\alpha) \sum_{i=1}^{m} x_i\left(-1+\frac{y_i}{a}\right)^2 & \alpha^2\frac{1}{a^2}\sum_{i=1}^{m} x_i\left\{\left(\frac{y_i}{a}\right)^2 - 4\frac{y_i}{a} +2-\alpha\left(-1+\frac{y_i}{a}\right)^2\right\} \end{pmatrix}$$

$$(128)$$

Notice that the zeroth order term in the right-hand side of eq.(127) is, in fact, independent from α as follows from eq.(116). For a local maximum in the point $\alpha,\hat{\hat{a}},\hat{\hat{a}}$, it is sufficient that the matrix $H_1(\alpha,\hat{\hat{a}},\hat{\hat{a}})$ is negative definite. This is the case when its upper left element is negative and its determinant is positive. Hence

$$\frac{1}{n}(1-\alpha)^2 \frac{1}{\hat{\hat{a}}^4}\left\{-(1+\alpha)\left(\sum_{i=1}^{m} x_iy_i\right)^2 + \alpha \left(\sum_{i=1}^{m} x_iy_i^2\right)\left(\sum_{i=1}^{m} x_i\right)\right\} < 0 \qquad (129)$$

and

$$\det(H_1(\alpha,\hat{\hat{a}},\hat{\hat{a}})) = -\alpha^2(1-\alpha)^2\frac{2}{\hat{\hat{a}}^4}\left\{-2\left(\sum_{i=1}^{m} x_iy_i\right)^2 + \left(\sum_{i=1}^{m} x_i\right)\left(\sum_{i=1}^{m} x_iy_i^2\right)\right\} > 0 \quad (130)$$

Notice that the sign of $\det(H_1(\alpha,\hat{\hat{a}},\hat{\hat{a}}))$ does not depend on α. The validity of the relations (129) and (130) can easily be verified numerically for the given numbers of data points x_i, $i=1,2,\ldots,m$.

All the considerations above apply to the log likelihood function for grouped data, eq.(116). This, and all the following equations can formally be converted into their equivalents for non-grouped data by means of the following set of substitutions:

$$\left.\begin{aligned} m &\rightarrow n \\ A &\rightarrow 0 \\ (i-\tfrac{1}{2})w &\rightarrow |d_i-\mu| \\ 2w &\rightarrow 1 \end{aligned}\right\} \qquad (131)$$

Hence, the results obtained for grouped data, apply to the case of non-grouped data as well.

Consider now the numerical solution of the likelihood equations (117)-(119) by means of Newton's method. The linearized versions of these equations can be derived from the general case discussed in Section 4.3, i.e. eqs. (59), (60) and (61) by ignoring the terms with Δb_1 and Δb_2. This gives

$$\sum_{i=1}^{m} \frac{x_i}{p_i} \left[\frac{\partial p_i}{\partial \alpha} + \left(\frac{\partial^2 p_i}{\partial \alpha^2} - \frac{1}{p_i} \frac{\partial p_i}{\partial \alpha} \frac{\partial p_i}{\partial \alpha} \right) \Delta \alpha + \right.$$

$$\left. + \left(\frac{\partial^2 p_i}{\partial \alpha \partial a_1} - \frac{1}{p_i} \frac{\partial p_i}{\partial a_1} \frac{\partial p_i}{\partial \alpha} \right) \Delta a_1 + \left(\frac{\partial^2 p_i}{\partial \alpha \partial a_2} - \frac{1}{p_i} \frac{\partial p_i}{\partial a_2} \frac{\partial p_i}{\partial \alpha} \right) \Delta a_2 + \frac{\partial p_i}{\partial \alpha} \right] = 0 \tag{132}$$

$$\sum_{i=1}^{m} \frac{x_i}{p_i} \left[\frac{\partial p_i}{\partial a_1} + \left(\frac{\partial^2 p_i}{\partial \alpha \partial a_1} - \frac{1}{p_i} \frac{\partial p_i}{\partial \alpha} \frac{\partial p_i}{\partial a_1} \right) \Delta \alpha + \right.$$

$$\left. + \left(\frac{\partial^2 p_i}{\partial a_1^2} - \frac{1}{p_i} \frac{\partial p_i}{\partial a_1} \frac{\partial p_i}{\partial a_1} \right) \Delta a_1 + \left(\frac{\partial^2 p_i}{\partial a_1 \partial a_2} - \frac{1}{p_i} \frac{\partial p_i}{\partial a_2} \frac{\partial p_i}{\partial a_1} \right) \Delta a_2 + \frac{\partial p_i}{\partial a_1} \right] = 0 \tag{133}$$

$$\sum_{i=1}^{m} \frac{x_i}{p_i} \left[\frac{\partial p_i}{\partial a_2} + \left(\frac{\partial^2 p_i}{\partial \alpha \partial a_2} - \frac{1}{p_i} \frac{\partial p_i}{\partial \alpha} \frac{\partial p_i}{\partial a_2} \right) \Delta \alpha + \right.$$

$$\left. + \left(\frac{\partial^2 p_i}{\partial a_1 \partial a_2} - \frac{1}{p_i} \frac{\partial p_i}{\partial a_1} \frac{\partial p_i}{\partial a_2} \right) \Delta a_1 + \left(\frac{\partial^2 p_i}{\partial a_2^2} - \frac{1}{p_i} \frac{\partial p_i}{\partial a_2} \frac{\partial p_i}{\partial a_2} \right) \Delta a_2 + \frac{\partial p_i}{\partial a_2} \right] = 0 \tag{134}$$

All quantities other than $\Delta \alpha$, Δa_1 and Δa_2 in eqs.(132)-(134) depend explicitly on the values $\alpha_0, a_{1_0}, a_{2_0}$, say, about which the original equations have been linearized. After determining a solution $\Delta \alpha$, Δa_1, and Δa_2 of eqs.(132)-(134), new values of α, a_1, and a_2 are computed by means of

$$\left. \begin{aligned} \alpha_1 &= \alpha_0 + \Delta \alpha \\ a_{1_1} &= a_{1_0} + \Delta a_1 \\ a_{2_1} &= a_{2_0} + \Delta a_2 \end{aligned} \right\} \tag{135}$$

This process is repeated until either a desired accuracy criterion is achieved or the process breaks down. The central question is, therefore, for which triples of initial guesses α_0, a_{1_0}, a_{2_0} the iteration process converges. Although a full analysis is presently not possible, some useful insight can be obtained from the special case that the initial guesses of the two scale parameters are identical. Many partial derivatives occurring in the linearized equations then reduce to relatively simple expressions. Thus, consider the system of eqs.(132)-(134) evaluated for an initial guess α,a,a with $a>0$ and $0\leq\alpha\leq1$. Using again $y_i=(i-\frac{1}{2})w$ the system reduces to

$$
\left(\; X \;\right) \left(\begin{array}{c} \Delta\alpha \\ \Delta a_1 \\ \Delta a_2 \end{array}\right) = \left(\; Y \;\right) \tag{136}
$$

where the matrix X and the vector Y are given by

$$
X = \begin{pmatrix} 0 & -\frac{1}{a}\sum_{i=1}^{m} x_i\left(\frac{y_i}{a}-1\right) & \frac{1}{a}\sum_{i=1}^{m} x_i\left(\frac{y_i}{a}-1\right) \\[2ex] -(1-\alpha)\frac{1}{a}\sum_{i=1}^{m} x_i\left(\frac{y_i}{a}-1\right) & (1-\alpha)^2\frac{1}{a^2}\sum_{i=1}^{m} x_i\left\{\left(\frac{y_i}{a}\right)^2-4\frac{y_i}{a}+2-(1-\alpha)\left(\frac{y_i}{a}-1\right)^2\right\} & -\alpha(1-\alpha)^2\frac{1}{a^2}\sum_{i=1}^{m} x_i\left(\frac{y_i}{a}-1\right)^2 \\[2ex] \alpha\frac{1}{a}\sum_{i=1}^{m} x_i\left(\frac{y_i}{a}-1\right) & -\alpha^2(1-\alpha)\frac{1}{a^2}\sum_{i=1}^{m} x_i\left(\frac{y_i}{a}-1\right)^2 & \alpha^2\frac{1}{a^2}\sum_{i=1}^{m} x_i\left\{\left(\frac{y_i}{a}\right)^2-4\frac{y_i}{a}+2-\alpha\left(\frac{y_i}{a}-1\right)^2\right\} \end{pmatrix}
$$

$$\tag{137}$$

$$
Y = \begin{pmatrix} 0 \\[2ex] -(1-\alpha)^2\frac{1}{a}\sum_{i=1}^{m} x_i\left(\frac{y_i}{a}-1\right) \\[2ex] -\alpha^2\frac{1}{a}\sum_{i=1}^{m} x_i\left(\frac{y_i}{a}-1\right) \end{pmatrix} \tag{138}
$$

The solution of the system (136) to (138) is dependent on the values of α and a occurring in the coefficient matrix X and the right-hand side vector Y. Table 4 summarizes this dependence. For all the cases other than case 1, the matrix X is singular. Before considering the regular case, the

singular cases are considered. For each of the cases ②, ⑤ and ⑧, the initial guesses of a_1 and a_2 are the same and equal to the maximum likelihood estimate of the scale parameter of a (single) Double Exponential probability density defined by eq.(125). As a DE density with that particular parameter value already satisfies the (non-linear) likelihood equations, the Newton algorithm does not change the scale parameters of the densities involved. For example, when the initial value of the weighting factor is $\alpha=0$, only the core density plays a part. When the initial guesses of both the scale parameters are equal to twice the maximum likelihood estimate of a DE density, as for the cases ③, ⑥ and ⑨, the algorithm breaks down. Notice that case ⑥, starting, from a weighting factor of $\alpha=0$, transfers into case ⑨, with a weighting factor of 1, and similarly the other way around. The cases ④ and ⑦ yield a completely undetermined solution.

Consider now case ①. It follows from Table 4 that, when the initial guesses of the scale parameters of the core and tail density making up the Double Double Exponential mixture density are the same, these will remain the same for each iteration in the Newton process. Moreover, the weighting factor α does not change during the iteration process. This is correct because the weighting does not have any real significance when the two scale parameters are equal. It remains to analyze what the correction Δ of the scale parameter is. It holds that for $0<\alpha<1$ and $a_j\neq\hat{\hat{a}}$ as well as $a_j\neq2\hat{\hat{a}}$:

$$\Delta a_j = a_j\left(\frac{1-a_j/\hat{\hat{a}}}{2-a_j/\hat{\hat{a}}}\right) \qquad , j=0,1,2,\ldots \qquad (139)$$

The integer j in eq.(139) counts the iterations of the Newton process and the quantities Δa_j and a_j are abbreviations for $\Delta a_{1_j}=\Delta a_{2_j}$ and $a_{1_j}=a_{2_j}$ respectively. The convergence of the iteration process (139) follows from the following proposition.

Proposition

Consider the iteration process given by

$$\Delta x_i = x_i\left(\frac{1-x_i}{2-x_i}\right) \qquad , i=0,1,2,\ldots \qquad (140)$$

or, equivalently

$$x_{i+1} = x_i \left(\frac{3-2x_i}{2-x_i} \right) \qquad , \ i=0,1,2,\ldots \qquad (141)$$

under the condition $x_i > 0$ for every i.

This iteration process converges to a finite limit iff the initial value satisfies the relation $0 < x_0 < 3/2$. Under this condition the limit value is $x_\infty = 1$.

Proof

The proof is elementary and is illustrated in Figure 8. Define $C_i = (3-2x_i)/(2-x_i)$. Figure 8A shows that for $x_i > 2$ and $x_i < 3/2$, the factor C_i is positive. C_i must be positive in order to yield a sequence of positive numbers x_i, $i=0,1,2,\ldots$. Figure 8B shows that for $1 < x_i < 2$, the factor C_i is less than one, possibly negative. It follows that for every x_i with $2 < x_i$ holds $x_i < x_{i+1}$, leading to divergence of the iteration process.

For $3/2 < x_i < 2$ holds that $x_{i+k} < 0$, for $k \geq 1$, leading to divergence as well. It remains to consider $0 < x_i < 3/2$. Figure 8C shows that for every $x_i < 2$ holds $x_{i+1} < 1$ (possibly negative). Hence, for $1 < x_i < 3/2$ holds $x_{i+1} < 1$. But for every x_i with $0 < x_i < 1$ holds $x_i < x_{i+1} < 1$, leading to monotone convergence to $x_\infty = 1.0$.

See Figure 8D for a summary of the convergence properties.

Application of the Proposition to the iteration process eq.(139) leads to the conclusion that the process will converge to $a_{1_\infty} = a_{2_\infty} = \hat{\hat{a}}$ when the initial guesses a_{1_0} and a_{2_0} are equal and satisfy the condition $0 < a_{1_0} = a_{2_0} < 3/2\hat{\hat{a}}$. A similar convergence will occur when the initial values a_{1_0} and a_{2_0} are unequal, but not too far away from each other. An important practical consequence is that a Double Double Exponential mixture density with two different scale parameters rather than a Double Exponential density will be produced by the Newton iteration process of eqs.(132) to (134) only, if the initial guesses of the scale parameters of the core and tail density are sufficiently different.

5   SIMULATION OF RANDOM VARIABLES HAVING A DOUBLE GENERALIZED LAPLACE
     PROBABILITY DISTRIBUTION

Computer simulation of random variables with a Double Generalized
Laplace probability distribution is useful for two reasons. Firstly, it
offers the possibility to validate the parameter estimation algorithm
under controlled conditions. When the parameter values used in the
simulation are recovered with a sufficient precision, the algorithm is
likely to have been implemented correctly. Secondly, it offers the
possibility to investigate all kinds of properties of the parameter
estimation algorithm, for example: accuracy of the estimated parameters as
a function of the sample size, accuracy as a function of the number of
estimated parameters, convergence of the Newton iteration process as a
function of the initial estimate of the parameters, etc.

The usual way to generate a sample of size n of a random variable $\underline{d}$
with a specified probability distribution is to generate n random numbers
u_i, i=1,2,...,n, independently and homogeneously distributed on the
interval [0,1], and to use the cumulative distribution function F(d),

$$F(d) = \int_{-\infty}^{d} f(x)dx \qquad (142)$$

by putting

$$F(d_i) = u_i \qquad , i=1,2,...,n \qquad (143)$$

The realizations d_i are obtained by inverting the cumulative distribution
function. This inversion can not always be done analytically. A possible
solution, which will in fact be used in this report, is to solve eq.(143)
numerically for d_i. Before considering the numerical approach, the
cumulative distribution function of a random variable with a Double
Generalized Laplace probability distribution is described.

Let the variable $\underline{d}$ have the Double Generalized Laplace probability
density $f_{DGL}(d)$ defined by eq.(11) of Section 3.2. The cumulative
distribution function $F_{DGL}(d)$ can be expressed in two ways. The first is

$$F_{DGL}(d) = \begin{cases} (1-\alpha)\dfrac{1}{2\Gamma(b_1)} \ \Gamma\!\left(b_1, \left(\dfrac{\mu-d}{a_1}\right)^{1/b_1}\right) + \alpha\dfrac{1}{2\Gamma(b_2)} \ \Gamma\!\left(b_2, \left(\dfrac{\mu-d}{a_2}\right)^{1/b_2}\right) \ , \ d \leq \mu \\[4mm] 1-(1-\alpha)\dfrac{1}{2\Gamma(b_1)} \ \Gamma\!\left(b_1, \left(\dfrac{d-\mu}{a_1}\right)^{1/b_1}\right) - \alpha\dfrac{1}{2\Gamma(b_2)} \ \Gamma\!\left(b_2, \left(\dfrac{d-\mu}{a_2}\right)^{1/b_2}\right), \ d \geq \mu \end{cases}$$

$$(144)$$

The symbol $\Gamma(b)$ denotes, as before, the Gamma function for the argument value b and the symbol $\Gamma(b,x)$ denotes the incomplete Gamma function defined by

$$\Gamma(b,x) = \int_x^\infty y^{b-1} \ e^{-y} \ dy \qquad , \ x \geq 0 \tag{145}$$

The following relation holds

$$\Gamma(b) = \gamma(b,x) + \Gamma(b,x) \qquad , \ x \geq 0 \tag{146}$$

where

$$\gamma(b,x) = \int_0^x y^{b-1} \ e^{-y} \ dy \qquad , \ x \geq 0 \tag{147}$$

denotes another incomplete Gamma function. Either of the two may be defined to be the incomplete Gamma function while the other becomes the complementary incomplete Gamma function. Sometimes, the incomplete Gamma functions are scaled by $\Gamma(b)$ and then called incomplete Gamma function ratios. Using the incomplete Gamma function defined by eq.(147), the cumulative distribution function $F_{DGL}(d)$ is expressed as

$$F_{DGL}(d) = \begin{cases} \frac{1}{2} - \left\{(1-\alpha)\dfrac{1}{2\Gamma(b_1)} \ \gamma\!\left(b_1, \left(\dfrac{\mu-d}{a_1}\right)^{1/b_1}\right) + \alpha\dfrac{1}{2\Gamma(b_2)} \ \gamma\!\left(b_2, \left(\dfrac{\mu-d}{a_2}\right)^{1/b_2}\right)\right\} \\ \hspace{10cm} , \ d \leq \mu \\[4mm] \frac{1}{2} + \left\{(1-\alpha)\dfrac{1}{2\Gamma(b_1)} \ \gamma\!\left(b_1, \left(\dfrac{d-\mu}{a_1}\right)^{1/b_1}\right) + \alpha\dfrac{1}{2\Gamma(b_2)} \ \gamma\!\left(b_2, \left(\dfrac{d-\mu}{a_2}\right)^{1/b_2}\right)\right\} \\ \hspace{10cm} , \ d \geq \mu \end{cases}$$

$$(148)$$

Subroutines for the computation of the Gamma function $\Gamma(b)$ are readily available in many subroutine libraries, like the NAG Fortran Library, Mark 10 (and higher). Subroutines for the computation of the incomplete Gamma functions are less frequently included in such libraries.

The NAG library for instance does not include such a subroutine. Fortunately, it is not necessary to develop such a subroutine from scratch, because various algorithms for the computation of incomplete Gamma functions are available from within the literature, see Refs. 24 to 30 inclusive. Some of these provide derivatives with respect to the parameter as well. These algorithms are based on continued fraction expansions, and/or Taylor series expansions. The algorithms of Refs. 24 to 27 inclusive all compute the incomplete Gamma function (ratio) $\gamma(b,x)$ of eq.(147). The algorithms used in Refs. 26 and 27 for the computation of the incomplete Gamma function are based on one and the same Taylor series expansion for each pair of arguments (b,x). The algorithms in Refs. 24 and 25 both use a Taylor series expansion for $b \leq x \leq 1$, and also for $x < b$, and a continued fraction expansion otherwise. The latter two, therefore, are generally to be preferred above the former two algorithms. The algorithm of References 28 and 29 also uses Taylor series and continued fraction expansions. An extension of this algorithm compared with the previous four algorithms is, that it is also applicable for negative values of the variable b. Table 5 summarizes some characteristics of the five algorithms. The algorithm of Ref. 24 has been used in the current simulations.

Taking $F_{DGL}(d)$ defined by eq.(148), a sample of size n of the vertical distance d is constructed by numerically solving the separate non-linear equations

$$F_{DGL}(d_i) = u_i \qquad\qquad , \; i=1,2,\ldots,n \qquad\qquad (149)$$

Two considerations play a part with regard to the selection of an appropriate numerical method for solving eq.(149) for $i=1,2,\ldots,n$. These are efficiency and convergence of the method. Efficiency is mainly determined by the number of function evaluations and the convergence rate, provided that convergence as such appears. As only a limited number of simulations was envisaged for the present study, efficiency was given a lower weight than guaranteed convergence, independent of the initial guess. Thus, both Newton's method and the secant method were discarded and a combination of the methods of Regula Falsi and Bisection has been chosen. For random numbers between 0.1 and 0.9, the method of Regula Falsi is used. To start the method, an interval $[\ell,u]$ is needed on which the sign changes of the function of which a zero is to be determined. When the

random number u_i is less than 0.5, u can be taken as the mean value μ whereas otherwise ℓ can be taken to be the mean μ. It remains to define an associated lower bound ℓ and an upper bound u for the two cases $u_i \leq 0.5$ and $u_i \geq 0.5$ respectively. One possibility would be to use the mean minus or plus three to five times the standard deviation of the DGL mixture distribution. A more efficient value, although still being rather conservative, can be obtained as follows. Let $u_i \geq 0.5$ and let an upper bound u of the interval have to be determined. Define a Double Exponential probability density $f_{DE}(x)$ with the standard deviation σ equal to the standard deviation of the tail density within the Double Generalized Laplace mixture density. Thus

$$f_{DE}(x) = \frac{1}{\sigma\sqrt{2}} e^{-\left|\frac{x-\mu}{a}\right|\sqrt{2}} \tag{150}$$

where (compare eq.(13))

$$\sigma = a_2 \sqrt{\frac{\Gamma(3b_2)}{\Gamma(b_2)}} \tag{151}$$

Because of the thick tail of the DE density as compared with the DGL density, the following two requirements

$$\text{Prob } \{\underline{x}_{DE} \leq x_{DE}(0.9)\} = 0.9 \tag{152}$$

and

$$\text{Prob } \{\underline{x}_{DGL} \leq x_{DGL}(0.9) = 0.9\} \tag{153}$$

will certainly result in $x_{DGL}(0.9) < x_{DE}(0.9)$, see Figure 9. Note that the value of 0.9 is the upper bound for the application of the method of Regula Falsi. Hence, for each random number u_i with $0.5 \leq u_i \leq 0.9$ will hold that d_i as a solution of eq.(149) satisfies

$$d_i < x_{DE}(0.9) = -\frac{\sigma}{\sqrt{2}} \log(2(1-0.9)) \tag{154}$$

where σ is defined by eq.(151). So, $x_{DE}(0.9)$ may be taken as the upper bound u of the interval in which the solution d_i is sought. Similarly, $2\mu - x_{DE}(0.9)$ may be used as the lower bound ℓ when $0.1 \leq u_i \leq 0.5$.

For random numbers close to zero or close to one, the method of Regula Falsi was found to converge extremely slowly, because of the fact that the endpoint at the mean μ stayed fixed and the function was very flat near the other endpoint. This phenomenon is well known, see for example Reference 31. As for these small and large values of the random numbers the corresponding realizations d are known to be much smaller or larger than the mean value μ, a practical solution is to use a method that does not retain the mean value μ as an endpoint. The simple Bisection algorithm satisfies this requirement and, therefore, has been used to solve eq.(149) for those values of the random numbers u_i satisfying $0 < u_i \leq 0.1$ or $0.9 \leq u_i < 1$.

6   SOME MAXIMUM LIKELIHOOD PARAMETER ESTIMATION RESULTS FOR SIMULATED DATA

6.1   Introduction

In this section, some maximum likelihood parameter estimation results are described based on the algorithms for grouped data and non-grouped data presented in Section 4. The algorithms are applied to simulated data, which have been generated in the way described in Section 5. Four different probability distributions have been used to generate the data: a Double Exponential, a Gaussian Double Exponential, a Double Double Exponential and a Double Generalized Laplace distribution. For each of these are successively described the associated data sample and the parameter estimates obtained from grouped and non-grouped data. The probability distributions used for parameter estimation are not necessarily the same as the original distributions. For example, the two parameters of a Generalized Laplace distribution are estimated from data generated by the single parameter Double Exponential distribution. The resulting distribution should be close, in some sense, to the original one. When exactly the original distribution is estimated, the comparison can be made at hand of the original and the estimated parameter values. This leads to the validation of the parameter estimation algorithms, while the general case provides insight into the robustness with respect to the underlying probability distribution and the achievable accuracy as a function of the number of estimated parameters. The achievable accuracy,

of course, depends also on the number of data points used for estimating the parameters. Because for the application to the real pairs data the sample size is fixed to 4960 data points and can not be increased to an arbitrarily large value, a fixed sample size of 5000 data points is used for each simulation case.

A second property that might be studied at hand of simulations is the convergence of the Newton iteration process as a function of the initial estimate of the solution. This, however, goes beyond the scope of the present report. The point of view taken here is that it should be shown that convergence of the iteration process occurs for at least one set of initial estimates of the pertinent parameters. A set of initial estimates will, in principle, consist of the true parameter values used in each simulation.

6.2   A Double Exponential probability distribution

6.2.1 The simulated data sample

The Double Exponential probability density is given by

$$f_{DE}(d) = \frac{1}{2a} e^{-\left|\frac{d-\mu}{a}\right|} \tag{107}$$

and the Double Exponential cumulative distribution function by

$$F_{DE}(d) = \begin{cases} \frac{1}{2} e^{(d-\mu)/a} & , d \leq \mu \\ \\ 1 - \frac{1}{2} e^{-(d-\mu)/a} & , d \geq \mu \end{cases} \tag{155}$$

Because of the simple form of this cumulative distribution function, the inversion of eq.(143) can be done fully analytically to obtain a sample of the variable d. A sample of size 5000 was simulated based on the following parameter values:

$$\left.\begin{array}{l} \mu = 0.0 \\ \\ a = 30 \end{array}\right\} \tag{156}$$

The random numbers were, as in all the following simulations, generated by means of the subroutine G05CAF from the NAG subroutine library. The sample statistics are:

$$
\left.\begin{array}{lll}
\text{sample mean} & m = -0.01531 \\
\text{sample standard deviation} & s = 41.364 \\
\text{sample skewness} & \beta_1 = 0.0211 \\
\text{sample kurtosis} & \beta_2 = 5.743
\end{array}\right\} \quad (157)
$$

These sample values correspond fairly well with the population values of 0, 42.426, 0, and 6 respectively. Based on the value of the sample standard deviation s, the sample value of the scale parameter is 29.249. The data grouped into classes of 10 ft are given in Table 6. To generate the 5000 individual data points and to construct the associated frequency table together with the sample statistics, 0.7 CP seconds execution time were needed on the NLR Cyber 180/855 computer.

As in the sections to follow, a comparison will be made between the data samples and the probability densities evaluated on the basis of the maximum likelihood estimates of their parameters. It is worthwhile to make such a comparison between the data sample and the density evaluated for the true parameter values first. Figures 10 and 11 show the Double Exponential density defined by eqs.(107) and (156) and the folded data sample. In general, the correspondence between the two is good. It becomes somewhat worse in the tail region, due to the few data points found there. A statistical criterion for this comparison is given by the standardized log likelihood value defined by eq.(68). Taking m = 27, it holds that twice the standardized log likelihood value is distributed according to a χ^2-distribution with m-1 = 26 degrees of freedom (d.f.). The probabilities p_i, i=1,2,...,m in eq.(68) are defined by eq.(44) where, in this section, the density $f_{DGL}(x)$ reduces to the Double Exponential density $f_{DE}(x)$. However, as in the parameter estimation algorithm for grouped data the exact probabilities p_i are approximated by eq.(46), this approximation could also be used in eq.(68). Table 7 shows the effect of using the approximation (46) rather than the true expression (44) for the probabilities p_i, i=1,2,...,m. Notice that a second simulated sample is included in this table. With regard to the sum of the approximated probabilities it should be realized that it differs from one for two reasons, namely due to the approximations used and due to the fact that,

for the DE density, there is a non-zero probability of deviations which in absolute value are larger than 270 ft. The last probability is in fact equal to 0.000134, showing that the effect of the approximation is the dominant effect. The 95% critical values of a χ^2-test with 26 and 53 degrees of freedom are 38.89 and 71 respectively. Applying these values to twice the standardized log likelihood values of Table 7, would lead to rejecting the true Double Exponential density for each of the cases in which the approximated probabilities are used in the computation of the test statistic. Hence, it is concluded that the exact probabilities should be used in the computation of the test statistic. Notice that for the larger sample, the correspondence between the data and the true DE density, as measured by the test statistic is better.

6.2.2 Parameter estimation based on grouped data

The parameters of three different probability density models have been estimated from the grouped data from the true DE density given by the eqs.(107) and (156). These models are: a DE density, a GL density and a DDE density. The resulting parameter estimates obtained from eq.(113), eqs.(90) and (91), and eqs.(54)-(56) respectively, are:

$$
\left.
\begin{array}{l}
\text{DE} \quad \text{model}: \hat{\hat{a}} = 29.580000 \\
\text{GL} \quad \text{model}: \hat{a} = 31.583465, \quad \hat{b} = 0.95301432 \\
\text{DDE model}: \hat{a}_1 = 29.580000, \quad \hat{a}_2 = 29.580013, \quad \hat{\alpha} = 0.13776644*10^{-8}
\end{array}
\right\}
$$
(158)

Each of these three maximum likelihood solutions gives a negative definite matrix of second partial derivatives of the log likelihood function and defines a (local) maximum of the log likelihood function over the pertinent parameter space.

Parameter estimation based on the DDE model was performed for three sets of initial estimates of α, a_1 and a_2, namely:

i)	0.01	30	90
ii)	0.01	60	90
iii)	0.01	60	60

The estimated values given in eq.(158) were obtained from case i). The Newton iteration process was stopped after 22 iterations, when $\max\{|E_1|, |E_2|, |E_3|\} < 1.0 * 10^{-10}$, with E_1, E_2, and E_3 defined by eqs.(54)

to (56) inclusive. After the first iteration, the maximum of these three quantities was of the order of one hundred. The results show that the maximum likelihood estimates of the two scale parameters of the DDE density converge to the same value, which is, according to the analysis given in Section 4.4.4, the maximum likelihood estimate $\hat{\hat{a}}$ of the scale parameter of the DE density. In this example, the maximum likelihood estimate $\hat{\alpha}$ of the weighting factor converges to zero. As a result, the DDE model reduces to the DE model. Because the numerical values of $\hat{a}_1$ and $\hat{a}_2$ are slightly different and because $\hat{\alpha}$ is not exactly equal to zero, the matrix of second partial derivatives of the log likelihood function remains negative definite rather than becoming singular as in eq.(126) for $\hat{\hat{a}}_1 = \hat{\hat{a}}_2 = \hat{\hat{a}}$. For case ii) holds that the initial estimates are equal and slightly larger than twice the critical value $\hat{\hat{a}} = 29.580000$ (Recall Figure 8). The parameter estimation process was found to diverge in this case, in agreement with the analysis of the iteration process eq.(139) given in Section 4.4.4. The same divergence occurred for case iii). Although the initial estimates of the two scale parameters were not equal in case iii), both are larger than twice $\hat{\hat{a}} = 29.580000$. In this case, apparently, these values were sufficiently close to lead to the same divergence as for identical initial estimates larger than twice the value of $\hat{\hat{a}}$.

For the DE model, the maximum likelihood estimate of the scale parameter can be computed directly from the data. For the GL model, an initial estimate of the shape parameter is needed to start the iterative solution of the non-linear equation $f(b) = 0$ (eq.(75)). This was obtained from a diagram of $f(b)$ against b. The accuracy of the parameter estimates for these two models may be judged at hand of their estimated standard deviations. These are:

$$
\left.
\begin{array}{ll}
\text{DE model : } \sigma(\hat{\hat{a}}) = 0.418 & \\
\text{GL model : } \sigma(\hat{a}) = 1.13, \qquad \sigma(\hat{b}) = 0.0239 &
\end{array}
\right\} \tag{159}
$$

Notice the decrease in accuracy as measured by $\sigma(\hat{a})$ when the shape parameter of the GL density is estimated in addition to the scale parameter. The estimation errors in the scale parameters are about one estimated standard deviation for these two models, whereas the estimation error of the shape parameter is approximately two estimated standard deviations for the GL model. The maximum likelihood estimates of the standard deviation of the DE and the GL model are 41.832 and 41.334 respectively, compared with the population value of 42.426 and a sample value of 41.364.

Figures 12 and 13 show the true DE density and the estimated DE and GL densities on a linear and a logarithmic scale, with the (folded) grouped data superimposed. The true and estimated DE densities are seen to be very close over the entire domain, with the estimated GL density being less accurate in the tail area. Because the shape parameter of this GL density is less than one, the shape of the logarithm of the density is concave in Figure 13. Table 8 gives the corresponding χ^2-test statistic values. Notice the large discrepancy between the numbers based on the exact and the approximated probabilities. The degrees of freedom of each case are equal to the number of classes minus one (m-1), minus the number of estimated parameters. The 95% critical values of a χ^2-test with 25, 24 and 23 degrees of freedom are 37.65, 36.42 and 35.17 respectively. When the exact probabilities are used for computing the test statistic, none of the estimated densities is rejected at the 95% level.

6.2.3 Parameter estimation based on non-grouped data

The DE, GL, and DDE density were also fitted to the non-grouped simulated data sample from the DE density given by eqs.(107) and (156). The parameter estimates based on eq.(109), eqs.(73) and (75), and eqs.(20)-(22) are:

$$
\left.
\begin{aligned}
&\text{DE}\quad \text{model} : \hat{\hat{a}} = 29.285575 \\
&\text{GL}\quad \text{model} : \hat{a} = 28.139422, \hat{b} = 1.027912 \\
&\text{DDE model} : \hat{a}_1 = 29.285575, \hat{a}_2 = 29.286032, \hat{\alpha} = 0.366340*10^{-7}
\end{aligned}
\right\} \quad (160)
$$

The maximum likelihood solutions for the DE and the GL models give a negative definite matrix of second partial derivatives of the log likelihood function over the pertinent parameter space. For the DDE model, the matrix of second partial derivatives is not negative definite. The estimates given by eq.(160) and based on non-grouped data are to be compared with the estimates eq.(158) based on grouped data. For the DE density, the maximum likelihood estimate of the scale parameter is seen to be about one per cent smaller in the former case. For the GL density the estimates of the scale and the shape parameter are now smaller and larger respectively than the true values of 30 and 1.0.

Parameter estimation based on the DDE model was performed only for the set of initial estimates of α, a_1 and a_2 for which convergence occurred in the grouped data case, i.e. $\{\hat{\alpha}_0, \hat{a}_{1_0}, \hat{a}_{2_0}\} = \{0.01, 30, 90\}$. The iteration process was stopped after 19 iterations, when

$\max\{|E_1|, |E_2|, |E_3|\} < \epsilon = 1.0 * 10^{-8}$ with E_1, E_2 and E_3 defined by eqs.(20) to (22) inclusive for the case of non-grouped data. The results show again that the maximum likelihood estimates of both the scale parameters of the DDE density are converging to the same value $\hat{\hat{a}}$ and that $\hat{\alpha}$ converges to zero. Thus, the DDE density reduces to the DE density.

Parameter estimation using a DE model does not require an initial estimate of the scale parameter. As before, an initial estimate of the shape parameter of the GL density was obtained from a diagram of f(b) defined by eq.(75) against b. The accuracy of the parameter estimates of the DE and the GL probability density models may be judged at hand of their estimated standard deviations. These are:

$$\left. \begin{array}{l} \text{DE model} : \sigma(\hat{\hat{a}}) = 0.414 \\ \text{GL model} : \sigma(\hat{a}) = 1.20, \qquad \sigma(\hat{b}) = 0.0280 \end{array} \right\} \quad (161)$$

These values are very similar to those obtained utilizing grouped data. The maximum likelihood estimates of the standard deviation of the DE and th GL density are 41.416 and 41.713 respectively, compared with a population value of 42.426. Because the standard deviation of a GL density depends on both the scale parameter and the shape parameter, its maximum likelihood estimate can be closer to the true value, although its scale parameter is further away from the true value.

Figures 14 and 15 show the true DE density and the estimated DE and GL densities utilizing non-grouped data, on a linear and a logarithmic scale. Just for reference purposes, the histogram has been superimposed. The three densities, although not being identical coincide on the linear scale of Figure 14. On the logarithmic scale of Figure 15, a small difference between the two estimated densities and the true DE density is visible. Notice that the curve representing the estimated GL density is convex (on a logarithmic scale) and lies above the true DE density while the estimated DE density lies below the true DE density. Compare with Figure 13 based on grouped data. The standardized log likelihood values for the estimated DE and GL density, based on non-grouped data, are 5000 and 5139.56 respectively (see eq.(85)).

6.3   A Gaussian Double Exponential probability distribution

6.3.1 The simulated data sample

The Gaussian Double Exponential probability density is given by

$$f_{GDE}(d) = (1-\alpha) \ \frac{1}{a_1 \Gamma(0.5)} \ e^{-\left(\frac{d-\mu}{a_1}\right)^{1/b_1}} + \alpha \ \frac{1}{2a_2} \ e^{-\left|\frac{d-\mu}{a_2}\right|} \tag{162}$$

A sample of size n = 5000 was simulated, based on the following parameter values (the GDE density being a particular DGL density):

$$\left.\begin{array}{ll} \alpha = 0.3 & \mu = 0.0 \\ a_1 = 30.0 & \sigma_1 = 21.2 \\ a_2 = 90.0 & \sigma_2 = 127.3 \\ b_1 = 0.5 & \\ b_2 = 1.0 & \end{array}\right\} \tag{163}$$

The sample statistics are:

$$\left.\begin{array}{lll} \text{sample mean} & m = -0.5438 \\ \text{sample standard deviation} & s = 68.872 \\ \text{sample skewness} & \beta_1 = 0.207 \\ \text{sample kurtosis} & \beta_2 = 16.76 \end{array}\right\} \tag{164}$$

The corresponding population values are 0, 71.937, 0, and 17.655. The data grouped into classes of 10 ft are given in Table 9. To generate the 5000 individual data points, and to form the frequency table together with the sample statistics, 11.4 CP seconds execution time were needed on NLR's Cyber 180/855 computer.

6.3.2 Parameter estimation based on grouped data

The maximum likelihood parameter estimation algorithm as described in Section 4.3 was applied to the grouped, simulated data of Table 9. To this end it has been assumed that the model was a Double Generalized Laplace probability density, of which 2, 3, 4 or 5 parameters were to be estimated corresponding to the cases G_2, G_3, G_4 and G. During the

iteration process, the parameters which were not estimated were kept at their true values. Table 10 shows the estimated parameter values, whereas Table 11 shows the estimated standard deviations. Each row of Table 10 defines a (local) maximum of the likelihood function over the pertinent parameter space. It can be seen that the estimation errors are of the order of one estimated standard deviation. By comparing $\sigma(\hat{a}_1)$ and $\sigma(\hat{a}_2)$, it is seen that the relative accuracy of the scale parameter of the core density is higher than that of the tail density. Intuitively, this is to be expected because the data sample contains more information on the core density than on the tail density. Considering Table 11 further, it is seen that the accuracy for two or three estimated parameters is nearly the same. This may be interpreted as a_1 and a_2 using different parts of the information available in the data sample. However, when the fourth parameter is added to be estimated, the accuracy of the scale parameter of the core density decreases by about fourthy-three per cent. This is due to the fact that the information on the core, available in the data sample, is redistributed over both the scale parameter and the shape parameter of the core density. Similarly, the accuracies of the scale parameter and of the weighting factor of the tail density decrease when the fifth parameter is added to be estimated, while leaving the accuracy of the core density parameters unaffected. In other words, the information on the tail of the distribution, available in the data sample, is redistributed over three parameters (α, a_2, b_2) rather than over two (α, a_2).

When the sample size is sufficiently large, the maximum likelihood parameter estimates should be close to the true parameter values. Hence, a set of good initial estimates for the Newton iteration process should be formed by the true parameter values. These were used for each of the cases G_2, G_3, G_4 and G, giving convergence in 4, 4, 5 and 11 iterations and requiring 2.0, 2.0, 2.3 and 4.5 CP seconds execution time on the NLR Cyber 180/855. The iteration process was stopped whenever the maximum of the pertinent error values $|E_i|$ was smaller than $\epsilon = 1.0*10^{-10}$. In practice, more iterations and computer time will be needed because the initial estimates will usually not be so close to the final estimates. The results show, however, that convergence can occur when the initial estimates are sufficiently good.

In addition to a comparison on the parameter level, a direct comparison of the estimated probability densities with the simulated density and the data is useful. See Figures 16 and 17. From both these diagrams, the true and estimated probability densities are seen to be very close. Notice the difference in the shape of the core and the tail densities in Figure 17. On the logarithmic scale, the core densities are (exactly or approximately) quadratic whereas the tail densities are (exactly or approximately) linear. Except for the five isolated data points in the tails of the histogram, both the true and the estimated probability densities are seen to correspond well with the data.

A statistical measure for the closeness of the analytical probability densities and the grouped data is minus twice the standardized log likelihood value. Table 12 provides this value for the true GDE density as well as for the four estimated densities. These values can be used in a χ^2-test with the number of degrees of freedom as given also in Table 12. The corresponding 95% critical values vary about linearly between 90.53 and 79.08 (these being the 95% critical values of a χ^2-test with 70 and 60 degrees of freedom respectively). The test values are well below the corresponding critical values and none of the analytical probability density models would be rejected at the 95% level. Notice finally the close correspondence between the values of the test statistic based on the exact and the approximated probabilities p_i, i=1,2,...,69. The reason for this is that the approximation error in the probabilities is smaller for each of the five analytical densities in this simulation, compared with the simulation based on a DE density in Section 6.2.2 (see Table 8). For the true GDE density, for example, the sum of the exact probabilities over the 69 intervals equals 0.9998595 giving a total approximation error over these intervals of only 0.0001495. For the DE simulation of Section 6.2.2. the total approximation error over the pertinent 25 intervals was about 32 times larger.

6.3.3 Parameter estimation based on non-grouped data

Assuming a Double Generalized Laplace probability density model of which 2, 3, 4 or 5 parameters were to be estimated, the maximum likelihood parameter estimation algorithm described in Section 4.2 was applied to the non-grouped, simulated data based on the Gaussian Double Exponential

probability density of eqs.(162) and (163). Tables 13 and 14 show the estimated parameter values and their estimated standard deviations respectively. The parameters which were not estimated were again kept at their true values. The results are very similar to those obtained using grouped data. In fact, the present ones are even slightly less accurate. Each row of Table 13 defines a (local) maximum of the likelihood function over the pertinent parameter space. Because the present results are so close to the previous ones, no separate diagrams of the estimated probability densities are given. The standardized log likelihood values for the cases S_2 to S inclusive are 4995.7, 5004.9, 5210.6, and 5217.2 respectively.

The initial estimates of the parameters needed to start the Newton iteration process for each of the cases S_2, S_3, S_4 and S were again equal to the true parameter values. From these, convergence occurred in 4, 4, 5 and 11 iterations, requiring 99.0, 98.9, 120.7 and 226.8 CP seconds execution time. These numbers are about 50 times larger than the amount of computer time needed for the corresponding grouped data cases. This is roughly equal to the ratio of the number of terms in the respective likelihood equations, $5000/69 \approx 72$. The comparison is not completely fair, because the stop citerion ε (recall eq.(32)) was taken as $1.0*10^{-8}$ for the non-grouped data case, whilst being $1.0*10^{-10}$ for the grouped data case.

In practice, a reasonably accurate initial estimate will not generally be available and several attempts may be necessary before convergence occurs. It is advantageous, therefore, to use grouped data first and to use the resulting parameter estimates as initial estimates of the parameters for a final run utilizing non-grouped data.

6.4   A Double Double Exponential probability distribution

6.4.1 The simulated data sample

The Double Double Exponential probability density is given by

$$f_{DDE}(d) = (1-\alpha) \frac{1}{2a_1} e^{-\left| \left(\frac{d-\mu}{a_1} \right) \right|} + \alpha \frac{1}{2a_2} e^{-\left| \frac{d-\mu}{a_2} \right|} \qquad (165)$$

A sample of size n = 5000 was simulated based on the following parameter
values (the DDE density being a particular DGL density):

$$\left.\begin{array}{ll} \alpha = 0.3 & \mu = 0.0 \\ a_1 = 30.0 & \sigma_1 = 42.4 \\ a_2 = 90.0 & \sigma_2 = 127.3 \\ b_1 = 1.0 & \\ b_2 = 1.0 & \end{array}\right\} \quad (166)$$

The sample statistics are:

$$\left.\begin{array}{lll} \text{sample mean} & m = -0.3168 \\ \text{sample standard deviation} & s = 75.362 \\ \text{sample skewness} & \beta_1 = 0.1243 \\ \text{sample kurtosis} & \beta_2 = 12.09 \end{array}\right\} \quad (167)$$

The corresponding population values are 0, 78.230, 0 and 12.98. The data
grouped into classes of 10 ft are given in Table 15. Compare these with
those obtained for the GDE density and given in Table 9. From the 36-th
interval onwards, i.e. from 350 ft onwards, the two tables contain exactly
the same number of data points per class. This is due to the fact that the
tail density is the same for both the mixture densities, and the tail
density becomes dominant in the region of the larger data points. Compared
with the GDE density the number of data points per interval is nearly the
same for the classes 19 to 35 inclusive, indicating that in this region
there is a very small contribution from the core density yet. On the first
three classes, the numbers of data points per class are much higher for
the GDE probability density, because the standard deviation of the Gaussian
core density is much smaller than that of the Double Exponential core
density in the DDE mixture, 21.213 against 42.426, giving a tighter core
for the GDE. As a result, the DDE density gives more data points in the
remaining classes 4 up to 18. To generate the 5000 individual data points,
together with the frequency table and the sample statistics, 10.0 CP
seconds execution time were needed.

6.4.2 Parameter estimation based on grouped data

A (single) Double Exponential probability density was fitted to the
grouped data of Table 15 first, because it is known from Section 4.4.4 to
satisfy the likelihood equations of the family of Double Double

Exponential probability densities. In addition, a (single) Generalized
Laplace density was fitted. The parameter estimation results for these two
densities are:

$$
\left.
\begin{array}{ll}
\text{DE model} : \hat{\hat{a}} & = 46.8000 \\
\qquad \sigma(\hat{\hat{a}}) & = 0.662 \\
\text{GL model} : \hat{a} & = 25.7574 \;,\; \hat{b} = 1.376686 \\
\qquad \sigma(\hat{a}) & = 1.41 \qquad \sigma(\hat{b}) = 0.0318
\end{array}
\right\} \tag{168}
$$

Both the solutions define a (local) maximum of the log likelihood func-
tion. It is difficult to compare these parameter estimates with the true
parameters of the DDE density, because the latter density is a mixture
density with a significant weighting factor of $\alpha = 0.3$ for the tail density,
whilst the DE and GL density are single densities only.

Further, 2, 3, 4 and 5 parameters of a DGL density were estimated
according to the cases G_2, G_3, G_4 and G. The parameters not being esti-
mated were kept at their true values. Tables 16 and 17 present the
estimated values and their estimated standard deviations. The parameter
estimates are consistent in the sense that the estimation errors are of
the order of one estimated standard deviation. Comparing the cases G_2 and
G_3, it is seen from Table 17 that the accuracy of the weighting factor α
decreases by a factor of 2.2 when the scale parameter is additionally
estimated. At the same time, the accuracy of the scale parameter of the
core density decreases by a factor of 1.4. This may be explained from the
fact that the ratio between the core standard deviation and the tail
standard deviation is relatively high (as compared with the GDE case)
making the discrimination between the two more difficult. The transition
from 3 to 4 estimated parameters leads to smaller decreases in accuracy of
the individual parameter estimates. Apparently, the density with the three
estimated parameters discriminates already fairly well between the core
and the tail and the introduction of one more degree of freedom mainly
leads to a redistribution of the information about the core over the
pertinent parameters. Finally estimating the fifth parameter as well
mainly leads to a redistribution of the information about the tail over
the weighting factor and the shape and scale parameter of the tail den-
sity, leaving the accuracy of the core density parameters relatively
unaffected.

It is interesting to make a comparison between the accuracy of the parameter estimates for the GDE and the DDE density. Table 18 presents the ratios between two corresponding estimated standard deviations. Except for the scale and shape parameter of the tail density in case G_5, the accuracy is, on average, about two times higher for the GDE density.

As for the GDE density, the true parameter values were used as initial estimates for the Newton iteration process. Convergence occurred for the cases G_2 and G_3 in 4 and 5 iterations respectively, based on a stop criterion of $\varepsilon=1.0*10^{-10}$, wheras divergence occurred for the cases G_4 and G. Convergence was obtained, by trial and error, for the latter two cases from the following sets of initial estimates α_{0_0}, a_{1_0}, a_{2_0}, b_{1_0}, b_{2_0}: 0.2, 30, 90, 1, 1 and 0.285, 30, 90, 0.9, 0.94. This required 11 and 28 iterations respectively.

A direct comparison between the estimated probability densities and the simulated density and the data can be made by means of Figures 18 to 21 inclusive. Figures 18 and 19 show the simulated DDE density and the corresponding data, together with the estimated DE and GL densities. Except for the last isolated data points, the true DDE corresponds well with the histogram. Figure 19 shows that the single DE probability density is unable to represent the tail of the underlying Double Double Exponential density. The fit is better for the single GL density which is more flexible as it has two parameters. Notice that the shape of the GL curve on the logarithmic scale of Figure 19 is convex due to the value of the shape parameter being larger than one. Although the GL density is closer to the true DDE density and the data than the DE density, it is still not a very good representation. Table 19 confirms the bad fit of the single DE and GL densities through the very large values of the χ^2-statistic, in particular for the DE density. Both single probability densities would clearly be rejected at the 95% level. Notice also the large difference between the two values of the χ^2-test statistic for the GL density when based on the exact and the approximated probabilities p_i, i=1,2,...,69. The difference of the sum of the approximated probabilities from one is indicative for the bad approximation.

Figures 20 and 21 show the true DDE density and the four estimated mixture densities. Three of the four estimated probability densities are very close to the true DDE density. All the five densities describe the

data well, except for the last isolated data points. Notice that for each of the five mixture densities the shape of the core and the tail density is (exactly or approximately) linear on the logarithmic scale of Figure 21, but with a different slope due to the different values of the scale parameters of the tail and core densities. Table 19 also provides the χ^2-test statistic values for the four estimated mixture densities. Case G_4 and G give the highest values. Compared with the results for the GDE simulation, the test values are approximately 20 per cent higher. This is consistent with the higher accuracy of the parameter estimates for the GDE simulation. The last column of Table 19 gives approximate 95% critical values for each case, obtained by linear interpolation of the critical values for 60 and 70 degrees of freedom. The true DDE density as well as the DDE densities with two and three estimated parameters are accepted on the basis of a χ^2-test at the 95% level. The estimated Generalized Laplace Double Exponential density (case G_4) would also be accepted, but the Double Generalized Laplace density (case G) would be rejected at the 95% level. Although this is not immediately clear from the Figures 20 and 21, the large uncertainty of the parameter estimates for the case G may make this plausible. As a final technical point, it should again be noticed that the computation of the χ^2-test statistic values on the basis of the approximated probabilities p_i leads to extremely inaccurate numerical results. It would, in fact, also lead to incorrect statistical conclusions.

6.4.3 Parameter estimation based on non-grouped data

Assuming a Double Generalized Laplace probability density model of which 2, 3, 4 or 5 parameters were to be estimated, the maximum likelihood parameter estimation algorithm was applied to the non-grouped, simulated data based on the Double Double Exponential probability density of eqs.(165) and (166). The parameters not being estimated were kept at their true values. The true values were also used as initial estimates for the Newton iteration process. Different results were obtained for the cases S_2 and S_3 on the one hand and S_4 and S on the other. The estimation results are given in Table 20 and 21.

Convergence to $\varepsilon=1.0*10^{-10}$ occurred in 4 and 7 iterations for the cases S_2 and S_3, requiring 98.8 and 162.0 CP seconds execution time. The estimation errors for these two cases are of the order of one estimated

standard deviation or less. The point estimates for case S_2 are slightly more accurate than those for the corresponding case G_2, while for case S_3 they are slightly less accurate than for the grouped data case G_3.

Twenty-four iterations were required to obtain the solution given in Table 20 for case S. This solution of the likelihood equations, however, does not define a maximum of the log likelihood function. The estimated standard deviations could not be computed, because the matrix of second partial derivatives was not negative definite. A number of heuristically chosen initial parameter estimates were tried further. None of them resulted in convergence of the iteration process. Similar divergence problems were met for case S_4. The only solution obtained for this case was a (single) Double Exponential solution (different from the "real" DE model given by eq.(168)). Notice that the estimate of the scale parameter a_1 is close to the true value of a_1, but that the tail density is completely ignored. Some further investigations of the cases S_4 and S are necessary.

6.5   A Double Generalized Laplace probability distribution

6.5.1 The simulated data sample

The Double Generalized Laplace probability density is given by

$$f_{DGL}(d) = (1-\alpha)\, \frac{1}{2a_1 b_1 \Gamma(b_1)}\, e^{-\left|\left(\frac{d-\mu}{a_1}\right)\right|^{1/b_1}} + \alpha\, \frac{1}{2a_2 b_2 \Gamma(b_2)}\, e^{-\left|\frac{d-\mu}{a_2}\right|^{1/b_2}} \qquad (170)$$

A sample of size n = 5000 was simulated, based on the following parameter values:

$$\begin{aligned}
\alpha &= 0.3 & \mu &= 0.0 \\
a_1 &= 30.0 & \sigma_1 &= 25.2 \\
a_2 &= 90.0 & \sigma_2 &= 93.0 \\
b_1 &= 0.65 & & \\
b_2 &= 0.80 & &
\end{aligned} \qquad (171)$$

The sample statistics are:

$$
\left.\begin{array}{lll}
\text{sample mean} & m = -0.1416 \\
\text{sample standard deviation} & s = 52.900 \\
\text{sample skewness} & \beta_1 = 0.0298 \\
\text{sample kurtosis} & \beta_2 = 10.06
\end{array}\right\} \tag{172}
$$

The corresponding population values are 0, 55.122, 0 and 11.100 respectively. The data grouped into classes of 10 ft are given in Table 22. Notice that, compared with the simulated GDE and DDE data, the range of values is much smaller, due to the thinner tail of a GL density with a shape parameter of 0.8. On the other hand, the range of this particular mixture density is larger than for the single DE density considered in Section 6.2 (see Table 6). To generate the individual data points, the sample statistics and the frequency table, a total of 85 CP seconds execution time was needed.

6.5.2 Parameter estimation based on grouped data

The maximum likelihood parameter estimation algorithm as described in Section 4.3 was applied to the grouped, simulated data of Table 22. In accordance with the cases G_2, G_3, G_4 and G, 2, 3, 4 and 5 parameters were estimated while keeping the parameters not being estimated at their true values. Tables 23 and 24 show the estimated parameter values and their associated standard deviations. The estimation errors are of the order of one standard deviation or smaller. The addition of one more parameter to be estimated leads in each of the cases to a considerable redistribution of the information available in the pertinent part of the sample over the parameters, see for example the increase in $\sigma(\hat{a})$ at the transition from case G_2 to G_3. The resulting accuracy for the DGL density is between that of the GDE and DDE densities, the only exception being the accuracy of the scale and shape parameter of the tail density in the DGL mixture.

The true parameter values were used as initial estimates for the Newton iteration process, for each of the cases G_2, G_3, G_4 and G. Convergence occurred in 4, 4, 7 and 15 iterations respectively, requiring 1.1, 1.2, 1.7 and 3.1 CP seconds execution time.

A graphical comparison of the simulated and the four estimated Double Generalized Laplace probability densities is given in Figures 22 and 23. The small difference between the densities are not visible on the linear scale of Figure 22. Even on the logarithmic scale of Figure 23, the differences between the true and the estimated DGL densities are hardly visible in the region between -400 and +400. The correspondence between all the densities and the data in that region is good. Outside this region, the DGL density corresponding with case G (5 estimated parameters) deviates considerably form the others. This is due to the large inaccuracy of the estimates of the scale and the shape parameter of the tail density within the mixture in this case. Notice that the logarithm of each DGL density consists of a concave core part and a concave tail part which are connected to each other in a transition region.

Table 25 provides the χ^2-test statistic values for the true and the four estimated DGL densities. The highest and the lowest value occur for the true DGL density, with no estimated parameters, and the estimated DGL density of case G with five estimated parameters. The 95 per cent critical values of a χ^2-test with 30 and 40 degrees of freedom are 43.77 and 55.76. Hence, each of the five probability densities would pass a χ^2-test at the 95 per cent level, when the exact probabilities p_i would be used. The difference between the values of the test statistic based on the exact and approximated probabilities varies between 16.4 and 19.2 per cent. Although a test based on the approximated probabilities would not lead to wrong statistical decisions, it is better not to use the approximated probabilities due to their limited accuracy.

6.5.3 Parameter estimation based on non-grouped data

The results of the application of the maximum likelihood parameter estimation algorithm for non-grouped data are given in Table 26 and 27. As before, the parameters which were not estimated were kept at their true values. For the cases S_2 and S_3, the parameter estimates as well as their estimated standard deviations are very similar to their G_2 and G_3 equivalents. For these two cases, the estimation errors are of the order of one estimated standard deviation or less. For the case S_4, the results based on non-grouped data are even less accurate than those for the grouped data based case G_4, except for the estimate of the scale parameter of the tail density. For this case, the estimation errors in the scale

parameter and the shape parameter of the core density are 1.5 and 2.3 times the estimated standard deviations respectively. These large ratios are probably due to a random fluctuation. The results for case S are rather bad, yielding estimation errors of the order of two estimated standard deviations, except for the scale parameter of the core density.

The initial estimates of the parameters for the Newton iteration process were again equal to the true parameter values for each of the four cases. For the cases S_2, S_3 and S_4, this led to convergence in 4, 4 and 6 iterations respectively, requiring 89.1, 99.2 and 142.4 CP seconds execution time when using a stop criterion of $\varepsilon=1.0*10^{-10}$. By some trial and error, convergence in 12 iterations occurred for case S starting from the following initial estimates α_0, a_{1_0}, a_{2_0}, b_{1_0}, b_{2_0}: 0.24, 30, 90, 0.65 and 0.80.

Graphically, the estimated DGL densities based on simulated non-grouped data look very similar to those based on grouped data and, therefore, are not separately presented. The standardized log likelihood values for the cases S_2 to S inclusive are 4826.1, 4846.1, 5053.8, and 5158.56 respectively.

7 RESULTS FOR THE REAL AIRCRAFT PAIRS DATA

7.1 Introduction

The results of the application of the maximum likelihood parameter estimation technique described in Section 4 to the aircraft pairs data summarized in Section 2 are given in this section. Several subfamilies of probability distributions belonging to the general family of Double Generalized Laplace probability distributions will be considered. Three single probability distributions will be analyzed first, namely the Gaussian, the Double Exponential and the Generalized Laplace probability distribution (Section 7.2). Three different mixture distributions, each having a Double Exponential tail distribution will be discussed next in Section 7.3. The reason for paying so much attention to mixture distributions with a Double Exponential tail is that these lead to a cautious extrapolation of the tails of the observed distribution. Finally, the Double Generalized Laplace probability distribution will be presented (Section 7.4).

The data used in this section are the grouped aircraft pairs data of Table 3. Because the parameter estimation results based on simulated data showed, in general, a close correspondence between the two cases of grouped and non-grouped data, only grouped data are used. This is similar to the data being used elsewhere (References 14 – 18). For each probability density model to be analyzed, a mean value of 1000 ft will be assumed. This is slightly larger than the sample value of 997.1 ft.

The main results to be presented consist of the estimated probability densities and the corresponding estimated probabilities of vertical overlap in a 1000 ft environment, $P_z(1000)$. The latter probabilities are defined by

$$P_z(1000) = \int_{-\lambda_z}^{\lambda_z} f_{\underline{d}}(x) \, dx \tag{173}$$

where $f_{\underline{d}}(x)$ denotes any probability density model of the vertical distance $\underline{d}$ between the aircraft in a pair. In practice, $P_z(1000)$ may be approximated by

$$P_z(1000) = 2\lambda_z \, f_{\underline{d}}(0) \tag{174}$$

The symbol λ_z in eqs.(173) and (174) denotes the average height of an aircraft. Following Ref. 6, the value to be used is $\lambda_z = 40.9$ ft.

A brief summary of the results obtained will be given in Section 7.5. Moreover, one particular probability density model will be selected as being the most useful with regard to the practical objectives of the study.

7.2   Results based on some single probability distributions

Tables 28 and 29 summarize the parameter estimation results for the Gaussian, the Double Exponential, and the Generalized Laplace probability densities, based on the grouped data of Table 3. The shape parameter of the Generalized Laplace density is 0.727811, that is, approximately midway between the values characterizing the Gaussian and the Double Exponential densities (0.5 and 1.0 respectively). The estimated standard deviation

$\sigma(\hat{a})$ of the scale parameter estimate $\hat{a}$ is about twice as large for the Generalized Laplace density as for the Gaussian and the Double Exponential densities. This is due to the fact that the Generalized Laplace density is characterized by two parameters rather than one, over which the information available in the data has to be distributed. The relative accuracy of the parameter estimates of the Generalized Laplace density is approximately 2.5 to 3 per cent, whereas for the Gaussian and the Double Exponential density this is approximately 1 and 1.5 per cent respectively. These accuracy figures, however, are only relevant on the assumption that the probability distributions do represent the data. This will be evaluated in more detail below.

Figures 24 to 29 inclusive pertain to the Gaussian density. Figure 24 shows the histogram of the pairs data with a class interval of 10 ft, with the estimated Gaussian density superimposed. Similarly, Figure 25 shows the logarithm (base e) of the histogram and of the Gaussian density, in order to provide more insight into the tail region. It is seen that the Gaussian density does not at all fit to the data. The core of the observed histogram around 1000 ft is underestimated. The histogram seems to be more peaked than the Gaussian density. This is confirmed by the kurtosis values of the histogram and the Gaussian density in Table 29: 4.33 and 6. The maximum likelihood estimate of the standard deviation of a Gaussian density is by definition equal to the observed sample standard deviation (see eqs.(103) and (106)). The tails of the histogram are also seen to be underestimated and, as a result, the intermediate part of the histogram is overestimated.

The quality of the fit can be studied in more detail by examination of the residual plots. The (linear) residuals on class $\pm i$, $i=1,2,\ldots,63$ (counted from 1000 ft onwards to the right and the left respectively) are defined by

$$r_i = p_i - h_i \tag{175}$$

where h_i denotes the height of the histogram on class i and p_i denotes the value of the probability density model, evaluated at the midpoint of class i and at the maximum likelihood parameter estimates. The logarithmic residuals on class $\pm i$ are defined by

$$R_i = \log p_i - \log h_i = \log \frac{p_i}{h_i} \qquad (176)$$

where p_i and h_i are defined as above (note: log denotes the natural logarithm, i.e. base e). Figures 26 and 27 show the linear and the logarithmic residuals of the Gaussian density. The log residual values on the isolated classes ±62 and 63 are approximately -7.

A somewhat different view of the fit, which is particularly useful for the purpose of extrapolation to the probability of vertical overlap $P_z(1000)$, is given by the "1-cumulative" curve, i.e. the probability of vertical distances between aircraft in a pair larger than 1000+x or smaller than 1000-x ft. Figure 28 shows this curve for the Gaussian density. The solid line marks the curve based on the Gaussian density, whereas the +signs mark the corresponding quantity based on the observed pairs data. As might have been expected from the foregoing discussion, the correspondence between the two curves is not good. Figure 29 amplifies this for the tail region, by means of the logarithm of the "1-cumulative" curves. The logarithmic "1-cumulative" curve drops off quadratically for the Gaussian density. Notice that the logarithmic "1-cumulative" curve based on the observed pairs data is flat on the intervals 51 to 61 inclusive, because there are not any observations in the intervals 52 to 61 inclusive (see Table 3).

The bad fit of the Gaussian probability density to the data as examined so far graphically, is expressed mathematically by the very high value of minus twice the standardized log likelihood value given in Table 28: 257.80. Compared with the 95 per cent critical value, the Gaussian density is formally rejected in a χ^2-test.

Although the Gaussian density is not acceptable as a probability density model for the data, it might, from a theoretical point of view, still be used for computing an estimate of the probability of vertical overlap. Table 28 provides a value of $3.78*10^{-15}$, which, however, has no practical significance.

Consider now the Double Exponential density in Figures 30 to 35 inclusive. Figure 30 shows the histogram and the estimated Double Exponential density. The value of the Double Exponential density at the

mean is $1/\hat{a} = 0.0105$ and exceeds the scale of Figure 30. The Double
Exponential density is seen to be more peaked than the histogram. This is
confirmed by the corresponding kurtosis values in Table 29: 6 and 4.33.
The Double Exponential density almost shows the reverse picture of the
Gaussian density. It overestimates the very core of the histogram,
underestimates the intermediate part and overestimates the remaining part,
except for the two isolated data points in the extreme tails (see Figs. 30
and 31). The linear and logarithmic residual plots, Figure 32 and 33,
demonstrate this nicely. Notice that the magnitude of the logarithmic
residuals of the Double Exponential density is considerably smaller than
that of the Gaussian density (compare Figs. 27 and 33). Table 29 explains
why the Double Exponential density overestimates the larger part of the
observed histogram in the tail area. The reason is the relatively large
value of the maximum likelihood estimate of the standard deviation
compared to the sample value: 134.097 and 125.2. Figures 34 and 35 show
the "1-cumulative" curves and their logarithms. Notice that the logarithm
of the "1-cumulative" curve of the Double Exponential density in Figure 35
is a straight line. The correspondence between the estimated and the
observed curves is not good. The extrapolation of the tail area is too
pessimistic.

It is concluded from Figures 30 to 35 inclusive, that the Double
Exponential density is not an appropriate model of the data. Mathematically,
the X^2-test confirms this view. The computed value of 211.18 of minus
twice the standardized log likelihood, far exceeds the 95 per cent critical
value of this test, see Table 28. Table 28 also gives a value of the
probability of vertical overlap, namely $P_z(1000) = 1.13*10^{-5}$. This value,
however, is deemed to be too pessimistic for the practical situation.

As a final single probability density model of the pairs data, the
Generalized Laplace density will be considered. Recall from Section 3.2 on
"Selection of a family of probability distributions", that the family of
Generalized Laplace densities includes both the Gaussian and the Double
Exponential density as a particular member. The previous analysis suggests,
in fact, two things. Firstly, the probability density model might be
sought in between the Gaussian and the Double Exponential density. Second-
ly, this density should be more flexible than a probability density
characterized by a single parameter only. The Generalized Laplace density,
therefore, would seem to be an appropriate next step.

Figures 36 to 41 show the results for the family of Generalized Laplace probability densities. Figure 36 of the histogram and of the estimated Generalized Laplace density shows a considerable improvement in the fit, compared to the Gaussian and the Double Exponential densities. Notice in particular the representation of the core of the histogram by the Generalized Laplace density. Table 29 demonstrates that the estimates of the kurtosis of the histogram and of the Generalized Laplace density are very close: 4.33 and 4.09. In addition, the maximum likelihood estimate of the standard deviation corresponds very well to the sample value. Figure 37 shows that the fit of the Generalized Laplace density has also markedly improved on a logarithmic scale. See also the residual and "1-cumulative" plots in Figures 38 to 41 inclusive. All these observations are confirmed mathematically by a great reduction of the value of minus twice the standardized log likelihood in Table 28: 77.17 rather than a value of the order of 200. The present value is seen to be below the critical value of a χ^2-test at the 95 per cent level and could lead to formally accepting the Generalized Laplace probability density model.

Before doing so, the ultimate objective of the modelling process should be recalled. This is to provide a best (but not too optimistic) estimate of the probability of vertical overlap $P_z(1000)$ in a 1000 ft environment. As should be clear from the various diagrams, this estimate has to be based on the extrapolation of the tails of the histogram by means of the probability density model. As the value of the shape parameter of the Generalized Laplace density is 0.727811, the logarithm of this density is concave (see Figure 37) and drops off relatively quickly in Figure 37. Therefore, the corresponding estimate of the probability of vertical overlap, being $P_z(1000) = 4.68*10^{-8}$ (see Table 28), might be somewhat too optimistic. In fact, by comparing the Figures 33 and 39, the log residuals of the Generalized Laplace density on the two extreme intervals ±62 and 63 of the histogram are seen to be twice as large as those of the Double Exponential density. The logarithmic "1-cumulative" curves in Figure 41 also show that the extrapolation using the estimated Generalized Laplace density is not by definition cautious. Taking these observations into account, it is concluded that the estimated Generalized Laplace probability density is not completely satisfactory for the practical objectives. Rather than using a single Generalized Laplace density, a mixture of two Generalized Laplace densities should be applied and analyzed. Perhaps one or more of the parameters of such a mixture could have a priori specified values.

7.3 Results based on some Double Generalized Laplace probability distributions
with a Double Exponential tail distribution

Three different mixture densities, each having a different core
density but with Double Exponential tail density will be analyzed in this
subsection. These probability densities are:

- the Gaussian Double Exponential (GDE)
- the Double Double Exponential (DDE)
- the Generalized Laplace Double Exponential (GLDE).

The Gaussian Double Exponential density has three unknown parameters α,
a_1, a_2 and two fixed parameters b_1=0.5 and b_2=1.0. The Double Double
Exponential density has the same unknown parameters but the fixed
parameters are b_1= b_2= 1.0. The Generalized Laplace Double Exponential
density has four unknown parameters α, a_1, a_2, b_1 and one fixed parameter
b_2 = 1.0. The complexity of the parameter estimation process is slightly
less for the Gaussian Double Exponential and Double Double Exponential
densities because of the smaller number of parameters to be estimated.
These mixture densities, therefore, are considered first.

As an alternative to a theoretical analysis of the existence and
uniqueness of a solution of the likelihood equations based on a mixture
probability density model, a grid may be constructed in the pertinent
parameter space, and each grid point be used as an initial estimate for
the Newton iteration process. The maximum dimension and the resolution of
the grid need to be sufficiently large for this approach to be successful.
On the other hand, these two quantities should be as small as possible in
order that the approach is computationally efficient.

For the Gaussian Double Exponential density, the alternative
approach means that a grid in the (α,a_1,a_2)-parameter space has to be
constructed. Rather than utilizing a full three-dimensional grid, a grid
for the scale parameters a_1 and a_2 is constructed in the plane α=0.1.
Although a_1-a_2 planes defined by different values of α could also be
considered, the results obtained from the present choice are believed to
be sufficient. The dimensions of the grid in the (a_1,a_2)-plane chosen may
be derived from the magnitude of the sample standard deviation, because
the scale parameters a_1 and a_2 are related to the standard deviation of
the core and the tail density respectively (see eqs.(12)-(14)). Based on
a sample standard deviation of s=125.2 ft, the following (a_1,a_2)-grid was
chosen in the plane α_0=0.1:

$$a_1 = 30,40,\ldots,250 \quad ; \quad (\sigma_1 = 21.2,\ 28.3,\ \ldots,\ 176.8)$$
$$a_2 = 30,40,\ldots,250 \quad ; \quad (\sigma_2 = 42.2,\ 56.6,\ \ldots,\ 353.6)$$
$$\left.\right\} \quad (177)$$

The numbers between the brackets denote the corresponding grid in the (σ_1,σ_2)-plane. Notice the difference in the range of σ_1 and σ_2. Using this grid and the grouped data of Table 3, six solutions of the likelihood equations pertaining to the Gaussian Double Exponential probability density model have been obtained which meet both the stop criteria: $\max\{|E_1|,|E_2|,|E_3|\} < \varepsilon = 1.0*10^{-10}$ and the number of iterations less than or equal to 25. These solutions are given in Table 30. In addition, Figure 42 shows which grid point(s) converged to which solution. It should be remarked that the line pieces do not indicate the path followed by the successive iterants between the initial and the final value; they just connect the initial and the final values. The initial value of α for each solution is 0.1 and the final value can be read from Table 30.

The fifth solution of the likelihood equations as shown in Table 30 does not define a maximum of the log likelihood function, but a stationary point. In addition, it does not meet the requirement $0 \leqq \alpha \leqq 1.0$. The sixth solution does define a maximum of the log likelihood equation, but does not either meet the requirement of the weighting factor being between zero and one. The fourth solution is exactly the maximum likelihood solution of a (single) Gaussian probability density. In this case, the value of $\hat{a}_2$ is not of any importance. The second and the third solution are exactly equal to the maximum likelihood solution of a single Double Exponential density based on grouped data with a class interval of 10 ft. The difference between the two estimates of a_1 does not have a physical meaning, because the weighting factor of the core density is zero for these two solutions. So, both the (single) Gaussian and the (single) Double Exponential probability density satisfy the likelihood equations of the family of Gaussian Double Exponential mixture densities. Both these single densities, however, were shown not to be practically acceptable in Section 7.2.

From a numerical analysis point of view, it is interesting to remark that the set of gridpoints for which the two stopcriteria are not met can be subdivided into two classes. The first class consists of those points for which $M = \max\{|E_1|,|E_2|,|E_3|\}$ either oscillates around a certain finite (non-zero) value or increases at each iteration. The second class consists of those points for which M decreases very slowly to zero at each

iteration. After a certain number of iterations, M decreases by a factor
of approximately two per step. In some cases, all the three parameter
estimates are seen to jointly diverge to plus or minus inifinity (by a
factor of two at each iteration). In the other cases, the estimates of
only one or two of the three parameters show such divergence, with the
remaining two or one parameter(s) converging to a finite value. For each
of the grid points belonging to the second class, the stop criterion $M<\varepsilon$
could be met when the number of iterations would be allowed to become
sufficiently large. The resulting solutions, however, do not have any
practical meaning.

The first solution in Table 30 was analyzed further. The parameter
and additional estimates are summarized in Table 31. These indicate that
the Double Exponential tail density has a relatively large weighting
factor in the overall mixture. Its standard deviation is approximately 20
feet larger than that of the Gaussian core density. Notice that the
accuracy of the scale parameter estimates as measured by $\sigma(\hat{a}_1)$ and $\sigma(\hat{a}_2)$
is approximately the same. The maximum likelihood estimate of the overall
standard deviation of the Gaussian Double Exponential density is close to
the sample standard deviation. The maximum likelihood estimate of the
kurtosis is relatively large compared to the sample value of 4.33. This is
due to the heavy weighting of the kurtosis of the Double Exponential tail
density.

Figures 43 to 48 inclusive show the estimation results for the
Gaussian Double Exponential density graphically. Notice that, in agreement
with the estimated kurtosis value, the Gaussian Double Exponential density
is relatively heavily peaked about the mean value of 1000 ft. Compared
with the (single) Gaussian and the (single) Double Exponential density,
the fit has considerably improved. See the Figures 43 to 46 inclusive. For
the Gaussian Double Exponential density, the logarithmic residuals on the
four extreme classes are somewhat larger than for the Double Exponential
density. Compared with the Generalized Laplace density (see Figures 36 to
41 inclusive) these residuals are seen to be smaller. The quality of the
fit as measured by minus twice the standardized log likelihood value is
78.25, which is comparable with that for the Generalized Laplace density,
but just exceeds the 95 per cent critical value of a χ^2-test with 59
degrees of freedom. Compare finally the linear and logarithmic "1-cumula-
tive" curves of Figures 28, 34, 40, 47 and 29, 35, 41, 48. The correspon-

dence between the estimated and the observed curve is seen to be much
better for the Gaussian Double Exponential density than for the (single)
Gaussian and Double Exponential density. The results for the Gaussian
Double Exponential and the Generalized Laplace density are similar for
deviations from the mean up to approximately 340 feet, with the former
being at least as good and more cautious for the larger deviations without
being as pessimistic as the (single) Double Exponential. Hence, the
Gaussian Double Exponential is considered to be well suited for extrapo-
lation purposes. The corresponding estimate of the probability of vertical
overlap is $P_z(1000) = 6.59*10^{-6}$ (see also Table 31).

So far, the discussion has concentrated on the point estimates,
obtained by means of the maximum likelihood method, of the parameters of
the Gaussian Double Exponential probability density model. Further insight
is provided by interval estimates or confidence regions of the parameters.
There are various ways to obtain these using asymptotic approximations
valid for large sample sizes (Ref. 23). One possibility is to use the
Gaussian asymptotic distribution of the maximum likelihood estimator. This
involves eq.(41) or eqs.(42) and (43), adapted to the Gaussian Double
Exponential density and grouped data. The result is a $\gamma*100\%$ confidence
ellipsoid in the (α, a_1, a_2)-parameter space, centred at the maximum
likelihood point estimates. A second possibility is to use the likelihood
ratio statistic. This involves eqs.(37) and (38) of Section 4, similarly
adapted. Its asymptotic distribution is a χ^2-distribution. The advantage
of the likelihood ratio statistic over the maximum likelihood estimator
for calculating confidence regions is, that the former usually approaches
its asymptotic distribution faster, and, as a result, gives tighter
confidence bounds. Because it is graphically impossible to represent the
joint confidence regions of all the three parameters in a single diagram,
confidence regions in the various co-ordinate planes of the (α, a_1, a_2)-para-
meter space will be given below.

Figures 49, 50 and 51 show the isocontours of the relative
likelihood function corresponding with the 90 and 95 per cent critical
values of a χ^2-distributed random variable with two degrees of freedom
(eq. 40). Notice that the contours are pretty regularly shaped and that in
particular the contours in the (a_1, a_2)-plane are close to an elliptical
shape. Figures 52, 53 and 54 show isocontours based on the 95 per cent
critical value of the asymptotic maximum likelihood estimator. These
contours are based on the following asymptotic, observed, sub-covariance
matrices:

$$\text{cov}\{\hat{\alpha},\hat{a}_1\} = \begin{pmatrix} 0.00237 & 0.0676 \\ 0.0676 & 28.9 \end{pmatrix}, \text{cov}^{-1}\{\hat{\alpha},\hat{a}_1\} = \begin{pmatrix} 452.772 & -1.05947 \\ -1.05947 & 0.0370987 \end{pmatrix}$$

$$\text{cov}\{\hat{\alpha},\hat{a}_1\} = \begin{pmatrix} 0.00237 & -0.0988 \\ -0.0988 & 20.5 \end{pmatrix}, \text{cov}^{-1}\{\hat{\alpha},\hat{a}_2\} = \begin{pmatrix} 529.099 & 2.55372 \\ 2.55372 & 0.0611868 \end{pmatrix}$$

$$\text{cov}\{\hat{a}_1,\hat{a}_2\} = \begin{pmatrix} 28.9 & -18.9 \\ -18.9 & 20.5 \end{pmatrix}, \text{cov}^{-1}\{\hat{a}_1,\hat{a}_2\} = \begin{pmatrix} 0.0872645 & 0.0805227 \\ 0.0805227 & 0.123163 \end{pmatrix}$$

The corresponding correlation coefficients are: $\rho\{\hat{\alpha},\hat{a}_1\}=0.26$, $\rho\{\hat{\alpha},\hat{a}_2\}=$ -0.45, and $\rho\{\hat{a}_1,\hat{a}_2\}=-0.78$. The isocontours defined by the above covariance matrices are known to be ellipses in the various co-ordinate planes, centred at the maximum likelihood point estimates. The ellipses in Figures 52 and 53 look somewhat deformed because of the stretching of the vertical scale. Consequently, their axes are not perpendicular. The difference in the orientation of the ellipses of Figures 52 and 53 is due to the different signs of the correlation coefficients $\rho\{\hat{\alpha},\hat{a}_1\}$ and $\rho\{\hat{\alpha},\hat{a}_2\}$. The true angles between the ξ-axes of the ellipses in the Figures 52 and 53 and the a_1 and the a_2 axis are 0.13° and 89.87° respectively. The two types of isocontours are very similar for the pair of parameters a_1 and a_2. To a somewhat lesser extent, this also holds for the pair of parameters α and a_1, with the contour based on the relative likelihood function being somewhat smaller. Given the similarity in the shape, this is in agreement with the theoretically expected result. There is a similar difference in the size and an additional difference in the orientation of the two types of isocontours for the pair of parameters α and a_1.

Approximate confidence bounds on the parameters α, a_1, and a_2 of the Gaussian Double Exponential probability density model may be defined by the minimum and maximum parameter values along the isocontours. Similarly, approximate confidence bounds on the probability of vertical overlap $P_z(1000)$ may be defined by the minimum and maximum values of $P_z(1000)$, evaluated along the various isocontours. Thus, the following appropriate 95 per cent confidence bounds (conservatively rounded) have been obtained:

parameter	approximate 95 per cent confidence bounds based on	
	relative likelihood function	maximum likelihood estimator
α	0.31 – 0.55	0.31 – 0.55
a_1	152 – 172	151 – 171
a_2	88 – 108	87 – 109
$P_z(1000)$	$2.3*10^{-6}$ – $14.8*10^{-6}$	$2.0*10^{-6}$ – $16.7*10^{-6}$

As was to be expected from the discussion on the diagrams of the isocontours, the bounds on the parameters based on the two approaches are practically the same. The bounds on the probability of vertical overlap $P_z(1000)$ derived from the relative likelihood function are about ten per cent tighter than those based on the maximum likelihood estimator. The major question with regard to deriving a bound on a quantity is to obtain a bound which is as tight as possible. Because the relative likelihood method generally gives tighter bounds and because the various isocontours are regularly shaped, the bounds based on this method are taken to define a 95 per cent confidence interval for the probability of vertical overlap $P_z(1000)$, namely $2.3*10^{-6}$ – $14.8*10^{-6}$. The upper bound is approximately 2.2 times larger than the point estimate of $6.59*10^{-6}$.

The second probability density model belonging to the class of densities considered in this subsection is the Double Double Exponential probability density. It is characterized by exactly the same parameters as the Gaussian Double Exponential density, namely a weighting factor α and two scale factors a_1 and a_2. Hence, the same grid as used for the Gaussian Double Exponential density might, in principle, be used to define initial guesses for the Newton iteration process to solve the likelihood equations pertaining to the family of Double Double Exponential probability densities. This was actually done for a sub-grid defined by $a_1 = 30, 40, \ldots, 140$. Most of the grid points gave convergence to the (single) Double Exponential density already described in Section 7.2. Divergence occurred in a few cases and in one case ($a_{1_0} = 110$, $a_{2_0} = 220$) the following solution defining a stationary point of the log likelihood function was found: $\alpha = -0.331625$, $a_1 = 111.255$, $a_2 = 160.813$. Two different explanations might be given for these somewhat unexpected results. Firstly, the resolution

and/or the size of the used grid might be insufficient. Secondly, it might just not be possible to model the data by means of a real Double Double Exponential probability density. Because of the successful application of the same grid to the Gaussian Double Exponential as well as to the Generalized Laplace Double Exponential density (see below), it is believed that the second explanation is the more likely. For the same reason, it has not been attempted to obtain any solutions of the likelihood equations starting from a finer grid.

The alternative approach referred to in the beginning of this subsection means for the Generalized Laplace Double Exponential probability density, that a grid has to be constructed in the four-dimensional (α, a_1, a_2, b_1)-parameter space. Using ten grid points in each dimension would result in 10000 different initial guesses for the Newton iteration process to be evaluated. In order to limit this very extensive procedure, the same grid as used for the Gaussian Double Exponential density in the (a_1, a_2)-plane (see eq.(177)) was utilized, together with only one pair of initial estimates of α and b_1, namely $\alpha_0 = 1.0$ and $b_{1_0} = 0.8$. Table 32 presents the seven different solutions found of the likelihood equations pertaining to the Generalized Laplace Double Exponential probability density. Figure 55 shows which grid points gave convergence to which solution. A solution is said to have converged when both $M = \max\{|E_1|, |E_2|, |E_3|, |E_4|\} < \epsilon = 1.0*10^{-10}$ and the number of iterations is not larger than twenty-five. In addition, many grid points gave in a few iterations a solution with $\hat{\alpha} = 0.0$, $\hat{a}_1 = 134.284$ or $\hat{\alpha} = 1.0$ and $\hat{a}_2 = 94.8206$ with a whole range of values of $\hat{a}_2$ and $\hat{a}_1$ respectively, without satisfying $M < \epsilon = 1.0*10^{-10}$. In those cases, the iteration process was also stopped. These gridpoints and their corresponding solutions are not shown in Figure 55. This phenomenon did not occur for the Gaussian Double Exponential density discussed earlier. Apparently, the convergence rate was higher there. The second solution is exactly the maximum likelihood estimate of a (single) Double Exponential density. Similarly, the third solution is exactly the maximum likelihood estimate of a (single) Generalized Laplace density. The estimates of the parameters of the core density of the second solution and the estimate of the scale parameter of the tail density of the third solution do not have any practical meaning, because the associated weighting factors are precisely zero. The fourth to the seventh solution inclusive, although satisfying the likelihood equations, do not define any local maximum of the log likelihood function.

Each of these defines a stationary point only. The fourth solution is not acceptable because of its negative weighting factor. Physically, these solutions seem to be close to the third solution. The fifth to the seventh solution inclusive may be regarded as being close to the (single) Double Exponential solution (solution number two). The small values of the shape parameter of the core density of these three solutions define core densities which are close to a homogenous probability density in some region about the mean value of 1000 ft. On the basis of the existence of these three stationary points of the log likelihood function, the existence of even more similar stationary points might be conjectured.

The first solution in Table 32 was analyzed further. The parameter and additional estimates are summarized in Table 31, together with those of the Gaussian Double Exponential density and the Generalized Laplace density discussed before. The core density of the Generalized Laplace Double Exponential density is seen to be very similar to the single Generalized Laplace density. Its tail density has, compared with the Gaussian Double Exponential density, a very small weighting factor, and a relatively large standard deviation. The Generalized Laplace Double Exponential density may thus be considered to be a Generalized Laplace density, which in the core region is slightly perturbed by the Double Exponential tail density, and which in the tail area is dominated by the Double Exponential tail density (see the discussion of the pertinent diagrams below). Compared with the Gaussian Double Exponential density, the maximum likelihood estimates of the overall standard deviation and of the kurtosis are closer to the sample values (compare e.g. Table 29).

Figures 56 to 61 inclusive show the estimation results for the Generalized Laplace Double Exponential density graphically. Compare these figures with Figures 36 to 41 inclusive and Figures 43 to 48 inclusive of the Generalized Laplace and the Gaussian Double Exponential density respectively. On a linear scale, the three densities look very similar. Consequently, this also holds for the linear residual plots (Figures 38, 45 and 58). Logarithmically, they look (very) different in the tail regions below 500 and above 1500 ft. By comparing Figures 37 and 57, the effect of adding a Double Exponential tail density is seen to be very significant. In Figure 57, the concave shape of the logarithm of the core density ranges from about 400 to 1400 ft. The linear shape of the logarithm of the tail density extends below and above approximately 200

and 1800 ft. In the intermediate regions, the shape is convex. Notice
that, compared with the Gaussian Double Exponential density of Figure 44,
the core and the intermediate region extend over a much larger range. This
is due to the small value of the weighting factor of the Double Exponen-
tial tail density in the present case. The main difference between the
logarithmic residual plots (Figures 39, 46 and 59) appears on the two
extreme classes on the left and on the right. The Generalized Laplace
density gives the worst fit on these classes, the Gaussian Double Exponen-
tial density giving the best fit there, whilst the Generalized Laplace
Double Exponential density is in between the two. The overall quality of
the fit as measured by minus twice the standardized log likelihood value
is the best for the Generalized Laplace Double Exponential density (see
Table 31). In fact, it is below the 95 per cent critical value of a
X^2-test with 58 degrees of freedom and would lead to a formal acceptation
of this probability density model. For the practical objectives of this
study, the "1-cumulative" curves are again of prime interest. On a linear
scale, these are very similar (Figures 40, 47 and 60). The so-called
logarithmic "1-cumulative" curves (Figures 41, 48 and 61) show some
important differences in the range of deviations from the mean of 340 feet
and greater. Compared with the Generalized Laplace density, the Genera-
lized Laplace Double Exponential probability density leads to more
cautious estimates of the probability of large deviations from the mean
value. On the other hand, it is less cautious than the Gaussian Double
Exponential density. Notice that the logarithmic "1-cumulative" curve of
the Generalized Laplace Double Exponential density matches the last data
point of the experimental logarithmic "1-cumulative" curve. The corres-
ponding estimate of the probability of vertical overlap would be $P_z(1000)$
$= 4.73*10^{-6}$ (see also Table 31).

Figures 62 to 67 inclusive show the isocontours of the relative
likelihood function derived from the 90 and 95 per cent critical values of
the corresponding likelihood ratio. Small irregularities in the shape of
the contours are due to the limited resolution used in the underlying
computation process and do not have any statistical significance. Five out
of the six diagrams show isocontours which are approximately of an ellip-
tical shape, the exception being the isocontours pertaining to the pair of
parameters α and a_2. Recall that these two parameters are the weighting
factor and the scale factor of the tail density, which have to be esti-
mated from a relatively little part of the total information available in

the data. The large uncertainty on α and a_2 is reflected by the estimated standard deviations $\sigma(\hat{\alpha})$ and $\sigma(\hat{a}_2)$ in Table 31. From a statistical point of view, negative values of the weighting factor α are possible, given the fact that $\sigma(\hat{\alpha})$ is about twice as large as $\hat{\alpha}$. Practically, however, the weighting factor should be greater than or equal to zero, and the contours in Figures 62 to 64 inclusive, have been cut at a lower value of α of zero. Consider now Figure 63, showing isocontours in an (a_2, α)-plane. Notice that the 90 per cent isocontour is closed, whereas the 95 per cent isocontour is not. This is due to the following. When $\alpha=0.0$, the tail density does no longer play a part in the mixture density. As a result, the relative likelihood function does not depend on the scale factor a_2, along the a_2-axis, and is determined completely by the parameters a_1 and b_1 of the core density. Thus the a_2-axis in Figure 63 is itself an isocontour. Its value turns out to be 0.052, corresponding with a confidence value of slightly less than 95 per cent of the associated χ^2-distributed likelihood ratio. Because two isocontours can not intersect each other, the 95 per cent contour bends away along the a_2-axis. Figures 68 to 73 inclusive show the 95 per cent isocontours based directly on the asymptotic maximum likelihood estimator. A lower bound of zero was taken again for α. The shape of the two types of contours is very similar for the pairs of parameters α and b_1, a_1 and a_2, b_1 and a_1. With regard to the dimensions, the contours based on the relative likelihood function are tighter, as they are expected to be. The contours of the relative likelihood function of the two pairs of parameters α and a_1, b_1 and a_2 are approximately elliptical in shape, but they are differently oriented and tighter than their counterparts of Figures 68 and 73 derived directly from the asymptotic maximum likelihood estimator. The main difference between the two types of contours occurs for the pair of parameters α and a_2.

Confidence bounds on the parameters of the Generalized Laplace Double Exponential density model and on the probability of vertical overlap $P_z(1000)$ can, in principle, be defined in the same manner as for the Gaussian Double Exponential probability density. Two points, however, require particular attention. The first one is the observation that the 95 per cent isocontour of the relative likelihood function is not closed. It turns out that its right-hand side is given by $\alpha \to 0$ and $a_2 \to \infty$. The probability of vertical overlap given by eq.(174) may be further approximated by considering only the tail density, i.e.

$$P_z(1000) \approx 2\lambda_z \, f_{GLDE}(0) \approx 2\lambda_z \, \alpha \, \frac{1}{2a_2} \, e^{-\left(\frac{1000}{a_2}\right)}$$

Clearly, the probability of vertical overlap decreases with decreasing α and increases with increasing values of a_2, for $a_2 < 1000$. Whether or not $P_z(1000)$ increases with a_2 along the 95 per cent isocontour for values of a_2 larger than 300, say, depends on the corresponding rate of decay of α along that part of the contour. It turns out that the maximum of $P_z(1000)$ along the 95 per cent isocontour of Figure 63 occurs for approximately $\alpha=0.135$ and $a_2=252$. The second point to be kept in mind when considering confidence bounds for the parameters is that the parameter α is limited to non-negative values. The associated confidence level, therefore, will effectively be lower than 95 per cent.

The values summarized below were obtained by taking again the minimum and maximum values along the contours.

parameter	approximate 95 per cent confidence bounds based on	
	relative likelihood function	maximum likelihood estimator
α	0.0 – 0.22	0.0 – 0.26
a_1	128 – 148	124 – 149
a_2	85 – 245 (90%)	27 – 212
b_1	0.63 – 0.74	0.6 – 0.79
$P_z(1000)$	$1.6*10^{-8} - 42*10^{-6}$	$1.6*10^{-11} - 98*10^{-6}$

Notice that the interval given for a_2 in the middle column is a 90 per cent confidence value rather than a 95 per cent value. The 95 per cent isocontour of Figure 63 would give an infinite interval for the parameter a_2, because the contour is not closed. There is a fairly good agreement between the intervals based on the two approaches for the parameters α, a_1 and b_1. The difference in the case of the scale parameter a_2 of the tail density is obvious from the foregoing discussion. The range for a_2 based on the maximum likelihood estimator is, in fact, also responsible for the difference in the size of the confidence regions of the probability of

vertical overlap, based on the two approaches. The lower bound of $1.6*10^{-11}$ corresponds with the lower bound of 27 of a_2 whereas the upper bound of $98*10^{-6}$ corresponds with the upper bound of 212 of a_2, combined with a relatively high value of the weighting factor α. Taking the theoretical aspects of the derivation of confidence intervals based on the two approaches into account, as well as the particular observations in the case of the Generalized Laplace Double Exponential probability density discussed here, the confidence interval for the probability of vertical overlap $P_z(1000)$ chosen here is the one based on the relative likelihood function, i.e. the interval $1.6*10^{-8} - 42*10^{-6}$. The upper bound is approximately nine times larger than the maximum likelihood point estimate.

7.4 Results based on the Double Generalized Laplace probability distribution

The Double Generalized Laplace probability density is characterized by five parameters α, a_1, a_2, b_1 and b_2, in addition to its mean value (assumed to be 1000 ft here). Rather than using a full grid in the five-dimensional parameter space, the two-dimensional grid in the (a_1, a_2)-plane defined by eq.(177) has been utilized together with one triple of initial estimates of the remaining parameters, namely $\alpha_0 = 0.1$, $b_{1_0} = 0.8$ and $b_{2_0} = 1.0$. Table 32 presents the six solutions of the likelihood equations pertaining to the family of Double Generalized Laplace probability densities, found by this approach. Some of these solutions had not yet fully converged to the two stop criteria $\max\{|E_1|, |E_2|, |E_3|, |E_4|, E_5\} < \varepsilon = 1.0*10^{-10}$ and the number of iterations less than or equal to 25. For this reason, a diagram showing the convergence of the grid points to each solution is omitted. The second solution in Table 33 defines again the maximum likelihood estimate of a (single) Double Exponential density, whereas the third defines the estimate of a (single) Generalized Laplace density (compare Table 28). The fourth solution defines a stationary point of the log likelihood function rather than a maximum. Physically, the fourth solution may be interpreted as a slightly perturbed version of the third solution, that is, the (single) Generalized Laplace. The value of the shape parameter b_2 indicates that, what is called the tail density, is in fact a more or less homogenous density (with standard deviation $\sigma \approx 278.7$ ft) on the interval of 500 to 1500 ft. The fifth and sixth solution in Table 33 also define stationary points rather than maxima of the log likelihood function. Notice that these two solutions are more or less complementary.

The first solution in Table 33 was analyzed further. The parameter and additional estimates are summarized in Table 31, together with those of some of the densities discussed before. It is seen that what is called the tail density, i.e. the density weighted by the factor α, is close to the (single) Generalized Laplace density with regard to its parameter estimates $\hat{a}_2$=135.873 and $\hat{b}_2$=0.755964. Taking the value of α into account, the tail density appears in fact to be the main component of the Double Generalized Laplace mixture density. The core density has a shape parameter of approximately 0.3, which means that the core density is relatively flat over the range of 850 to 1150 ft. Taking the estimated values of the core and tail standard deviations into account, the population may be said to be composed of a small proportion (17%) of relative accurate objects and a major proportion (83 per cent) of average objects. Notice finally that the maximum likelihood estimates of the overall standard deviation and of the kurtosis are close to the sample values (see Table 29).

Figures 74 to 79 inclusive show the estimation results for the Double Generalized Laplace density graphically. The main difference between the Double Generalized Laplace density on the one hand, and the Generalized Laplace, the Gaussian Double Exponential, and the Generalized Laplace Double Exponential on the other, appears in the tail area and, therefore, is most clearly visible in the diagrams showing the logarithms of the densities, the residuals and the "1-cumulative". Compared with a (single) Generalized Laplace density, the tail of the logarithm of the Double Generalized Laplace density has slightly shifted upwards. The differences with respect to the Gaussian Double Exponential and the Generalized Laplace Double Exponential are similar to those between the latter two densities and a (single) Generalized Laplace. Consider the log "1-cumulative" curves of Figures 41, 48, 61 and 79. For deviations from the mean up to 470 feet, the curve based on the Double Generalized Laplace density most closely follows the experimental curve. This is, of course, the result of the flexibility inherent to the five parameter Double Generalized Laplace probability density model. Due to this same flexibility, the former curve follows the experimental curve between 470 and 620 feet more closely than the curve based on the (single) Generalized Laplace density. Compared with the Generalized Laplace Double Exponential density, the Double Generalized Laplace density is more cautious in the range of deviations from the mean between 300 and approximately 500 feet. In the

extreme tail region, however, the Double Exponential tail density of the
Generalized Laplace Double Exponential density becomes more cautious. For
deviations from the mean up to 400 feet, the log "1-cumulative" curves
based on the Gaussian Double Exponential and the Double Generalized
Laplace densities nearly coincide. For the larger deviations, the heavily
weighted tail density of the Gaussian Double Exponential density is more
cautious. Based on the Double Generalized Laplace probability density
model, the estimate of the probability of vertical overlap would be
$P_z(1000) = 2.21*10^{-7}$.

7.5 Summary

A total of six different probability density models has been fitted
to the real aircraft pairs data by means of the maximum likelihood method
in the subsections 7.2 to 7.4 inclusive. Three of these six densities are
single densities, namely a Gaussian, a Double Exponential and a Generali-
zed Laplace density. The Gaussian and the Double Exponential density are
shown not be able to describe the data in an appropriate manner. Although
the Generalized Laplace density gives a satisfactory description of the
data, it is considered to be unsuitable for the extrapolation required to
estimate the probability of vertical overlap.

The remaining three densities are mixture densities, namely the
Gaussian Double Exponential, the Generalized Laplace Double Exponential
and the Double Generalized Laplace density. Each of the three gives a
satisfactory description of the data. As the differences between the fits
of these densities to the core of the data sample are very small, the
choice of a unique probability density mdoel from these three alternatives
depends on how well these densities extrapolate the tails of the data
sample. Because of the inherent uncertainty of the extrapolation process,
a cautious approach has to be taken. The best model then is the Gaussian
Double Exponential probability density, giving a point estimate of the
probability of vertical overlap of $P_z(1000) = 6.59*10^{-6}$ and a 95 per cent
interval estimate of $2.3*10^{-6} - 14.8*10^{-6}$.

8 CONCLUSIONS

The assessment of the risk of collision between aircraft due to the loss of vertical separation in a possible 1000 ft environment makes use of a particular collision risk model. The main parameter of this model is the probability of vertical overlap. It may be estimated from data on the vertical distance between aircraft in a pair.

Due to the limited amount of data that can be obtained in practice, a mathematical probability distribution model of the data is required. A methodology to arrive at such a model has been defined and elaborated. It consists of three elements. The first two elements are the selection of a suitable family of probability distributions and of a method to estimate the unknown parameters from the available data. The family of the Double Generalized Laplace distributions and the maximum likelihood parameter estimation technique have been selected. The third element consists of a careful appraisal of the numerical results obtained, given the uncertainties inherent to the overall process.

The methodology has been implemented in a computer package called Double Generalized Laplace Distribution Fitting, DGLDiF. The package has been written in FORTRAN77, which safeguards portability. The package may be applied in various fields, because of the generality of the family of Double Generalized Laplace probability distributions. It is further strengthened by the possibility of selecting any subset of the five parameters to be estimated, using either grouped data or non-grouped data. Various additional statistical quantities of interest are part of the standard output of the package. The package is presently being extended to mixtures of more than two Generalized Laplace distributions and to other parameter estimation techniques.

The package has been used on the Cyber 180-855 computer of the National Aerospace Laboratory NLR. It has been validated, to the extent possible, by means of simulated data. Some of the technical questions associated with the underlying methodology have been investigated through digital simulation, while some others have been investigated analytically.

Applied to the so-called grouped aircraft pairs data, the
methodology has resulted in the Gaussian Double Exponential probability
density as the best model with regard to the objectives posed. The
associated point estimate of the probability of vertical overlap amounts
$6.6*10^{-6}$ with a 95 per cent confidence interval of $2.3*10^{-6} - 14.8*10^{-6}$.

9   ACKNOWLEDGMENT

It is a pleasure to thank all the people who contributed, in one way
or another, to the work described in this report. Firstly, the many
colleagues of the Vertical Separation Studies Group (VSSG) of the
EUROCONTROL NAVSEP-Panel should be acknowledged. In fact, the conclusions
arrived at in this report are to be considered as conclusions of the VSSG
as a whole. Secondly, some colleagues at the National Aerospace Laboratory
NLR should be acknowledged: Mr. J.M. ten Have of the Flight Division for
managing this work as a part of the overall European Vertical Data Collec-
tion project at the NLR and Mrs. S. Mimoun of the Informatics Division for
designing and developing all the software used during the study.

10   REFERENCES

1.   Report of the fourth meeting of the Review of the
                            General Concept of Separation Panel, ICAO Doc. 9352,
                            RGCSP/4 (1980), ICAO, Montreal, Canada, 1980.

2.   Report of the fifth meeting of the Review of the General
                            Concept of Separation Panel, ICAO Doc. 9465, RGCSP/5
                            (1985), ICAO, Montreal, Canada, 1985.

3. R.N. Croxford   A European programme of work for 1987 associated
                            with the world-wide study of heightkeeping accuracy
                            at and above FL290, Review of the General Concept of
                            Separation Panel, Working Group B (Vertical), Second
                            Meeting, WP/54, ICAO, Montreal, Canada, January
                            1987.

4. P.G. Reich   Analysis of long range air traffic systems –
separation standards, Journal of the Institute of
Navigation, Vol. 19, No. 1, 2, and 3, 1966.

5. G. Moek and   Progress on collision risk modelling within the
    W.P. Cannell   Vertical Separation Studies Group (VSSG),
VSSG/DP/132, EUROCONTROL, Brussels, December 1985.

6. G. Moek   Estimation of collision risk due to loss of vertical
separation using paired aircraft data,
National Aerospace Laboratory NLR, Amsterdam,
The Netherlands, NLR TR 88... , 1988.

7. G. Moek and   Determination of the safety of a North Atlantic
    C.R. Traas   Organized Track System with reduced lateral
separation, AGARD Conf.Proc.No.273,
Air Traffic management – civil/military systems and
technologies, Copenhagen, Denmark, 9–12 October 1979.

8. J.M. ten Have   The processing and analysis of the "European
Vertical Main Data Collection", Volume I: Text,
Volume II: Tables and Figures,
NLR TR 87056 C, National Aerospace Laboratory NLR,
Amsterdam, The Netherlands, 1987.

9.   Vertical separation studies – progress report,
VSSG/DP/35, EUROCONTROL, Brussels, 1983.

10. G.E.P. Box and   Time series analysis, Holden Day, 1976.
    G.M. Jenkins

11. G. Moek   Estimating the probability of vertical overlap
$(P_z(1000))$ based on the results of the European
Vertical Radar Data Collection, Review of the
General Concept of Separation Panel, Working Group B
(Vertical), Second Meeting, WP/18, ICAO, Montreal,
Canada, January 1987.

12. M. Abramowitz and Handbook of mathematical functions, National Bureau
    I.A. Stegun   of Standards, 1964.

13. R.A. Redner and   Mixture densities, maximum likelihood and the EM
    H.F. Walker   algorithm, SIAM Review, Vol.26, No.2, April 1984,
                        pp. 195-239.

14. J.W. Bradley   Preliminary results of fitting distributions to
                        total vertical error and assigned altitude deviation
                        data available from results of U.S. data collection
                        activities, Review of the General Concept of
                        Separation Panel, Working Group B (Vertical), Second
                        Meeting, WP/28, ICAO, Montreal, Canada, January
                        1987.

15. S. Nagaoka   Analyses of the distribution of relative vertical
                        distances observed by NAMS, Fifth meeting of the
                        Review of the General Concept of Separation Panel,
                        WP/111, ICAO, Montreal, Canada, May 1985.

16. S. Nagaoka   Estimation of the probability of vertical overlap of
                        a proximate pair of aircraft based on NAMS data,
                        Review of the General Concept of Separation Panel,
                        Working Group B (Vertical), Second Meeting, WP/55,
                        ICAO, Montreal, Canada, January 1987.

17. S. Nagaoka   An estimation trial of collision risk due to loss of
                        vertical separation, Review of the General Concept
                        of Separation Panel, Working Group B (Vertical),
                        Second Meeting, WP/56, ICAO, Montreal, Canada,
                        January 1987.

18. D. Harrison   Estimating the probability of vertical overlap using
                        aircraft height keeping data obtained from the
                        European Vertical Data Collection, UKCAA DR...,
                        London, 1988.

19. H.W. Sörenson   Parameter estimation, principles and problems,
                        Marcel Dekker, 1980.

20. M.G. Kendall and   The advanced theory of statistics, Griffing, 1967.
    A. Stuart

21. G. Moek   Fitting Generalized Laplace densities to the results
                     of the European Vertical Radar Data Collection by
                     means of maximum likelihood, VSSG/DP/179,
                     EUROCONTROL, Brussels, 1987.

22. S.D. Silvey   Statistical inference, Chapman and Hall, 1975.

23. J.F. Lawless   Statistical models and methods for lifetime data,
                     John Wiley, 1982.

24. P.G. Bhattacharjee Algorithm AS32 - The incomplete Gamma integral,
                     Appl. Statistics, Vol. 19, 1970, pp. 285-287.

25. R.J. Moore   Algorithm AS187 - Derivatives of the incomplete
                     Gamma integral, Appl. Statistics, Vol. 31, No. 3,
                     1982, pp. 330-335.

26. C. Lau   Algorithm AS147 - A simple series for the incomplete
                     Gamma integral, Appl. Statistics, Vol. 29, No. 1,
                     1980, pp. 113-114.

27. F.T. Lindstrom   An algorithm for the evaluation of the incomplete
                     Gamma function and the first two partial derivatives
                     with respect to the parameter, Commun. Statist. -
                     Simula. Computa., B10(5), 1981, pp.465-478.

28. W. Gautschi   A computational procedure for incomplete Gamma
                     functions, ACM transactions on mathematical
                     software, Vol. 5, No. 1, December 1979, pp. 466-481.

29. W. Gautschi   Algorithm 542 - Incomplete Gamma functions (S14),
                     ACM transactions on mathematical software, Vol. 5,
                     No. 4, December 1979, pp. 482-489.

30. C.G. van der Laan Calculation of special functions: the Gamma
    and N.M. Temme   function, the exponential integrals and error-like
                     functions, CWI Tract 10, Centre for Mathematics and
                     Computer Science, Amsterdam, the Netherlands.

31. J.S. Vandergraft Introduction to numerical computations, Academic
                     Press, 1983.

APPENDIX A

DEFINITION OF SOME SAMPLE STATISTICS

The sample mean m and the sample moments s^2, $(m_3)^3$, $(m_4)^4$ centred around the sample mean are defined as follows:

$$m = \frac{1}{n} \sum_{i=1}^{n} d_i$$

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (d_i - m)^2$$

$$(m_3)^3 = \frac{1}{n-1} \sum_{i=1}^{n} (d_i - m)^3$$

$$(m_4)^4 = \frac{1}{n-1} \sum_{i=1}^{n} (d_i - m)^4$$

The sample skewness β_1 is defined as

$$\beta_1 = \frac{\{(m_3)^3\}^2}{(s^2)^3}$$

The sample kurtosis β_2 is defined as

$$\beta_2 = \frac{(m_4)^4}{(s^2)^2}$$

APPENDIX B

REVIEW OF PROBABILITY DENSITY MODELS BEING USED
IN THE US AND JAPANESE VERTICAL DATA COLLECTION STUDIES

A total of nine different families of probability densities is described in Ref. 14 for the United States data collection study. The most general family considered is the "Mixture of Two Power-Exponential Probability Density Functions". It can easily be verified that the Power-Exponential density $f_E(d;\phi,\beta)$ of Ref. 14 is identical to the Generalized Laplace density forming the elements of the Double Generalized Laplace density as defined in this paper. The relation between the parameters a and b on the one hand and ϕ,β on the other, is:

$$b = \frac{1 + \beta}{2} \tag{B.1}$$

$$a = \phi \, 2^{(1 + \beta)/2} \tag{B.2}$$

The "Mixture of Two Power-Exponential Probability Density Functions" is then identical to the Double Generalized Laplace density. The other eight families of Ref. 14 are special cases of the most general family, obtained by fixing the values of one or more of the five parameters. It should finally be noted that in the U.S. case the probability densities are used to model single aircraft data (versus paired aircraft data in the European case).

In the Japanese data collection, the so-called "relative vertical distance" constitutes the data (Refs. 15 - 17). For two aircraft assigned to the same flight level, the relative vertical distance is defined as the difference of their heights as measured by the radar. This is similar to the paired aircraft data used in the European case, but no correction needs to be made for the difference in the height of the assigned flight levels (provided that their temporal and spatial height variations are negligible). Each data point z_j is modelled by

$$z_j = e_{h,2} - e_{h,1} + e_{m,2} - e_{m,1} \tag{B.3}$$

where $e_{h,i}$ is the height keeping error of aircraft i (i = 1,2) in pair j
and $e_{m,i}$ is its measurement error (due to radar as well as changes in
atmospheric conditions). Each term in the right-hand side of eq.(B.3)
above, is assumed to be statistically independent of all the other terms.

Two approaches are distinguished viz. i) postulating separate
probability distributions for the height keeping and measurement errors,
followed by convolution to obtain the probability distribution of the
relative vertical data, and ii) postulating directly a probability
distribution for the relative vertical data. For the latter approach,
three families of probability distributions are considered in Ref. 15,
namely Gaussian, Double Exponential and Gaussian Double Exponential. Only
the last family, however, is elaborated in Ref. 16 and 17. For the former
approach, two cases are considered. Firstly, the case of Gaussian
measurement errors and Double Exponential height keeping errors is given
(Refs. 15 - 17). Secondly, the case of Double Double Exponential height
keeping errors with identically zero measurement errors is described
(Refs. 16 and 17). This corresponds with the Convoluted Double Double
Exponential case discussed for the European data collection in Ref. 11.
The combination of Gaussian measurement errors and Double Exponential
height keeping errors is not included in either the European or the United
States' approach.

APPENDIX C

SUMMARY OF FIRST AND SECOND PARTIAL DERIVATIVES OF
THE DGL PROBABILITY DENSITY TO THE PARAMETERS

C.1) $\displaystyle \frac{\partial}{\partial \alpha} f(d_i) = -\frac{1}{2a_1 b_1 \Gamma(b_1)}\ e^{-\left|\frac{d_i - \mu}{a_1}\right|^{1/b_1}} + \frac{1}{2a_2 b_2 \Gamma(b_2)}\ e^{-\left|\frac{d_i - \mu}{a_2}\right|^{1/b_2}}$

C.2) $\displaystyle \frac{\partial}{\partial a_1} f(d_i) = (1-\alpha)\left\{-1 + \frac{1}{b_1}\left|\frac{d_i - \mu}{a_1}\right|^{1/b_1}\right\}\ \frac{1}{2a_1^2 b_1 \Gamma(b_1)}\ e^{-\left|\frac{d_i - \mu}{a_1}\right|^{1/b_1}}$

C.3) $\displaystyle \frac{\partial}{\partial a_2} f(d_i) = \alpha\left\{-1 + \frac{1}{b_2}\left|\frac{d_i - \mu}{a_2}\right|^{1/b_2}\right\}\ \frac{1}{2a_2^2 b_2 \Gamma(b_2)}\ e^{-\left|\frac{d_i - \mu}{a_2}\right|^{1/b_2}}$

C.4) $\displaystyle \frac{\partial}{\partial b_1} f(d_i) = (1-\alpha)\left\{-\frac{1}{b_1} - \Psi(b_1) + \frac{1}{b_1^2}\left|\frac{d_i - \mu}{a_1}\right|^{1/b_1}\ \log\left|\frac{d_i - \mu}{a_1}\right|\right\}\ *$

$\displaystyle \frac{1}{2a_1 b_1 \Gamma(b_1)}\ e^{-\left|\frac{d_i - \mu}{a_1}\right|^{1/b_1}}$

C.5) $\displaystyle \frac{\partial}{\partial b_2} f(d_i) = \alpha\left\{-\frac{1}{b_2} - \Psi(b_2) + \frac{1}{b_2^2}\left|\frac{d_i - \mu}{a_2}\right|^{1/b_2}\ \log\left|\frac{d_i - \mu}{a_2}\right|\right\}\ *$

$\displaystyle \frac{1}{2a_2 b_2 \Gamma(b_2)}\ e^{-\left|\frac{d_i - \mu}{a_2}\right|^{1/b_2}}$

C.6) $\displaystyle \frac{\partial^2}{\partial \alpha^2} f(d_i) = 0$

C.7) $\displaystyle \frac{\partial^2}{\partial \alpha \partial a_1} f(d_i) = \left(1 - \frac{1}{b_1}\left|\frac{d_i - \mu}{a_1}\right|^{1/b_1}\right)\ \frac{1}{2a_1^2 b_1 \Gamma(b_1)}\ e^{-\left|\frac{d_i - \mu}{a_1}\right|^{1/b_1}}$

C.8) $\dfrac{\partial^2}{\partial\alpha\partial a_2} f(d_i) = \left(-1 + \dfrac{1}{b_2} \left|\dfrac{d_i-\mu}{a_2}\right|^{1/b_2}\right) \dfrac{1}{2a_2^2 b_2 \Gamma(b_2)}\ e^{-\left|\frac{d_i-\mu}{a_2}\right|^{1/b_2}}$

C.9) $\dfrac{\partial^2}{\partial\alpha\partial b_1} f(d_i) = \left(\dfrac{1}{b_1} + \Psi(b_1) - \dfrac{1}{b_1^2} \left|\dfrac{d_i-\mu}{a_1}\right|^{1/b_1}\ \log\left|\dfrac{d_i-\mu}{a_1}\right|\right) *$

$$\dfrac{1}{2a_1 b_1 \Gamma(b_1)}\ e^{-\left|\frac{d_i-\mu}{a_1}\right|^{1/b_1}}$$

C.10) $\dfrac{\partial^2}{\partial\alpha\partial b_2} f(d_i) = \left(\dfrac{1}{b_2} + \Psi(b_2) - \dfrac{1}{b_2^2} \left|\dfrac{d_i-\mu}{a_2}\right|^{1/b_2}\ \log\left|\dfrac{d_i-\mu}{a_2}\right|\right) *$

$$\dfrac{1}{2a_2 b_2 \Gamma(b_2)}\ e^{-\left|\frac{d_i-\mu}{a_2}\right|^{1/b_2}}$$

C.11) $\dfrac{\partial^2}{\partial a_1^2} f(d_i) = (1-\alpha) \dfrac{1}{a_1^3 b_1 \Gamma(b_1)}\ e^{-\left|\frac{d_i-\mu}{a_1}\right|^{1/b_1}} *$

$$\left\{-\dfrac{1}{2b_1^2} \left|\dfrac{d_i-\mu}{a_1}\right|^{1/b_1} + \left(-1 + \dfrac{1}{b_1} \left|\dfrac{d_i-\mu}{a_1}\right|^{1/b_1}\right) \left(-1 + \dfrac{1}{2b_1} \left|\dfrac{d_i-\mu}{a_1}\right|^{1/b_1}\right)\right\}$$

C.12) $\dfrac{\partial^2}{\partial a_1 \partial a_2} f(d_i) = 0$

C.13) $\dfrac{\partial^2}{\partial a_1 \partial b_1} f(d_i) = (1-\alpha) \dfrac{1}{2a_1^2 b_1 \Gamma(b_1)}\ e^{-\left|\frac{d_i-\mu}{a_1}\right|^{1/b_1}} *$

$$\left\{-\dfrac{1}{b_1^2} \left|\dfrac{d_i-\mu}{a_1}\right|^{1/b_1} \left(1 + \dfrac{1}{b_1} \log\left|\dfrac{d_i-\mu}{a_1}\right|\right)\ +$$

$$+ \left(-1 + \dfrac{1}{b_1} \left|\dfrac{d_i-\mu}{a_1}\right|^{1/b_1}\right) \left(-\dfrac{1}{b_1} - \Psi(b_1) + \dfrac{1}{b_1^2} \left|\dfrac{d_i}{a_1}\right|^{\mu\ 1/b} \log\left|\dfrac{d_i}{a_1}\right|^{\mu}\right)\right\}$$

C.14) $\dfrac{\partial^2}{\partial a_1 \partial b_2} f(d_i) = 0$

C.15) $\dfrac{\partial^2}{\partial a_2^2} f(d_i) = \alpha \ \dfrac{1}{a_2^3 b_2 \Gamma(b_2)} \ e^{-\left|\frac{d_i-\mu}{a_2}\right|^{1/b_2}} \ *$

$$\left\{- \frac{1}{2b_2^2} \left|\frac{d_i-\mu}{a_2}\right|^{1/b_2} + \left(- 1 + \frac{1}{b_2} \left|\frac{d_i-\mu}{a_2}\right|^{1/b_2}\right)\left(- 1 + \frac{1}{2b_2} \left|\frac{d_i-\mu}{a_2}\right|^{1/b_2}\right)\right\}$$

C.16) $\dfrac{\partial^2}{\partial a_2 \partial b_1} f(d_i) = 0$

C.17) $\dfrac{\partial^2}{\partial a_2 \partial b_2} f(d_i) = \alpha \ \dfrac{1}{2a_2^2 b_2 \Gamma(b_2)} \ e^{-\left|\frac{d_i-\mu}{a_2}\right|^{1/b_2}} \ *$

$$\left\{- \frac{1}{b_2^2} \left|\frac{d_i-\mu}{a_2}\right|^{1/b_2} \left(1 + \frac{1}{b_2} \log\left|\frac{d_i-\mu}{a_2}\right|\right) \ + \right.$$

$$\left. + \left(- 1 + \frac{1}{b_2} \left|\frac{d_i-\mu}{a_2}\right|^{1/b_2}\right)\left(- \frac{1}{b_2} - \Psi(b_2) + \frac{1}{b_2^2} \left|\frac{d_i-\mu}{a_2}\right|^{1/b_2} \log\left|\frac{d_i-\mu}{a_2}\right|\right)\right\}$$

C.18) $\dfrac{\partial^2}{\partial b_1^2} f(d_i) = (1-\alpha) \ \dfrac{1}{2a_1 b_1 \Gamma(b_1)} \ e^{-\left|\frac{d_i-\mu}{a_1}\right|^{1/b_1}} \ *$

$$\left[\frac{1}{b_1^2} - \Psi'(b_1) + \left\{- \frac{1}{b_1^3} \left|\frac{d_i-\mu}{a_1}\right|^{1/b_1}\left(2 + \frac{1}{b_1} \log\left|\frac{d_i-\mu}{a_1}\right|\right)\right\} \log\left|\frac{d_i-\mu}{a_1}\right| \ + \right.$$

$$\left. + \left(- \frac{1}{b_1} - \Psi(b_1) + \frac{1}{b_1^2} \left|\frac{d_i-\mu}{a_1}\right|^{1/b_1} \log\left|\frac{d_i-\mu}{a_1}\right|\right) ** \ 2\right]$$

C.19) $\dfrac{\partial^2}{\partial b_1 \partial b_2} f(d_i) = 0$

C.20) $\dfrac{\partial^2}{\partial b_2^2} f(d_i) = \alpha \ \dfrac{1}{2a_2 b_2 \Gamma(b_2)} \ e^{-\left|\frac{d_i-\mu}{a_2}\right|^{1/b_2}}$ *

$$\left[\dfrac{1}{b_2^2} - \Psi'(b_2) + \left\{ -\dfrac{1}{b_2^3} \left|\dfrac{d_i-\mu}{a_2}\right|^{1/b_2} \left(2 + \dfrac{1}{b_2} \log\left|\dfrac{d_i-\mu}{a_2}\right|\right) \right\} \log\left|\dfrac{d_i-\mu}{a_2}\right| \ + \right.$$

$$\left. + \left(-\dfrac{1}{b_2} - \Psi(b_2) + \dfrac{1}{b_2^2} \left|\dfrac{d_i-\mu}{a_2}\right|^{1/b_2} \log\left|\dfrac{d_i-\mu}{a_2}\right| \right) ** 2 \right]$$

C.21) $\Psi(b) = \dfrac{1}{\Gamma(b)} \dfrac{d\Gamma(b)}{db}$

C.22) $\Psi(b) = -C - \dfrac{1}{b} + b \sum_{k=1}^{\infty} \dfrac{1}{k(b+k)}$

C.23) $C = 0.57721566490...$, Euler's constant

C.24) $\Psi'(b) = \dfrac{1}{b^2} + \sum_{k=1}^{\infty} \dfrac{1}{(b+k)^2}$

See for C.21 to C.24 inclusive: I.S. Gradshteyn and I.M. Ryzhik, Table of integrals, series, and products, Academic Press, 1980.

interval	# data points
1	1447
2	1174
3	891
4	565
5	368
6	203
7	134
8	78
9	40
10	25
11	14
12	12
13	5
14	0
15	0
16	4

Table 1   Number of data points per class interval of 40 ft
           (Note: intervals counted from 1000 ft onwards, on the assumption
           of symmetry)

interval	½ data points	interval	½ data points
1	766	16	37
2	681	17	25
3	611	18	15
4	563	19	15
5	479	20	10
6	412	21	12
7	340	22	2
8	225	23	7
9	209	24	5
10	159	25	4
11	120	26	1
12	83	27	0
13	75	28	0
14	59	29	0
15	41	30	0
		31	2
		32	2

Table 2   Number of data points per class interval of 20 ft
         (Note: intervals counted from 1000 ft onwards, on the assumption
                of symmetry of the distribution)

interval	# data points	interval	# data points	interval	# data points
1	404	27	25	53	0
2	362	28	34	54	0
3	358	29	18	55	0
4	323	30	23	56	0
5	328	31	21	57	0
6	283	32	16	58	0
7	299	33	11	59	0
8	264	34	14	60	0
9	238	35	8	61	0
10	241	36	7	62	2
11	226	37	8	63	2
12	186	38	7	64	0
13	192	39	5		
14	148	40	5		
15	124	41	3		
16	101	42	9		
17	111	43	2		
18	98	44	0		
19	90	45	3		
20	69	46	4		
21	58	47	2		
22	62	48	3		
23	44	49	3		
24	39	50	1		
25	33	51	1		
26	42	52	0		

Table 3   Number of data points per class interval of 10 ft
(Note: intervals counted from 1000 ft onwards, on the assumption
of symmetry of the distribution)

α \ a	$\neq \hat{\hat{a}} \wedge \neq 2\hat{\hat{a}}$	$\hat{\hat{a}}$	$2\hat{\hat{a}}$
$0<\alpha<1$	① $\Delta a_1 = \Delta a_2 = \Delta$ $\Delta\alpha = 0$	② $\Delta a_1 = \Delta a_2 = 0$ $\Delta\alpha$ undetermined	③ $\Delta a_1 = \Delta a_2$ undetermined $\Delta\alpha$ contradictory
$\alpha=0$	④ $\Delta a_1 = \Delta a_2$ undetermined $\Delta\alpha$ undetermined	⑤ $\Delta a_1 = 0$ $\Delta\alpha$ undetermined Δa_2 undetermined	⑥ $\Delta a_1 = \Delta a_2$ undetermined $\Delta\alpha = 1$
$\alpha=1$	⑦ $\Delta a_1 = \Delta a_2$ undetermined $\Delta\alpha$ undetermined	⑧ Δa_1 undetermined $\Delta\alpha$ undetermined $\Delta a_2 = 0$	⑨ $\Delta a_1 = \Delta a_2$ undetermined $\Delta\alpha = -1$

Table 4 The solution of eqs.(136) to (138) as a function of the values of the weighting factor α and the scale parameter a.

method	algorithm	number of terms	accuracy
Taylor series expansion	Bhattacharjee	variable	absolute, first neglected term $< 10^{-8}$
	Moore	variable	relative, first neglected term $< 10^{-6}$
	Lau	variable	relative, first neglected term $< 10^{-6}$
	Lindstrom	200	nine or more significant digits
	Gautschi	variable	any prescribed number of significant digits
Continued fraction expansion	Bhattacharjee	variable	relative difference of two successive iterations $\leq 10^{-8}$
	Moore	variable	relative difference of two successive iterations $\leq 10^{-6}$
	Gautschi	variable	any prescribed number of significant digits

Table 5   Summary of the characteristics of five algorithms for the computation of the incomplete Gamma function.

interval	# data points
1	1492
2	963
3	735
4	504
5	408
6	253
7	183
8	139
9	84
10	68
11	48
12	48
13	17
14	16
15	17
16	9
17	6
18	1
19	2
20	2
21	1
22	1
23	1
24	0
25	1
26	0
27	1

Table 6   Number of data points per class interval of 10 ft.
Simulation based on DE probability distribution of Section 6.2.
(Note: intervals counted from the mean onwards, on the
assumption of symmetry of the distribution.)

sample size n	class interval w	degrees of freedom d.f.	2* standardized log likelihood based on		$\Sigma_i \; p_i$(approximated)
			p_i(exact) eq.(44)	p_i(approximated) eq.(46)	
5000	5	53	63.03	74.60	0.99872
5000	10	26	33.54	79.80	0.99526
10000	5	53	68.58	91.72	0.99872
10000	10	26	29.07	121.6	0.99526

Table 7 χ^2-test statistic values based on exact and approximated probabilities, for various combinations of sample size and class interval.

type of probability density	degrees of freedom d.f.	2* standardized log likelihood based on		$\Sigma_i \; p_i$(approximated)
		p_i(exact) eq.(44)	p_i(approximated) eq.(46)	
DE	25	31.23	78.81	0.99515
GL	24	35.21	72.84	0.99615
DDE	23	31.23	78.81	0.99515

Table 8 χ^2-test statistic values based on exact and approximated probabilities, for the three probability density models fitted to the simulated data.

interval	# data points	interval	# data points	interval	# data points
1	1499	26	5	51	0
2	1111	27	6	52	0
3	791	28	6	53	0
4	453	29	7	54	1
5	270	30	4	55	0
6	129	31	6	56	0
7	90	32	9	57	1
8	72	33	3	58	0
9	61	34	5	59	0
10	49	35	2	60	0
11	60	36	3	61	0
12	46	37	4	62	1
13	39	38	4	63	0
14	37	39	1	64	0
15	29	40	2	65	0
16	22	41	0	66	0
17	27	42	1	67	0
18	23	43	0	68	0
19	15	44	0	69	1
20	16	45	0	70	
21	21	46	1	71	
22	16	47	3	72	
23	17	48	0	73	
24	14	49	0	74	
25	16	50	1	75	

Table 9   Number of data points per class interval of 10 ft.

Simulation based on GDE probabiliity distribution of Section 6.3.

case	number of parameters estimated	$\hat{\alpha}$	$\hat{a}_1$	$\hat{a}_2$	$\hat{b}_1$	$\hat{b}_2$
G_2	2	0.284507	30.4652			
G_3	3	0.288882	30.3581	88.1285		
G_4	4	0.283686	29.7019	89.0488	0.532338	
G	5	0.269350	29.748910	103.458	0.539197	0.918949

Table 10 Maximum likelihood parameter estimation results for various cases, based on grouped simulated data. Simulated probability density is GDE with parameter values:

α = 0.3 μ = 0.0

a_1 = 30 b_1 = 0.5

a_2 = 90 b_2 = 1.0

case	number of parameters estimated	$\sigma(\hat{\alpha})$	$\sigma(\hat{a}_1)$	$\sigma(\hat{a}_2)$	$\sigma(\hat{b}_1)$	$\sigma(\hat{b}_2)$
G_2	2	0.0101	0.529			
G_3	3	0.0129	0.564	3.34		
G_4	4	0.0137	0.804	3.50	0.0282	
G	5	0.0234	0.802	21.1	0.0298	0.109

Table 11 Estimated standard deviations of maximum likelihood parameter estimates, based on grouped simulated data. Simulated probability density is GDE.

case	degrees of freedom	2* standardized log likelihood based on		$\sum\limits_i p_i$(approximated)
		p_i(exact) eq.(44)	p_i(approximated) eq.(46)	
true GDE	68	67.66	69.24	0.99971
G_2	66	65.65	66.80	0.99979
G_3	65	65.26	66.49	0.99979
G_4	64	63.53	65.73	0.99968
G	63	62.49	64.26	0.99974

Table 12 X^2-test statistic values based on exact and approximated probabilities, for the true and estimated probability density models.

case	number of parameters estimated	$\hat{\alpha}$	$\hat{a}_1$	$\hat{a}_2$	$\hat{b}_1$	$\hat{b}_2$
S_2	2	0.284518	30.2682			
S_3	3	0.289635	30.1420	87.8291		
S_4	4	0.279873	28.8746	89.5754	0.561488	
S	5	0.265341	28.8822	103.905	0.570186	0.920811

Table 13 Maximum likelihood parameter estimation results for various cases, based on non-grouped simulated data. The simulated probability density is the GDE density with parameter values:

α = 0.3

a_1 = 30 b_1 = 0.5

a_2 = 90 b_2 = 1.0

case	number of parameters estimated	$\sigma(\hat{\alpha})$	$\sigma(\hat{a}_1)$	$\sigma(\hat{a}_2)$	$\sigma(\hat{b}_1)$	$\sigma(\hat{b}_2)$
S_2	2	0.0101	0.527			
S_3	3	0.0130	0.565	3.34		
S_4	4	0.0139	0.878	3.59	0.0326	
S	5	0.0284	0.877	26.7	0.0362	0.137

Table 14 Estimated standard deviations of maximum likelihood parameter estimates, based on non-grouped simulated data. Simulated probability density is GDE.

interval	# data points	interval	# data points	interval	# data points
1	1200	26	6	51	0
2	849	27	6	52	0
3	603	28	6	53	0
4	497	29	7	54	1
5	358	30	4	55	0
6	263	31	5	56	0
7	231	32	10	57	1
8	194	33	3	58	0
9	122	34	4	59	0
10	98	35	3	60	0
11	81	36	3	61	0
12	62	37	4	62	1
13	71	38	4	63	0
14	48	39	1	64	0
15	33	40	2	65	0
16	38	41	0	66	0
17	18	42	1	67	0
18	35	43	0	68	0
19	18	44	0	69	1
20	15	45	0		
21	24	46	1		
22	17	47	3		
23	14	48	0		
24	18	49	0		
25	15	50	1		

Table 15   Number of data points per class interval of 10 ft.
Simulation based on DDE probability distribution of Section 6.4.

case	number of parameters estimated	$\hat{\alpha}$	$\hat{a}_1$	$\hat{a}_2$	$\hat{b}_1$	$\hat{b}_2$
G_2	2	0.273494	30.6685			
G_3	3	0.286538	30.3169	87.8419		
G_4	4	0.351737	32.3065	82.2403	0.883416	
G	5	0.423585	32.0985	60.6733	0.867734	1.13398

Table 16 Maximum likelihood parameter estimation results for various cases, based on grouped simulated data. Simulated probability density is DDE with parameter values:

$$\alpha = 0.3 \quad \mu = 0.0$$
$$a_1 = 30 \quad b_1 = 1.0$$
$$a_2 = 90 \quad b_2 = 1.0$$

case	number of parameters estimated	$\sigma(\hat{\alpha})$	$\sigma(\hat{a}_1)$	$\sigma(\hat{a}_2)$	$\sigma(\hat{b}_1)$	$\sigma(\hat{b}_2)$
G_2	2	0.0189	0.983			
G_3	3	0.0415	1.40	5.85		
G_4	4	0.0507	1.70	5.33	0.0548	
G	5	0.101	1.81	19.8	0.0628	0.137

Table 17 Estimated standard deviations of maximum likelihood parameter estimates, based on grouped simulated data. Simulated probability density is DDE.

case	number of parameters estimated	estimated parameter				
		$\hat{\alpha}$	$\hat{a}_1$	$\hat{a}_2$	$\hat{b}_1$	$\hat{b}_2$
G_2'	2	1.9	1.9			
G_3	3	3.2	2.5	1.8		
G_4	4	3.7	2.1	1.5	1.9	
G	5	4.7	2.3	0.94	2.1	0.7

Table 18 Ratio between $\sigma(parameter)_{DDE}$ and $\sigma(parameter)_{GDE}$

| case | degrees of freedom d.f. | 2* standardized log likelihood based on | | $\sum_i p_i$(approx.) | $\chi^2_{.95}$,(d.f.) |
		p_i(exact) eq.(44)	p_i(approximated) eq.(46)		
DE	67	353.76	372.78	0.99810	87.1
GL	66	104.08	185.79	0.99190	86.9
true DDE	68	80.65	114.75	0.99648	88.2
G_2	66	79.07	112.65	0.99652	86.9
G_3	65	78.64	112.52	0.99651	84.8
G_4	64	83.18	104.50	0.99775	83.7
G	63	83.09	105.88	0.99757	82.5

Table 19 χ^2-test statistic values based on exact and approximated
probabilities, for the true and estimated probability density models.

case	number of parameters estimated	$\hat{\alpha}$	$\hat{a}_1$	$\hat{a}_2$	$\hat{b}_1$	$\hat{b}_2$
S_2	2	0.283337	29.7288			
S_3	3	0.320231	28.7000	84.4456		
S_4	4	0.0	28.1394	32.8299	1.02791	
S	5	0.139939	24.6116	110.678	1.26653	0.963159

Table 20 Maximum likelihood parameter estimation results for various cases, based on non-grouped simulated data. Simulated probability density is DDE with parameter values:

$\alpha = 0.3$ $\mu = 0.0$
$a_1 = 30.0$ $b_1 = 1.0$
$a_2 = 90.0$ $b_2 = 1.0$

case	number of parameters estimated	$\sigma(\hat{\alpha})$	$\sigma(\hat{a}_1)$	$\sigma(\hat{a}_2)$	$\sigma(\hat{b}_1)$	$\sigma(\hat{b}_2)$
S_2	2	0.0189	0.982			
S_3	3	0.0455	1.48	5.43		
S_4	4	-	-	-	-	
S	5	-	-	-	-	-

Table 21 Estimated standard deviations of maximum likelihood parameter estimates, based on non-grouped simulated data. Simulated probability density is DDE.

interval	# data points	interval	# data points
1	1449	26	8
2	1037	27	2
3	755	28	1
4	494	29	0
5	351	30	0
6	206	31	2
7	131	32	2
8	97	33	1
9	84	34	0
10	58	35	1
11	49	36	0
12	44	37	1
13	33	38	0
14	32	39	0
15	25	40	1
16	29		
17	26		
18	22		
19	9		
20	12		
21	7		
22	11		
23	9		
24	7		
25	4		

Table 22 Number of data points per class interval of 10 ft.

Simulation based on DGL probability distribution of Section 6.5

case	number of parameters estimated	$\hat{\alpha}$	$\hat{a}_1$	$\hat{a}_2$	$\hat{b}_1$	$\hat{b}_2$
G_2	2	0.279731	30.5115			
G_3	3	0.290766	30.2155	87.3888		
G_4	4	0.292612	30.3301	87.1795	0.643835	
G	5	0.244176	30.7618	113.568	0.661067	0.660112

Table 23   Maximum likelihood parameter estimation results for various cases, based on grouped simulated data. Simulated probability density is DGL with parameter values:

α = 0.3 μ = 0.0
a_1 = 30.0 b_1 = 0.65
a_2 = 90.0 b_2 = 0.80

case	number of parameters estimated	$\sigma(\hat{\alpha})$	$\sigma(\hat{a}_1)$	$\sigma(\hat{a}_2)$	$\sigma(\hat{b}_1)$	$\sigma(\hat{b}_2)$
G_2	2	0.0129	0.678			
G_3	3	0.0205	0.797	3.54		
G_4	4	0.0230	1.02	3.72	0.0348	
G	5	0.0523	1.11	30.5	0.0401	0.169

Table 24   Estimated standard deviations of the maximum likelihood parameter estimates, based on grouped simulated data. Simulated probability density is DGL.

| case | degrees of freedom d.f. | 2* standardized log likelihood based on | | $\Sigma\ p_i$ (approximated) |
		p_i(exact) eq.(44)	p_i(approximated) eq.(46)	
true DGL	39	45.00	52.38	0.99892
G_2	37	42.78	49.97	0.99896
G_3	36	42.08	49.45	0.99899
G_4	35	42.41	49.28	0.99904
G	34	39.84	47.47	0.99906

Table 25 χ^2-test statistic values based on exact and approximated probabilities, for the true and estimated probability density models.

case	number of parameters estimated	$\hat{\alpha}$	$\hat{a}_1$	$\hat{a}_2$	$\hat{b}_1$	$\hat{b}_2$
S_2	2	0.282860	30.0398			
S_3	3	0.297780	29.6362	86.5270		
S_4	4	0.276779	28.1992	88.8963	0.724012	
S	5	0.191361	28.7524	143.199	0.768766	0.542318

Table 26 Maximum likelihood parameter estimation results for various cases, based on non-grouped simulated data. Simulated probability density is DGL with parameter values:

$\alpha = 0.3$ $\mu = 0.0$
$a_1 = 30.0$ $b_1 = 0.65$
$a_2 = 90.0$ $b_2 = 0.80$

case	number of parameters estimated	$\sigma(\hat{\alpha})$	$\sigma(\hat{a}_1)$	$\sigma(\hat{a}_2)$	$\sigma(\hat{b}_1)$	$\sigma(\hat{b}_2)$
S_2	2	0.0129	0.674			
S_3	3	0.0206	0.798	3.46		
S_4	4	0.0247	1.18	4.05	0.0454	
S	5	0.0395	1.17	28.9	0.0530	0.114

Table 27 Estimated standard deviations of maximum likelihood parameter estimates. Simulated probability density is DGL.

density	parameters				quality of fit		
	$\hat{a}$	$\sigma(\hat{a})$	$\hat{b}$	$\sigma(\hat{b})$	d.f.	2 * 1) standard. log like- lihood	$\chi^2_{.95,d.f.}$
Gaussian	177.152	1.78	–	–	61	257.80	80.2
Double Exponential	94.8206	1.35	–	–	61	211.18	80.2
Generalized Laplace	134.284	3.72	0.727811	0.0203	60	77.17	79.08

Table 28   Maximum likelihood parameter estimation results for Gaussian, Double Exponential and Generalized Laplace probability distribution and grouped pairs data of Table 3

(Note 1: based on exact probabilities p_i)

density	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\beta}_1$	$\hat{\beta}_2$	$P_z(1000)$
observed	997.1	125.2	-0.00015	4.33	–
Gaussian	1000	125.265	0	3	$3.78*10^{-15}$
Double Exponential	1000	134.097	0	6	$1.13*10^{-5}$
Generalized Laplace	1000	125.205	0	4.09	$4.68*10^{-8}$

Table 29   Estimates of the characteristics of the observed and of the three estimated probability distributions

solution number	$\hat{\alpha}$	$\hat{a}_1$	$\hat{a}_2$
1	0.429494	163.990	97.9211
2	1.00000	97.9558	94.8206
3	1.00000	283.918	94.8206
4	0.00000	177.152	4.08656
5	1.02566	7.86915	92.5808
6	1.29950	50.4167	79.5968

Table 30 Some solutions of the likelihood equations
of the Gaussian Double Exponential probability
density model, using the grouped data of
Table 3 with a class interval of 10 ft.

probability density model / estimation result	Gaussian Double Exponential (GDE)	Generalized Laplace Double Exponential (GLDE)	Generalized Laplace (GL)	Double Generalized Laplace (DGL)
α	0.429494	0.0495614	-	0.827724
a_1	163.990	136.717	134.284	151.358
a_2	97.9211	122.526	-	135.873
b_1	-	0.697953	0.727811	0.271231
b_2	-	-	-	0.755964
$\sigma(\alpha)$	0.0487	0.0844	-	0.0815
$\sigma(a_1)$	5.37	4.90	3.72	14.6
$\sigma(a_2)$	4.52	35.2	-	7.53
$\sigma(b_1)$	-	0.0391	0.0203	0.130
$\sigma(b_2)$	-	-	-	0.00362
d.f.	59	58	60	57
$2*(\ell_2-\ell_1)$ [1]	78.25	74.99	77.17	66.95
$\chi^2_{.95,(d.f.)}$	77.9	76.8	79.08	75.7
$P_z(1000)$	$6.59*10^{-6}$	$4.73*10^{-6}$	$4.68*10^{-8}$	$2.21*10^{-7}$
μ	1000	1000	1000	1000
σ	126.126	125.428	125.205	125.423
σ_1	115.958	122.421	-	88.9614
σ_2	138.481	173.278	-	131.749
β_1	0.0	0.0	0.0	0.0
β_2	4.97	4.47	4.09	4.39

Table 31 Summary of maximum likelihood parameter estimation results for four different probability density models and the grouped data of Table 3 (Note 1: based on exact probabilities p_i, $i=1,2,\ldots,63$)

solution number	$\hat{\alpha}$	$\hat{a}_1$	$\hat{a}_2$	$\hat{b}_1$
1	0.0495614	136.717	122.526	0.697953
2	1.00000	102.541	94.8206	0.0409075
3	0.00000	134.284	95.9099	0.727811
4	-0.0228234	126.426	134.970	0.763878
5	0.833567	146.304	99.0175	0.0418090
6	0.846096	139.130	99.1601	0.0332417
7	0.872184	130.430	98.8452	0.0276456

Table 32   Some solutions of the likelihood equations of the
Generalized Laplace Double Exponential probability
density model, using the grouped data of Table 3 with a
class interval of 10 ft.

solution number	$\hat{\alpha}$	$\hat{a}_1$	$\hat{a}_2$	$\hat{b}_1$	$\hat{b}_2$
1	0.827724	151.358	135.873	0.271231	0.755964
2 *	1.00000	92.1233	134.284	0.654584	0.727811
3 *	0	134.284	162.952	0.727811	0.625939
4	0.00662023	134.836	496.629	0.715010	0.131680
5 *	0.449525	143.601	137.723	0.764259	0.554998
6 *	0.549530	137.693	143.642	0.555473	0.764234

Table 33 Some solutions of the likelihood equations of the
Double Generalized Laplace probability density model,
using the grouped data of Table 3 with a class interval
of 10 ft.
(Note: * not fully converged solution)

FLk

2000 ft   d

FLj

FLk'

1000 ft

FLj'
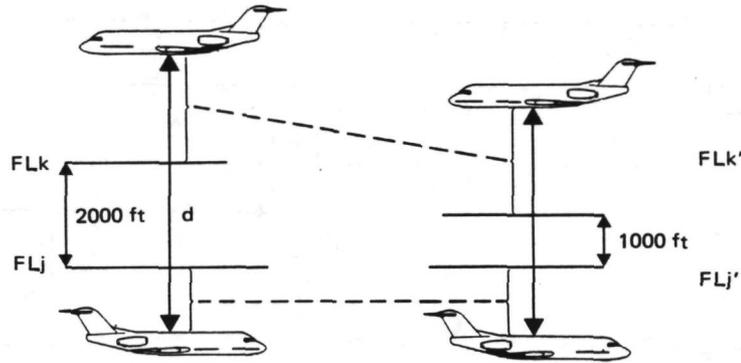
Figure 1   Transforming aircraft height keeping errors from present situation (2000 ft nominal separation) to possible future situation (1000 ft nominal separation)

λ_z

DELTA

DELTA

f (DELTA)

0

CFL** R

msl

WITH h = DELTA:

S_z

f (h)

f (h $-$ S$_z$)

0

h

$$P_z\{S_z\} = \int_{-\infty}^{+\infty} f(h) \left[\int_{h-\lambda_z}^{h+\lambda_z} f(t-S_z)\, dt \right] dh$$

$$P_z\{S_z\} \approx 2\lambda_z \int_{-\infty}^{+\infty} f(h)\, f(h-S_z)\, dh$$
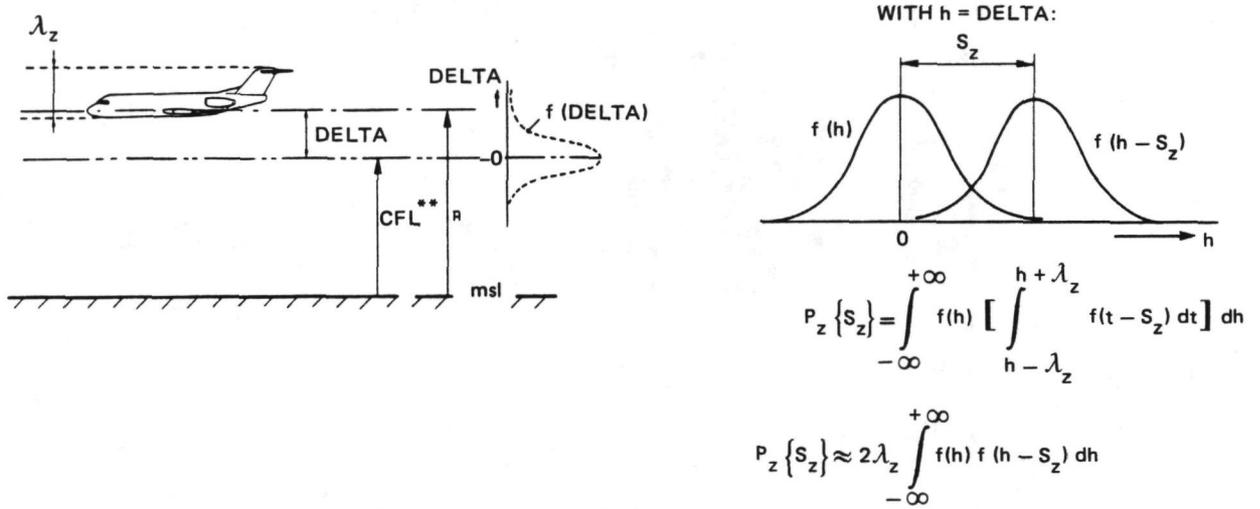
Figure 2   Single Aircraft Approach: determination of the probability of vertical overlap $P_z\{S_z\}$ from single aircraft measurements for a nominal separation S_z

λ_z

DELTA (2)

λ_z

CFL (2)

S_z

H (2)

DELTA (1)

H (1)

CFL (1)

msl

g (NOMSEP)

$-\lambda_z$ 0 $+\lambda_z$ S$_z$ NOMSEP

NOMSEP (S$_z$) = DELTA (2) $-$ DELTA (1) + S$_z$ =

$\left\{ H(2) - H(1) \right\} - \left\{ CFL(2) - CFL(1) \right\} + S_z$

WITH t = NOMSEP:

$$P_z\{S_z\} = \int_{-\lambda_z}^{+\lambda_z} g(t)\, dt$$
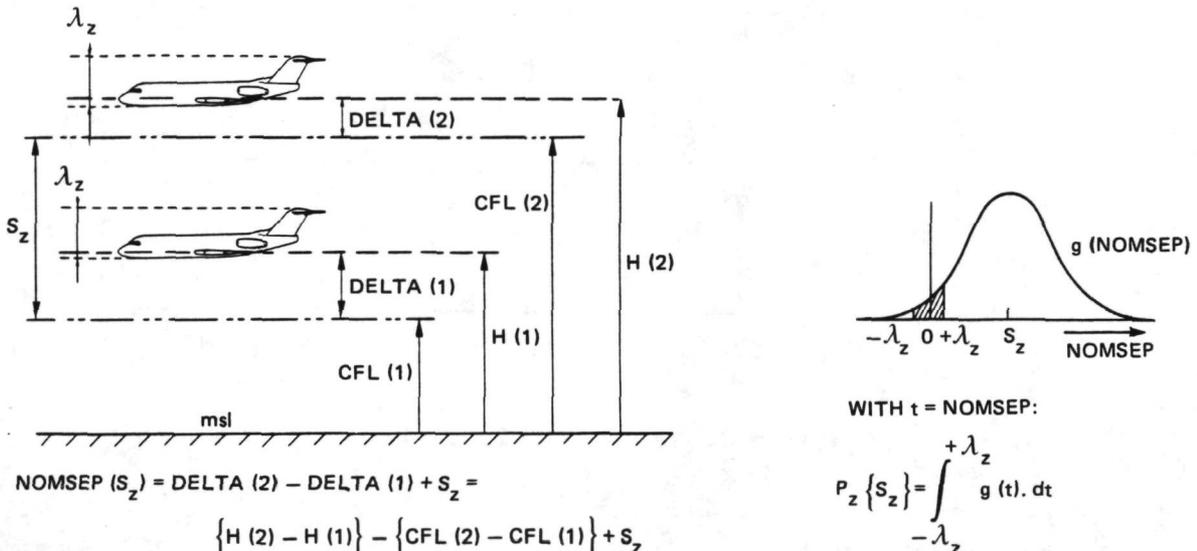
Figure 3   Pairing Approach: determination of the probability of vertical overlap $P_z\{S_z\}$ from paired aircraft measurements for a nominal separation S_z
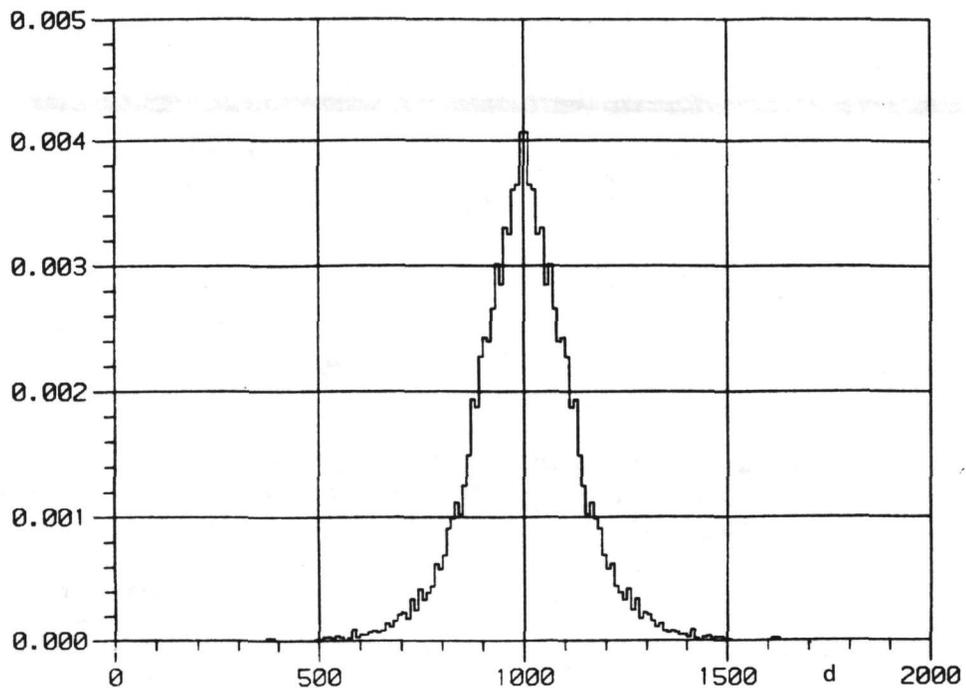
Figure 4 Folded histogram of pairs data
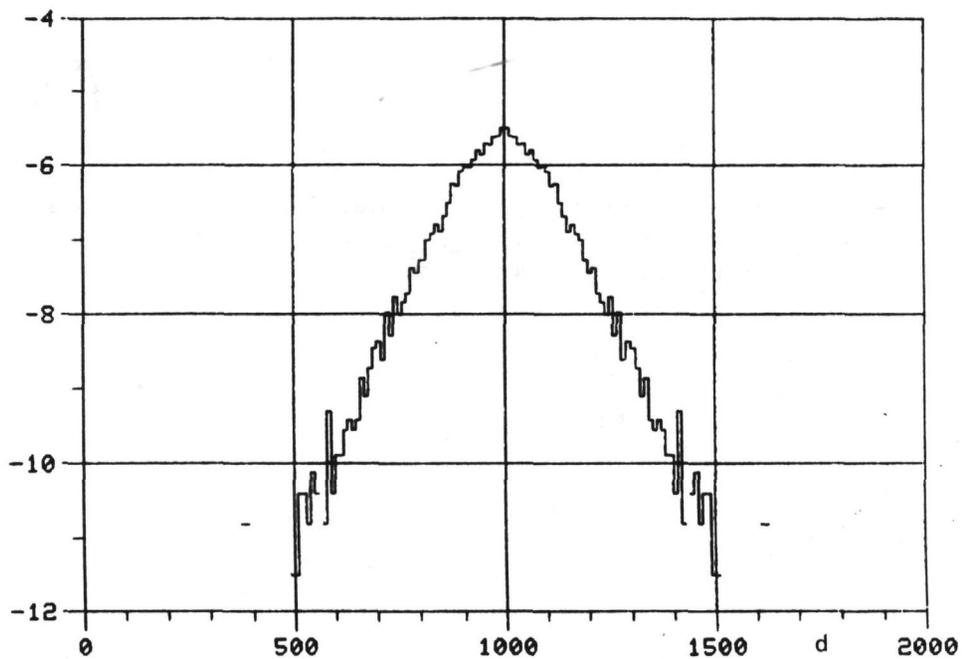
Figure 5 Logarithm (base e) of folded histogram of pairs data
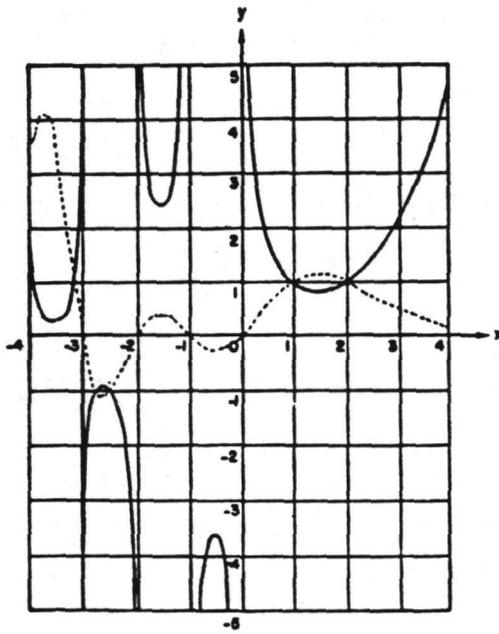
Figure 6.1   *Gamma function*.

————, $y = \Gamma(x)$,   - - - - , $y = 1/\Gamma(x)$
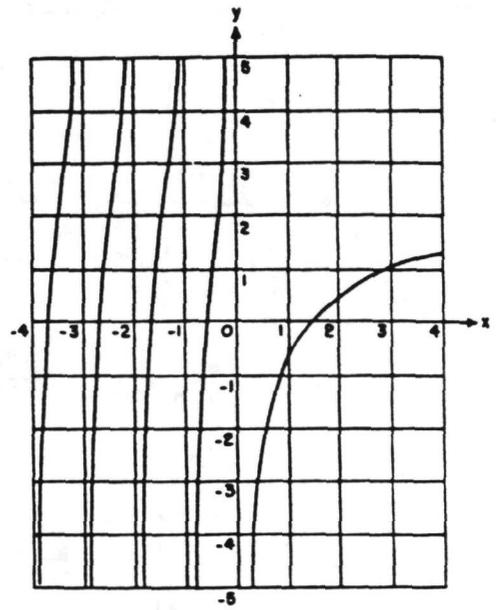
Figure 6.2   *Psi function*.

$y = \psi(x) = d \ln \Gamma(x)/dx$
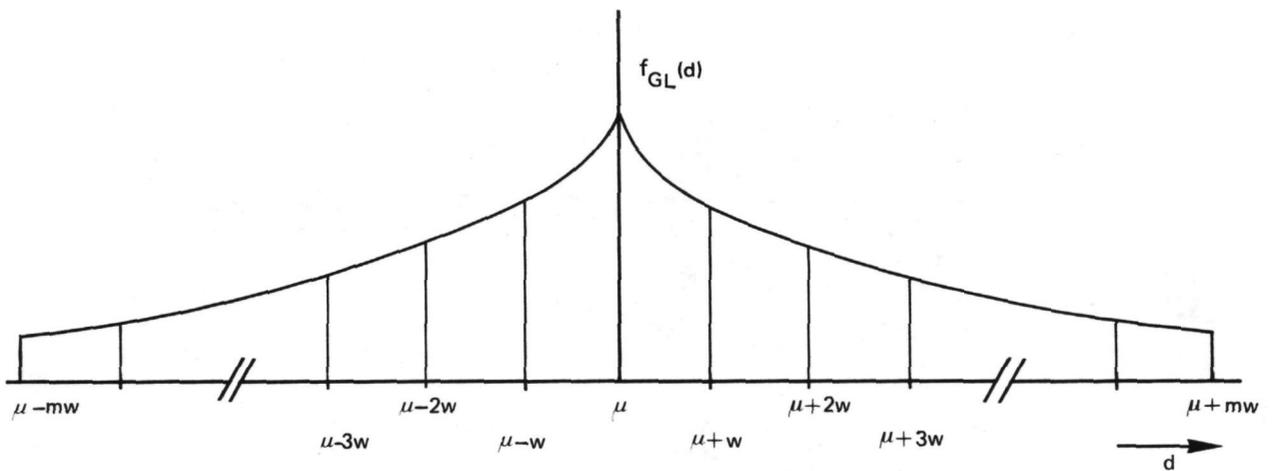
Figure 7   Subdivision of the vertical distance d axis into m classes
on the right and m on the left of the mean value μ.
Class interval w
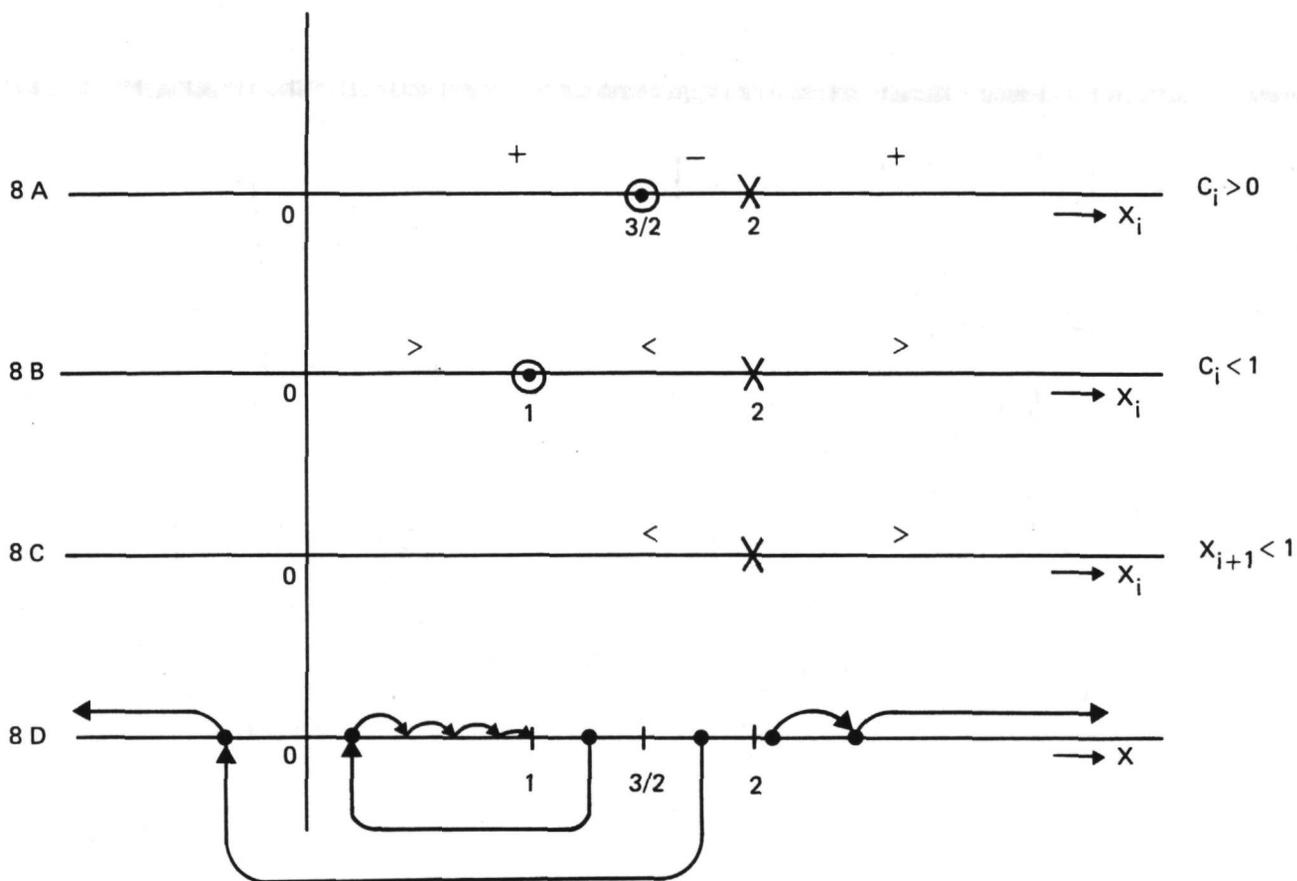
Figure 8 Analysis (8A - 8C) and summary (8D) of the convergence properties of the iteration process given by eq.(140)
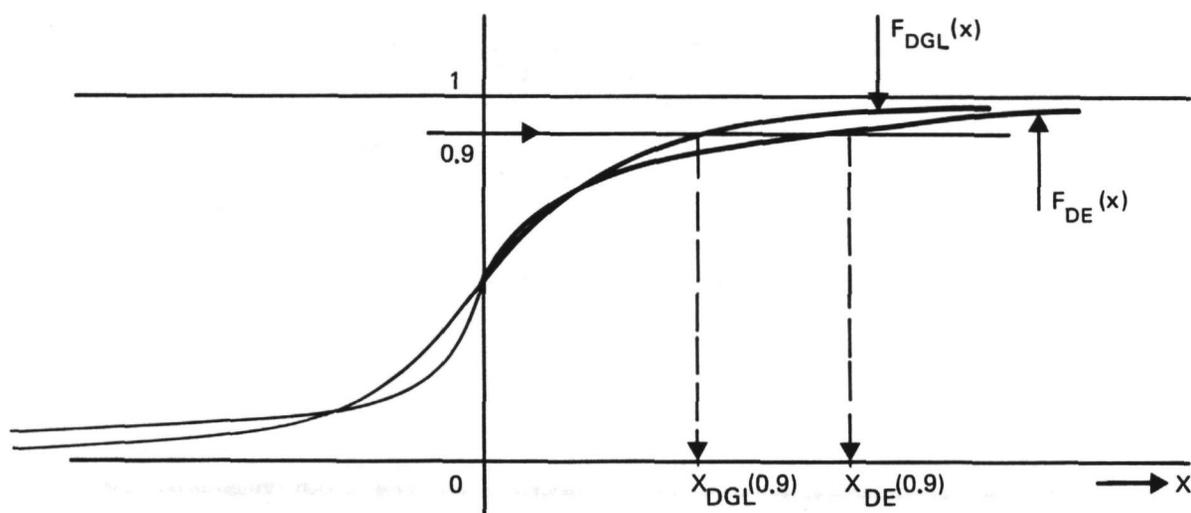
Figure 9 Using the thick tail of a Double Exponential probability distribution to determine an upperbound for the initial interval needed by the method of Regula Falsi
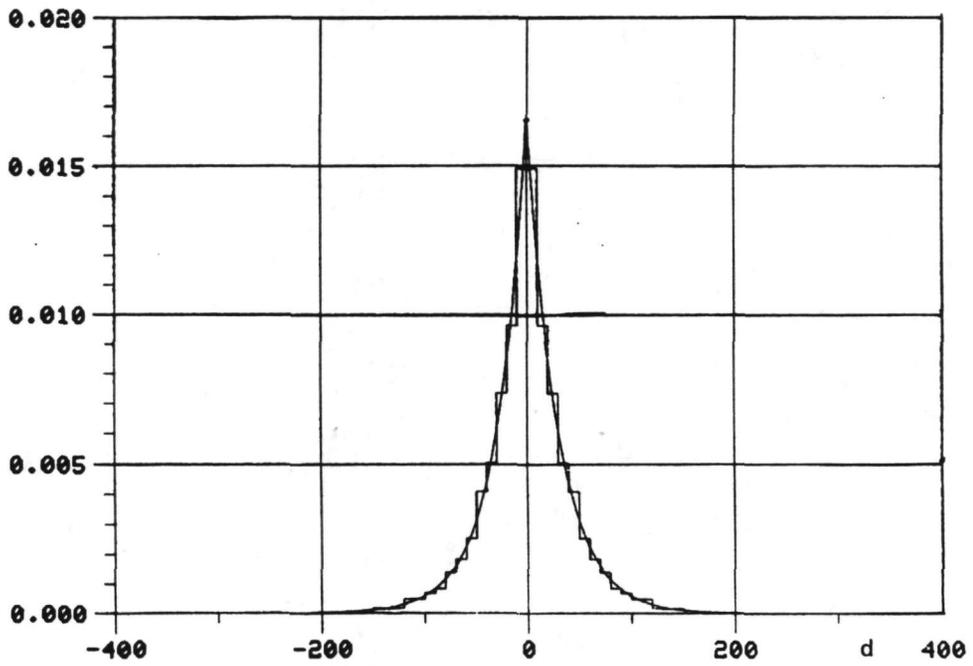
Figure 10   The simulated Double Exponential density with the folded
             histogram superimposed
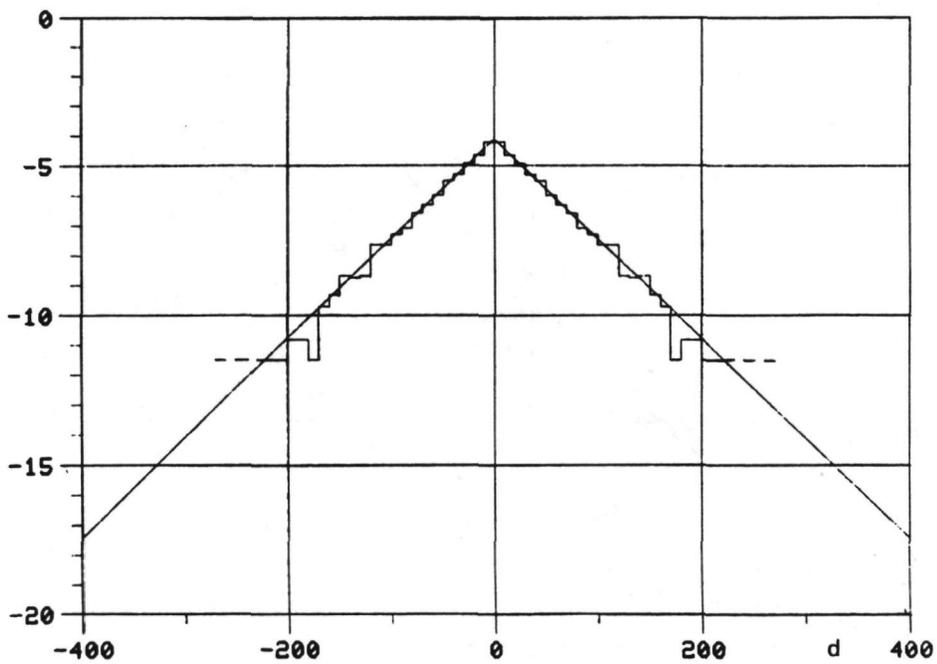
Figure 11   Logarithm (base e) of the simulated Double Exponential
             density and of the folded histogram
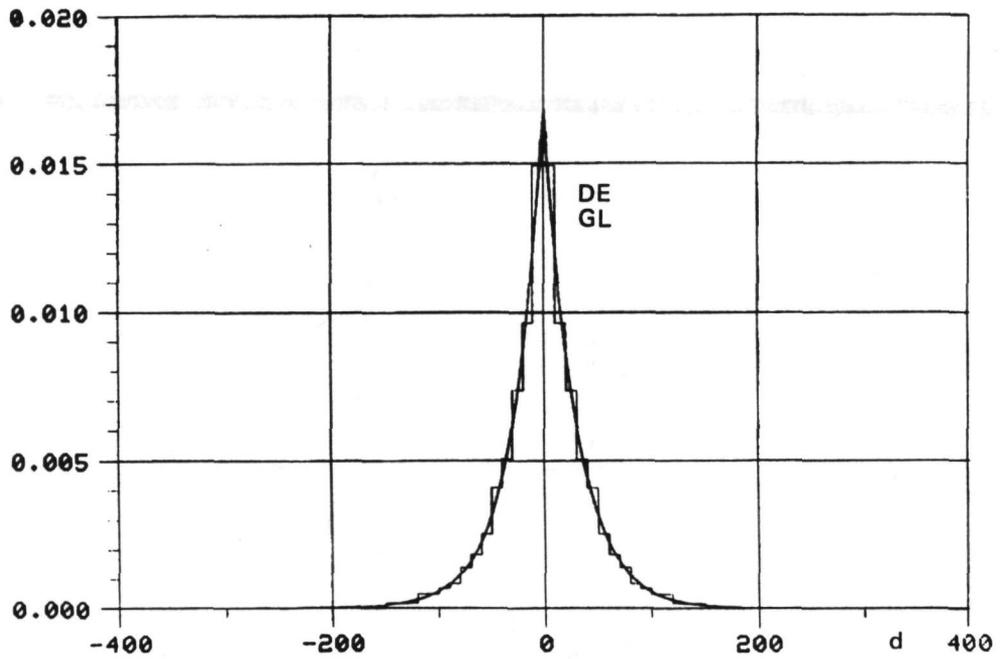
Figure 12 The simulated Double Exponential density, with the estimated
Double Exponential and Generalized Laplace densities as well
as the folded histogram superimposed
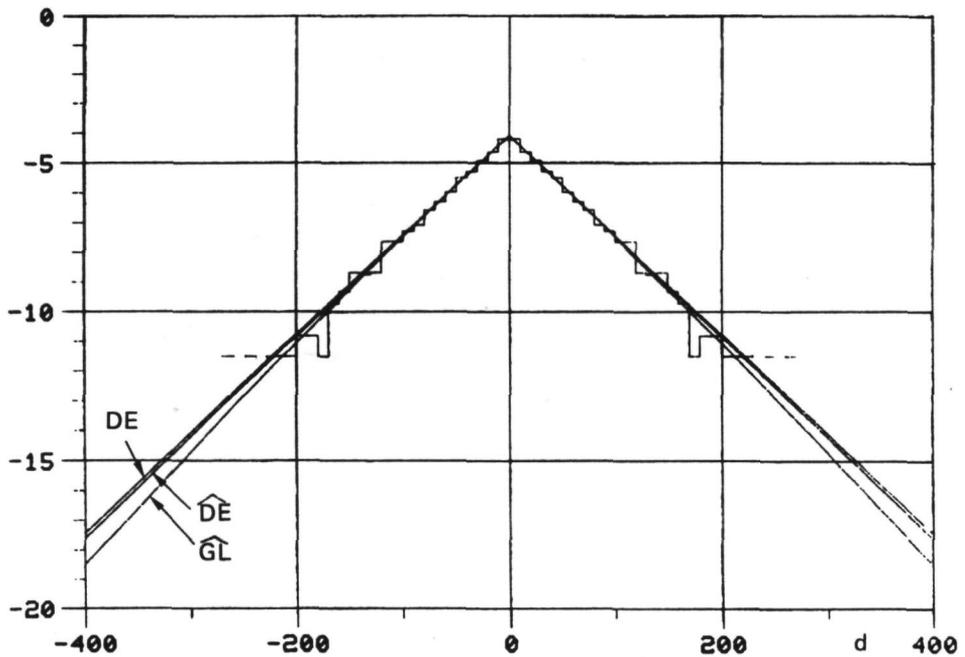
Figure 13 Logarithm (base e) of the simulated and estimated probability
densities, and of the folded histogram
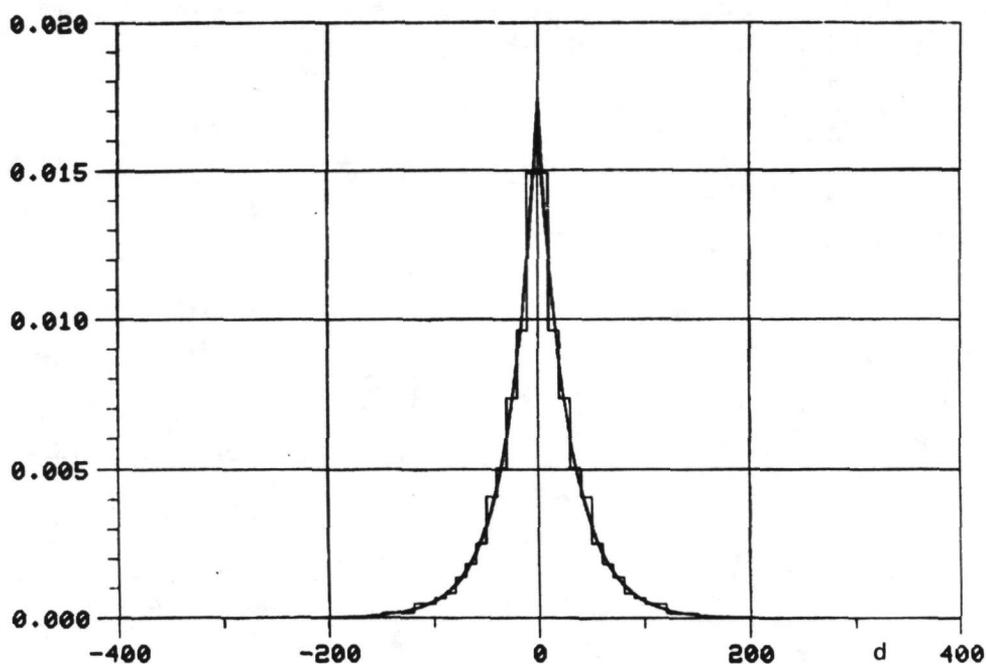
Figure 14   The simulated density, with the estimated Double Exponential
            and Generalized Laplace densities (based on non-grouped data)
            and the folded histogram superimposed
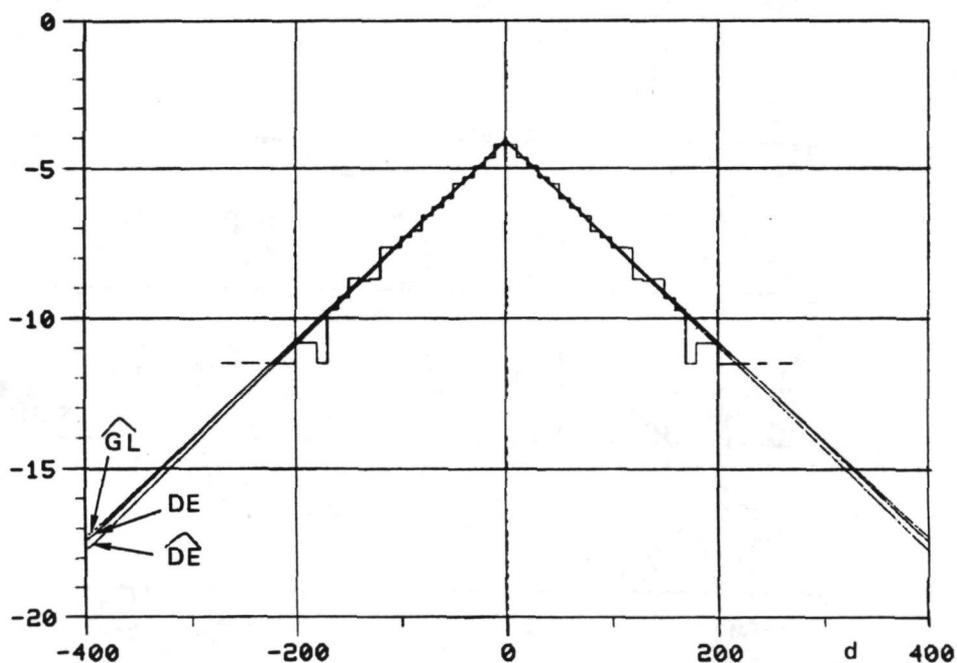
Figure 15   Logarithm (base e) of the simulated and estimated probability
            densities (based on non-grouped data), and of the folded
            histogram
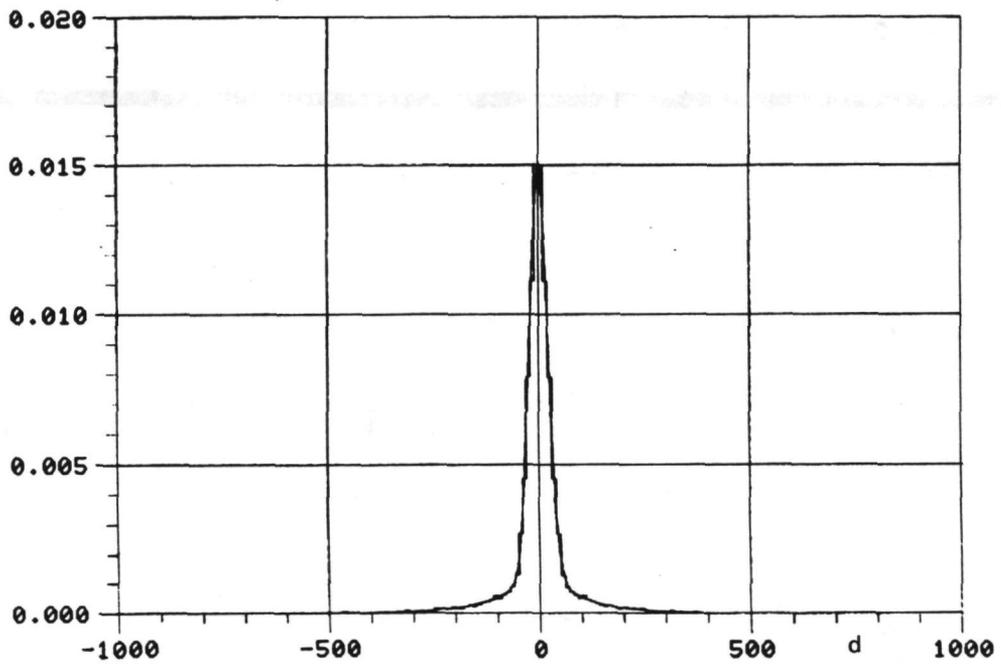
Figure 16   The simulated Gaussian Double Exponential density, with the
four estimated Double Generalized Laplace densities and the
folded histogram superimposed
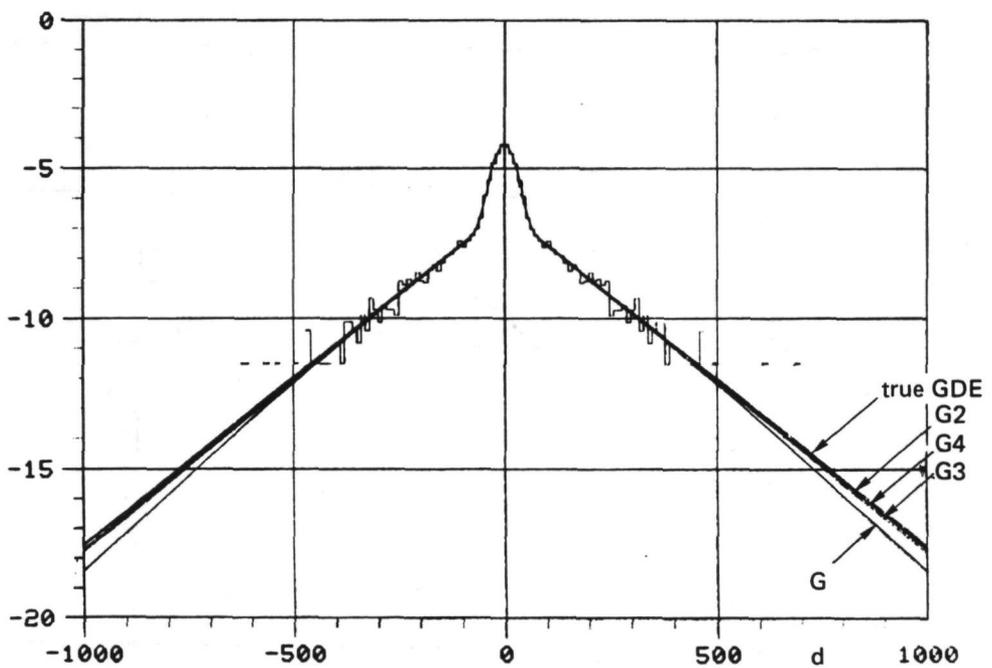
Figure 17   Logarithm (base e) of the simulated and estimated probability
densities, and of the folded histogram
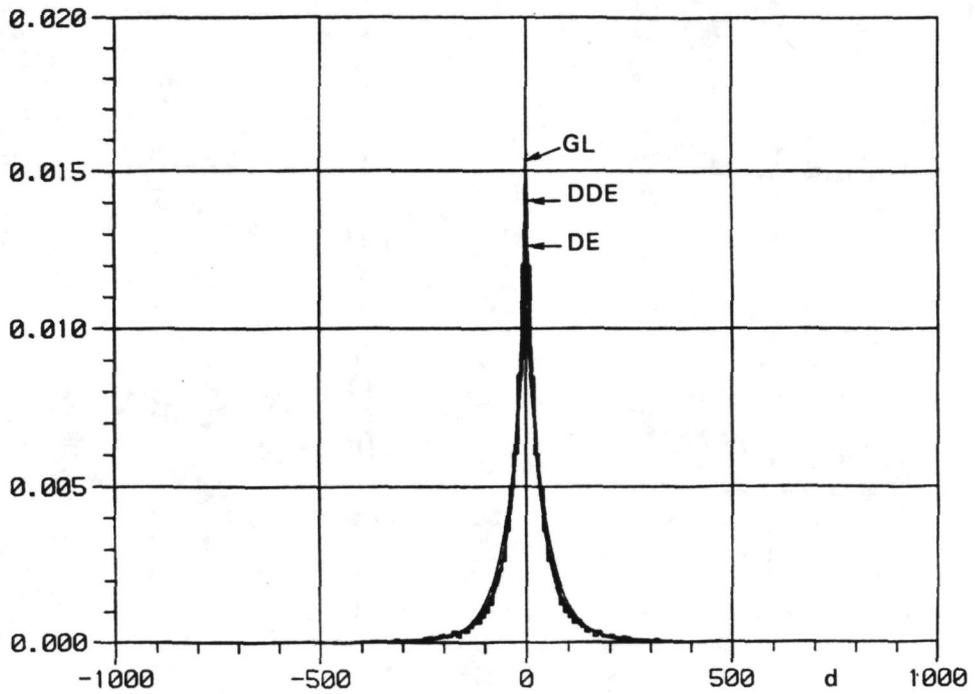
Figure 18   The simulated Double Double Exponential density, with the
estimated Double Exponential and Generalized Laplace densities
as well as the folded histogram superimposed
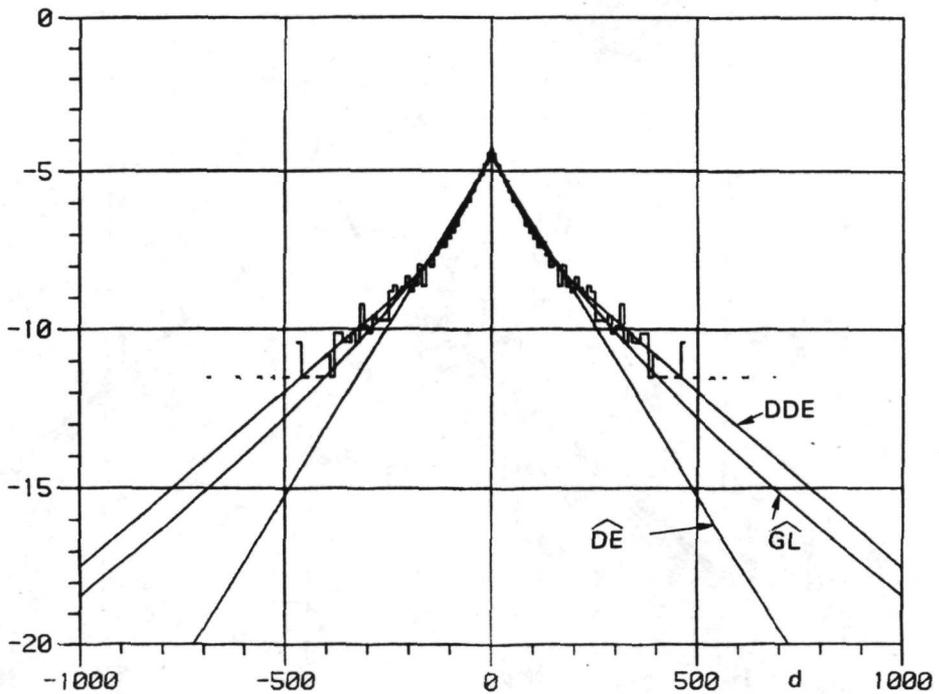
Figure 19   Logarithm (base e) of the simulated and two estimated
probability densities, and of the folded histogram
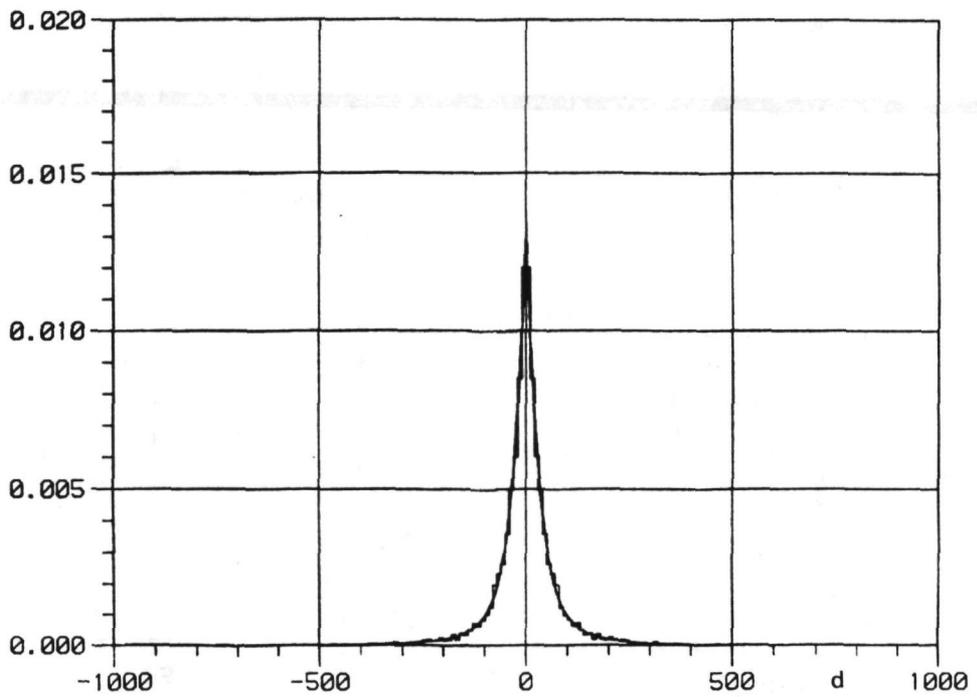
Figure 20   The simulated Double Double Exponential density, with the four
estimated Double Generalized Laplace densities and the folded
histogram superimposed
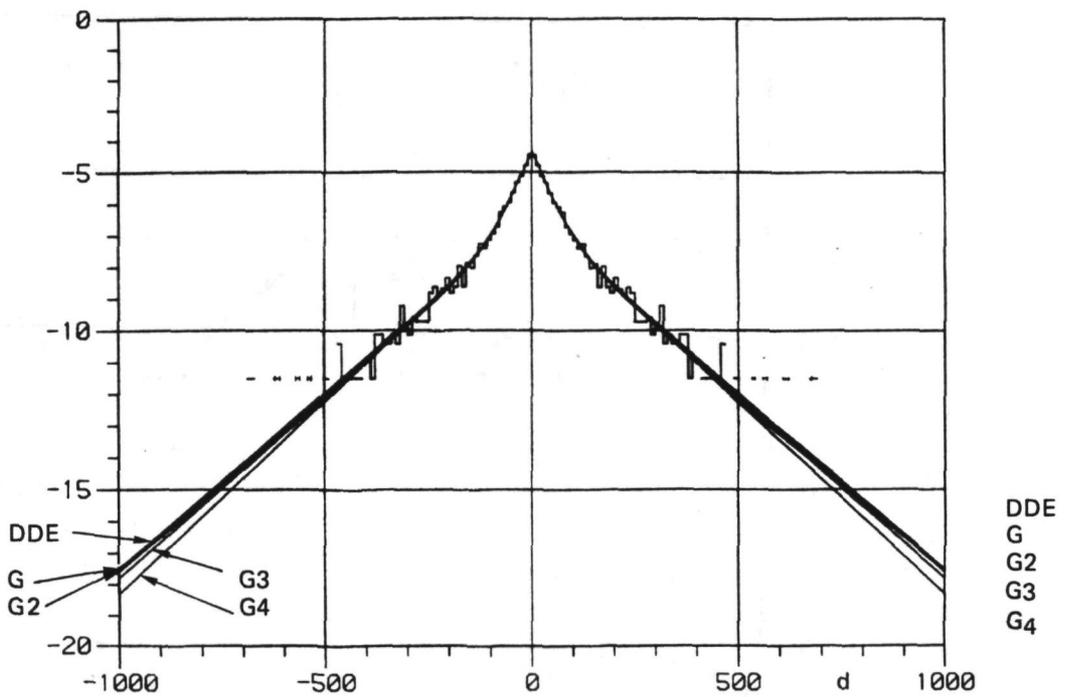
Figure 21   Logarithm (base e) of the simulated and estimated probability
densities, and of the folded histogram
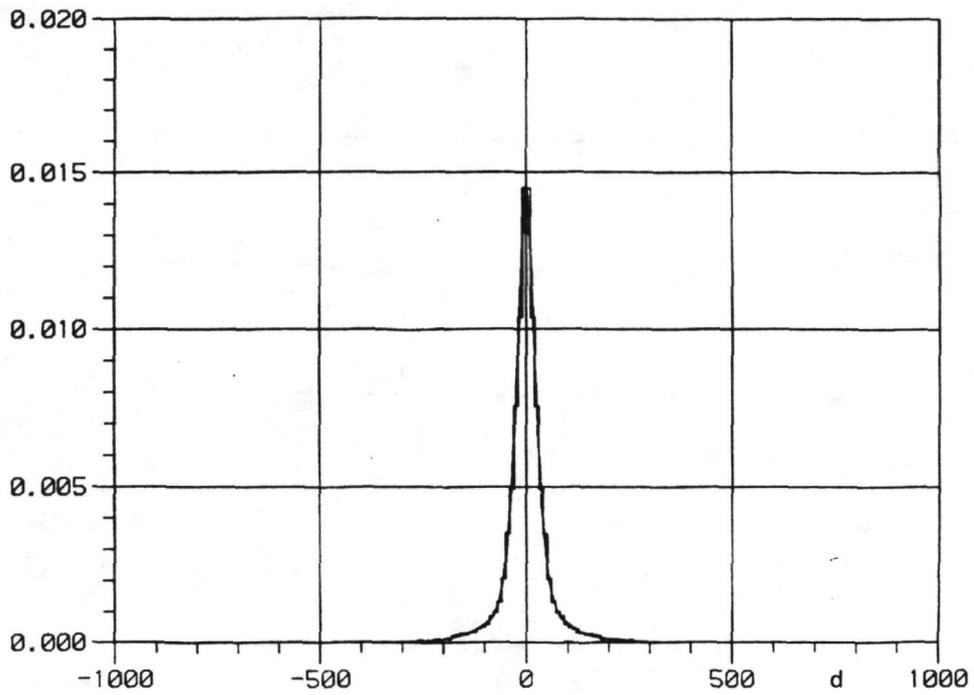
Figure 22   The simulated and the estimated Double Generalized Laplace
densities with the folded histogram superimposed (estimated
densities based on grouped data)
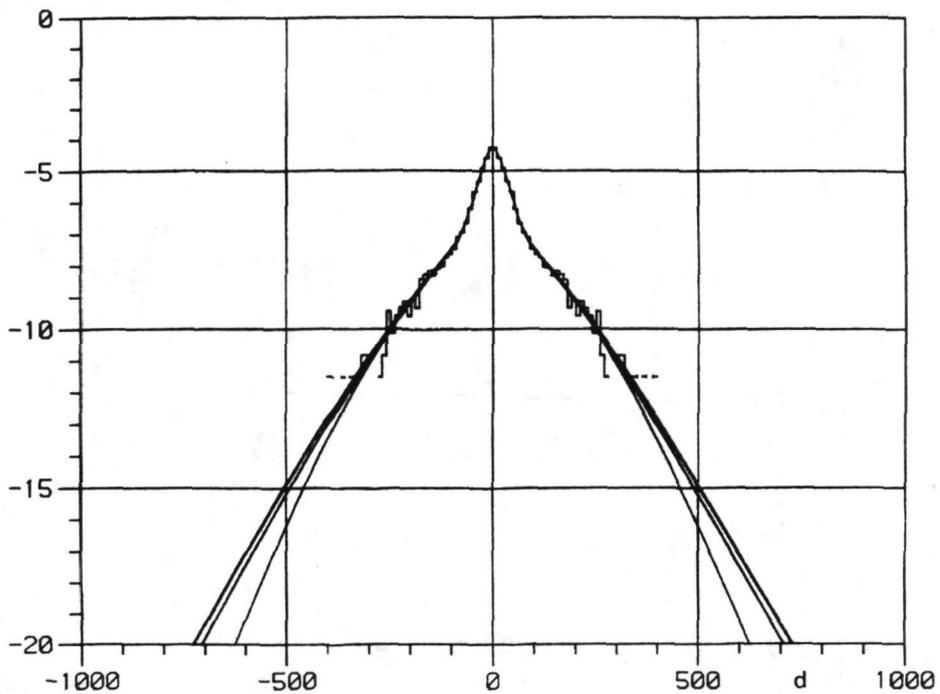
Figure 23   Logarithm (base e) of the simulated and estimated probability
densities (based on grouped data), and of the folded histogram
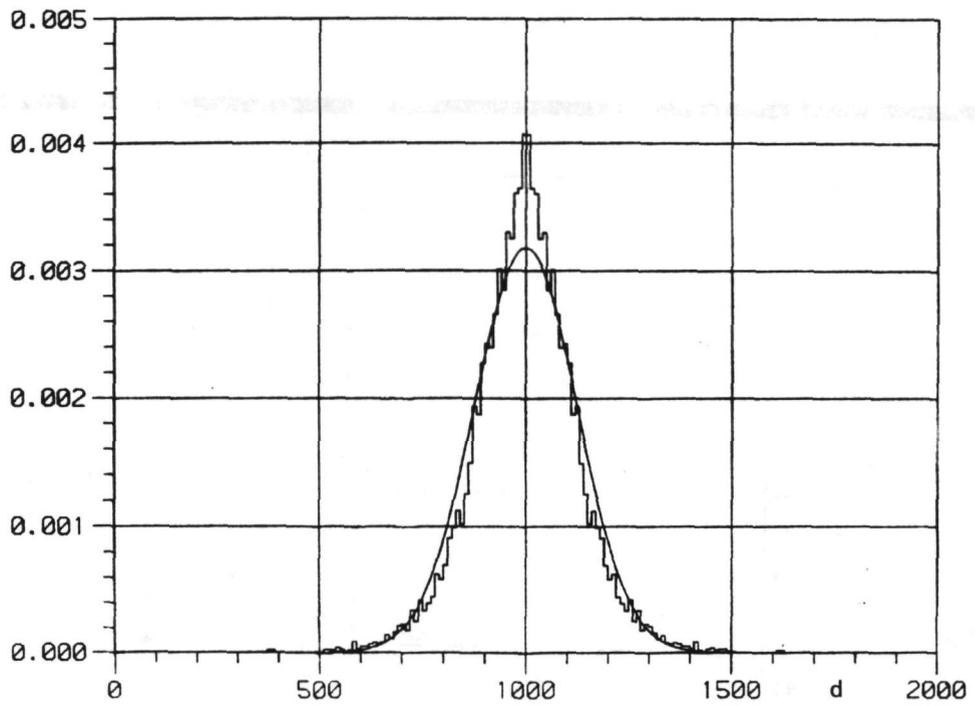
Figure 24 The estimated Gaussian density with the folded histogram superimposed
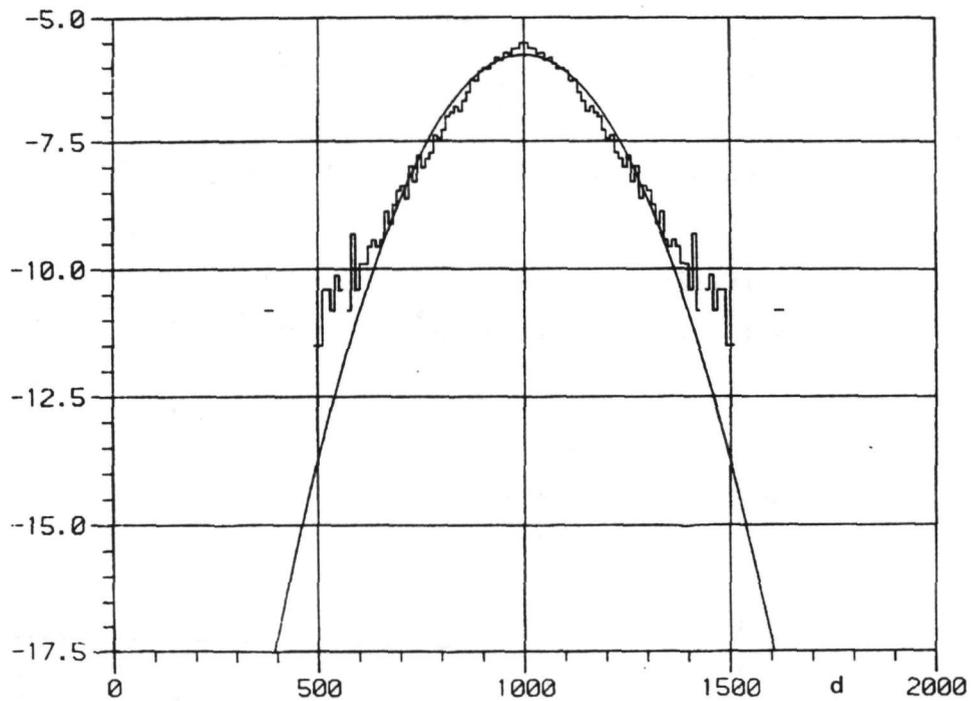
Figure 25 Logarithm (base e) of the estimated Gaussian density and of the folded histogram
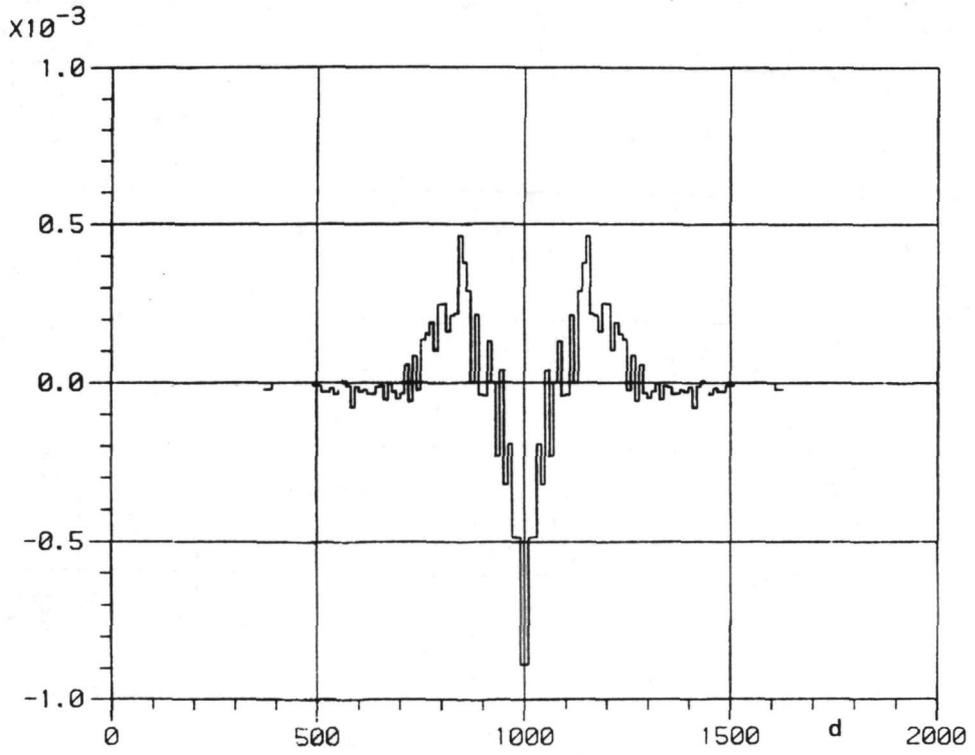
Figure 26 Linear residual plot of Gaussian density
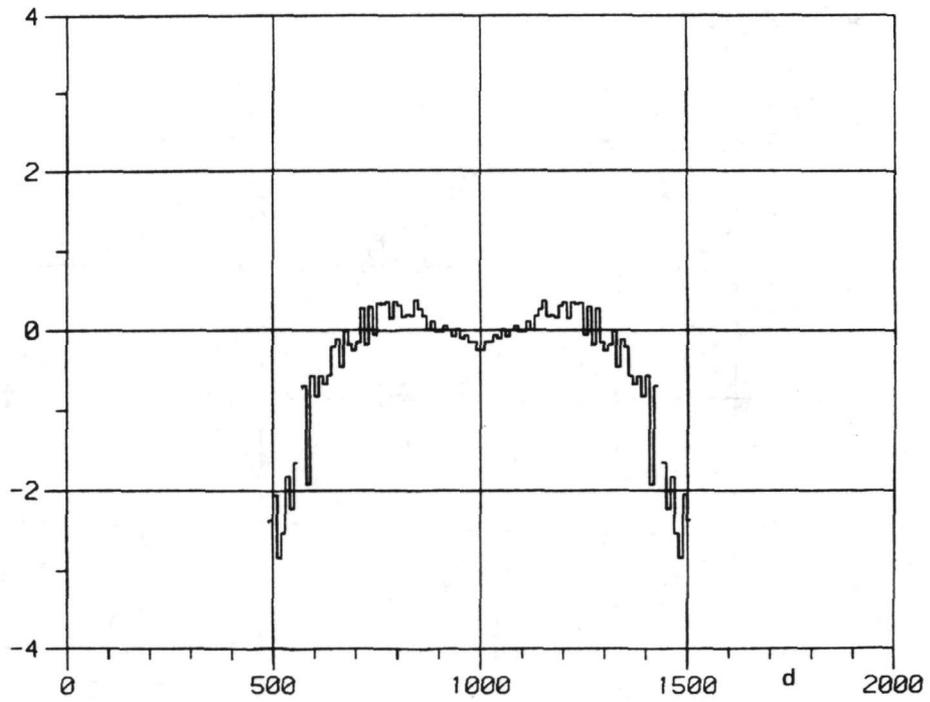
Figure 27 Logarithmic residual plot of Gaussian density
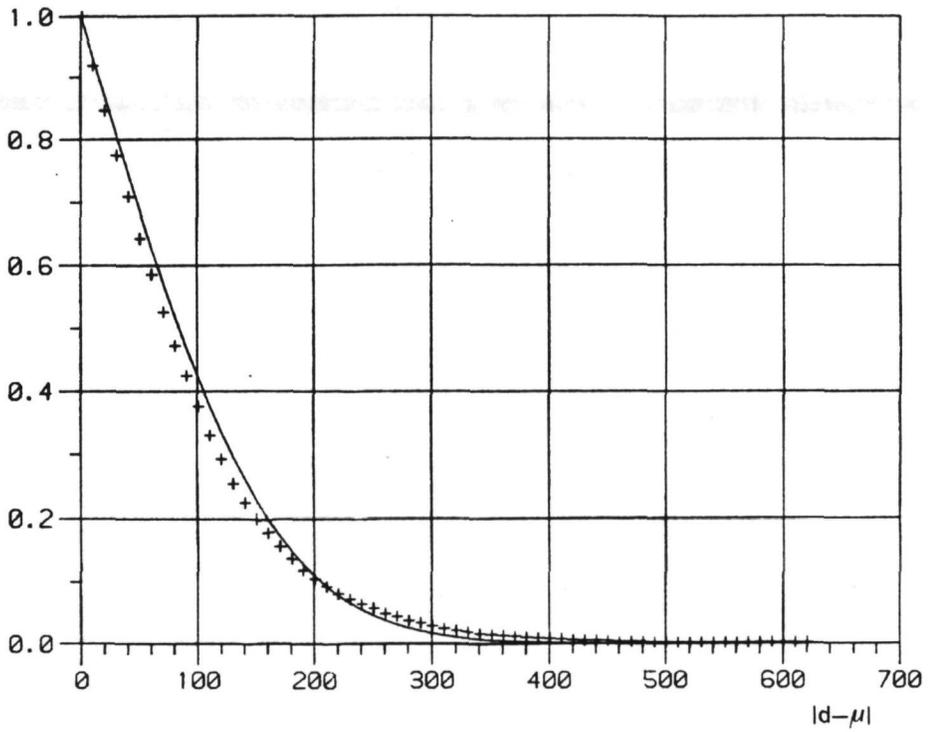
Figure 28   "1-cumulative" curve of Gaussian density
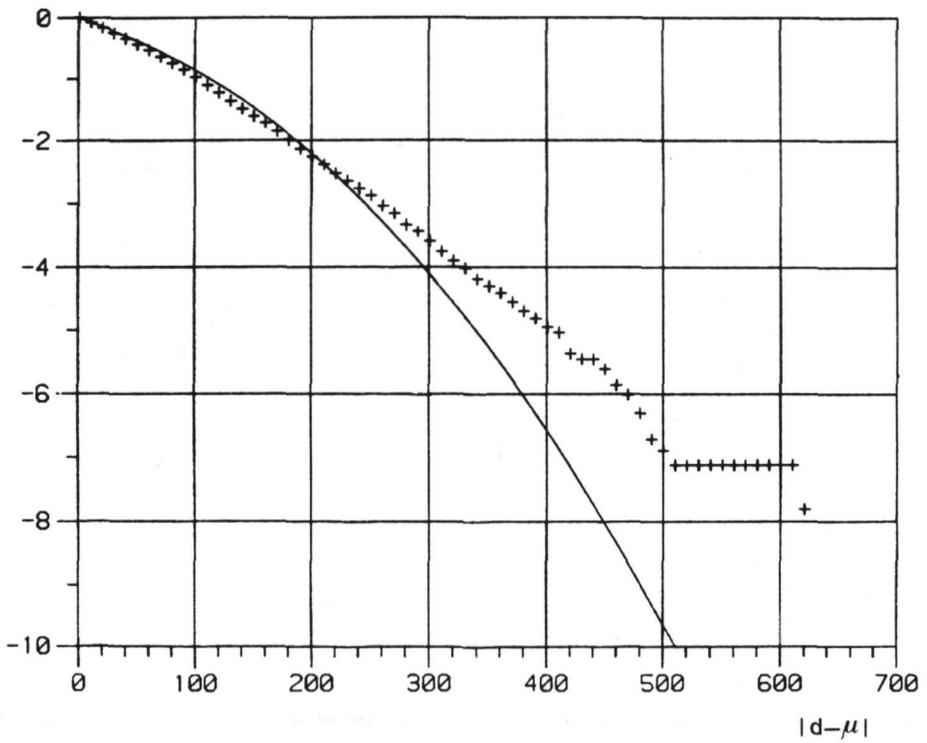
Figure 29   Logarithmic "1-cumulative" curve of Gaussian density
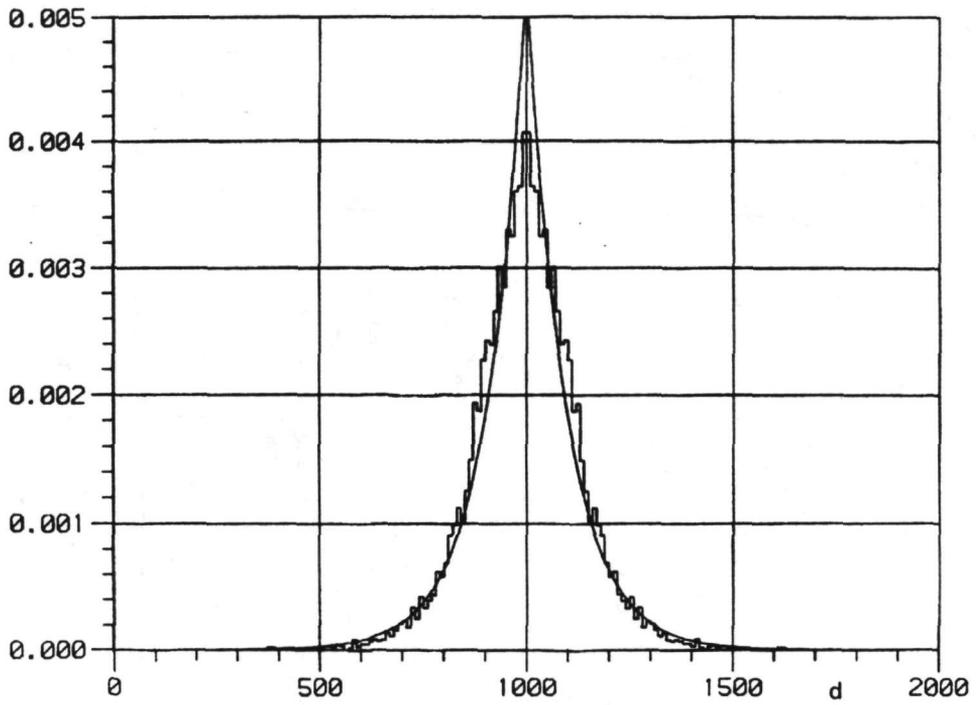
Figure 30   The estimated Double Exponential density with the folded
             histogram superimposed
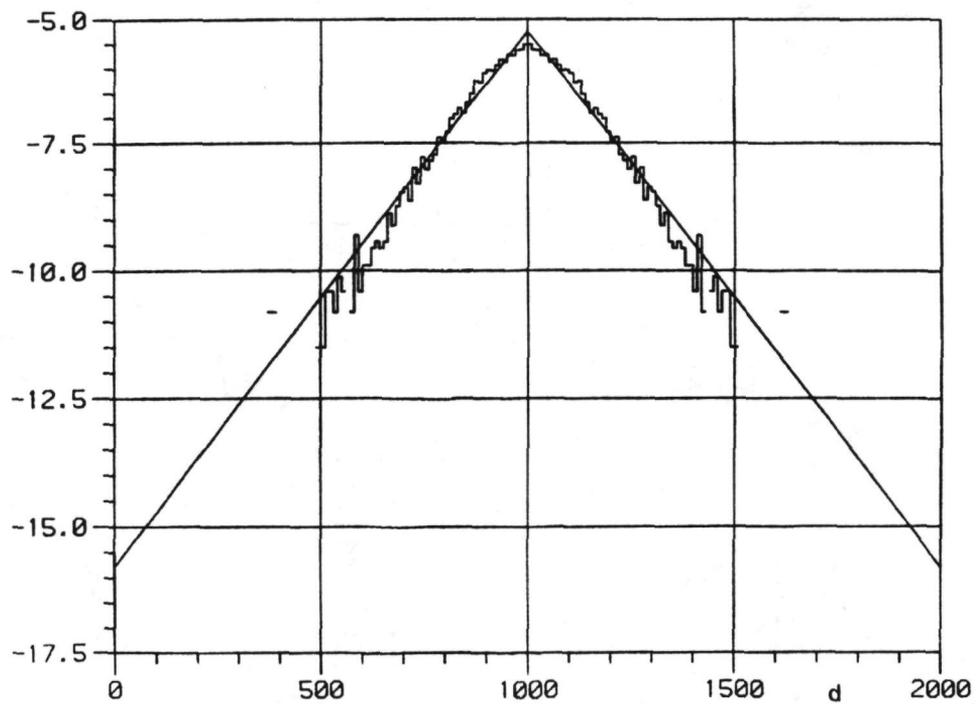
Figure 31   Logarithm (base e) of the estimated Double Exponential
             density and of the folded histogram
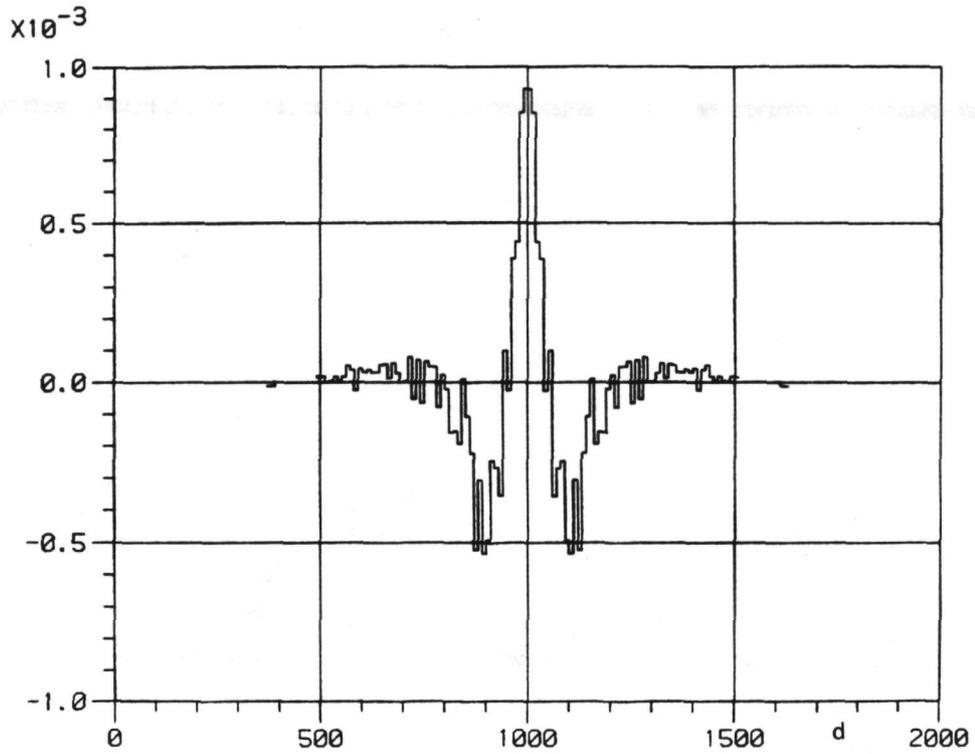
Figure 32 Linear residual plot of the Double Exponential density
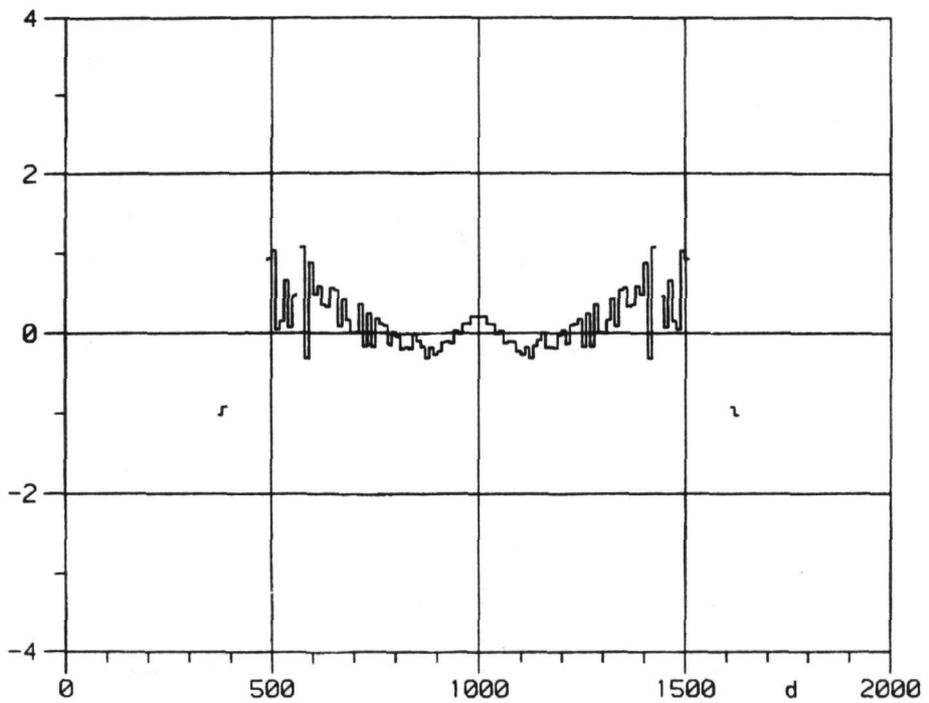
Figure 33 Logarithmic residual plot of the Double Exponential density
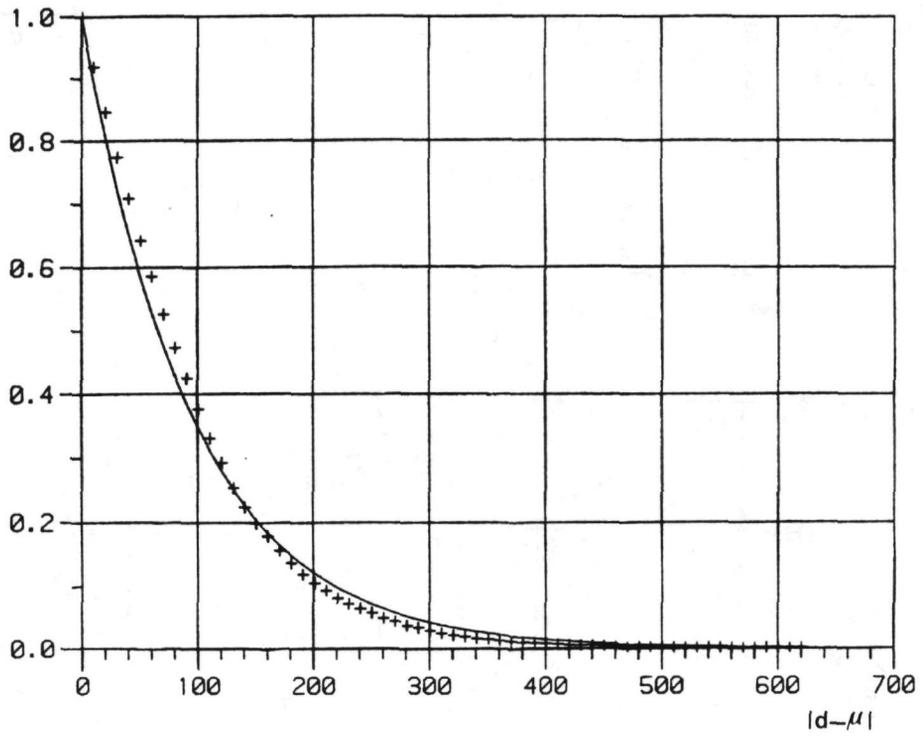
Figure 34 "1-cumulative" curve of the Double Exponential density
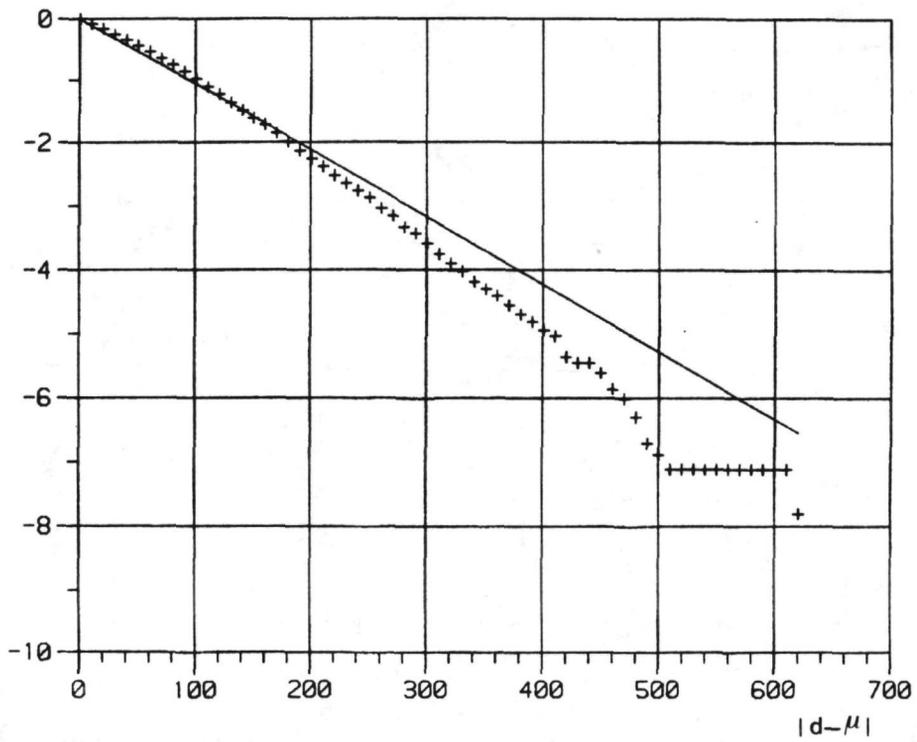
Figure 35 Logarithmic "1-cumulative" curve of the Double Exponential density
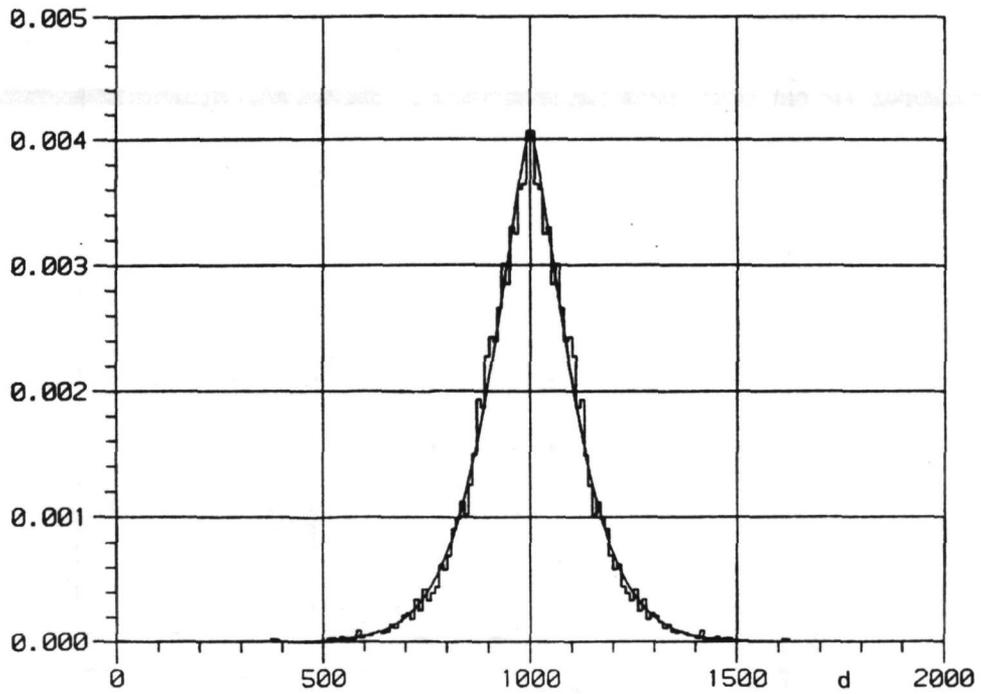
Figure 36 The estimated Generalized Laplace density with the folded histogram superimposed
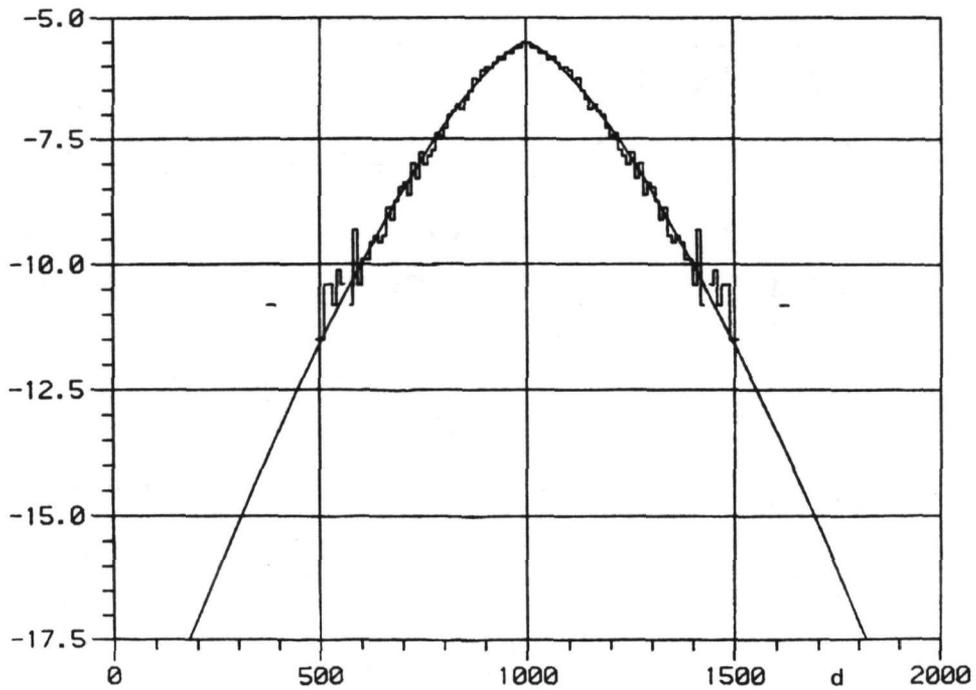
Figure 37 Logarithm (base e) of the estimated Generalized Laplace density and of the histogram
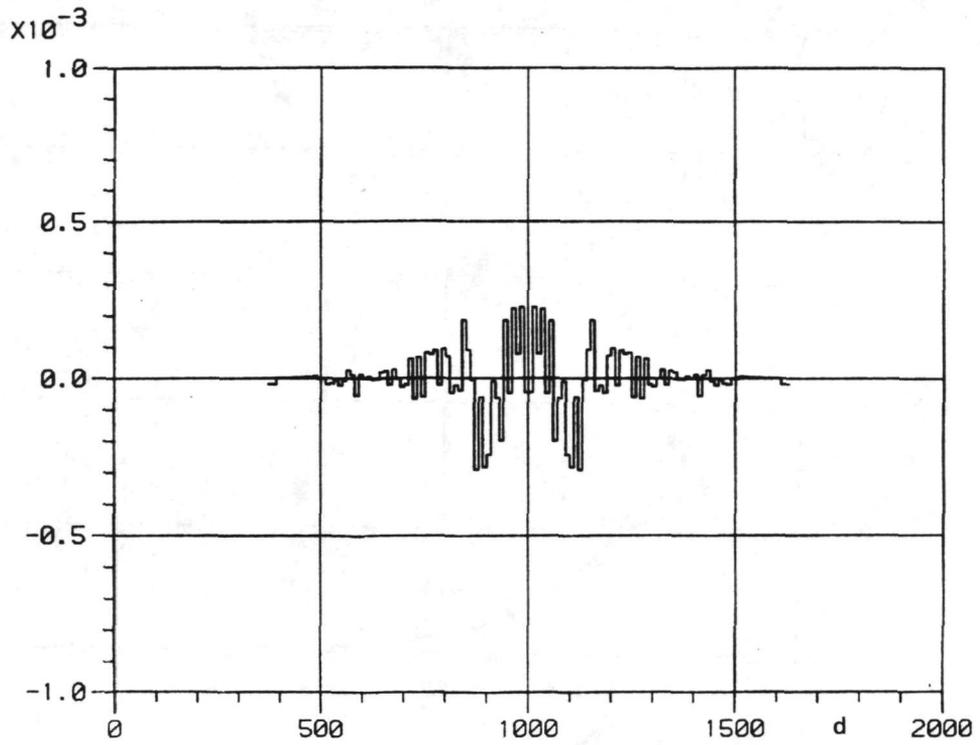
Figure 38   Linear residual plot of the Generalized Laplace density
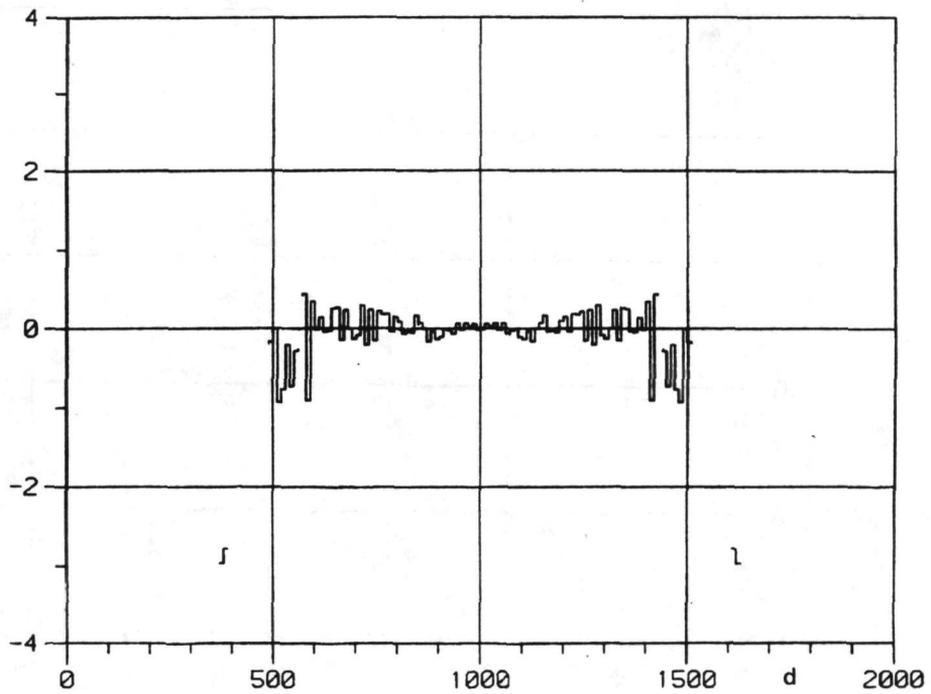
Figure 39   Logarithmic residual plot of the Generalized Laplace density
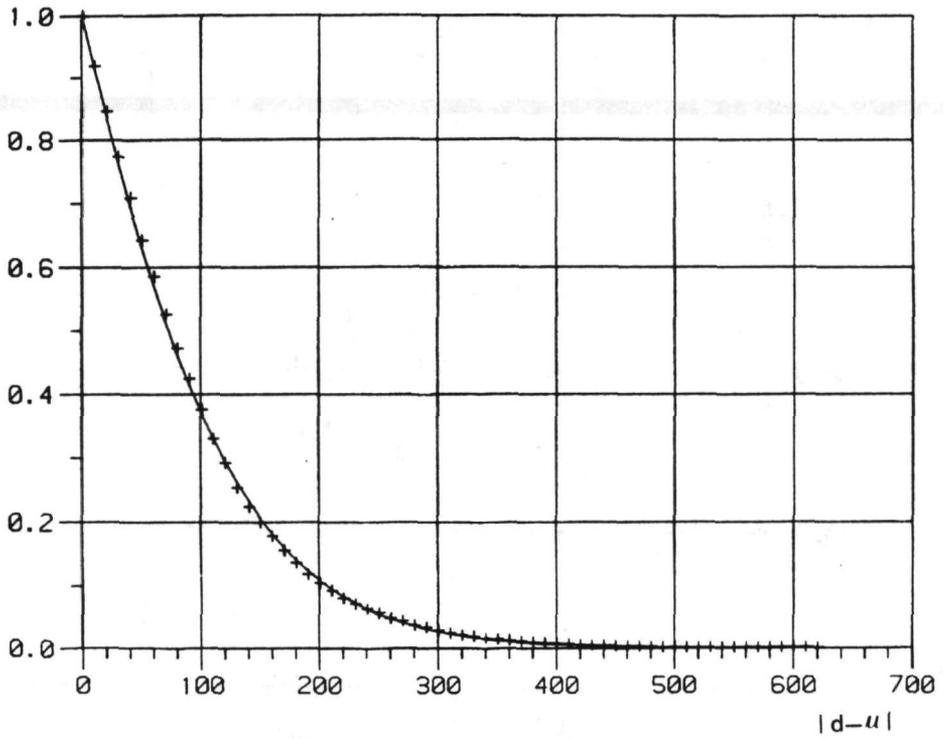
Figure 40 "1-cumulative" curve of the Generalized Laplace density
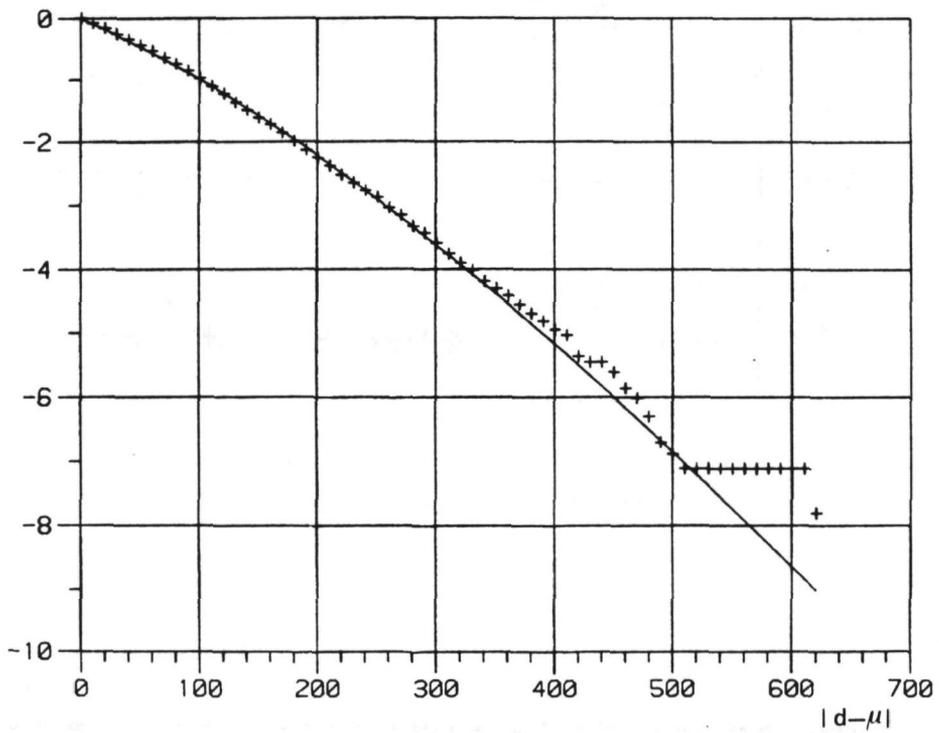
Figure 41 Logarithmic "1-cumulative" curve of the Generalized Laplace density
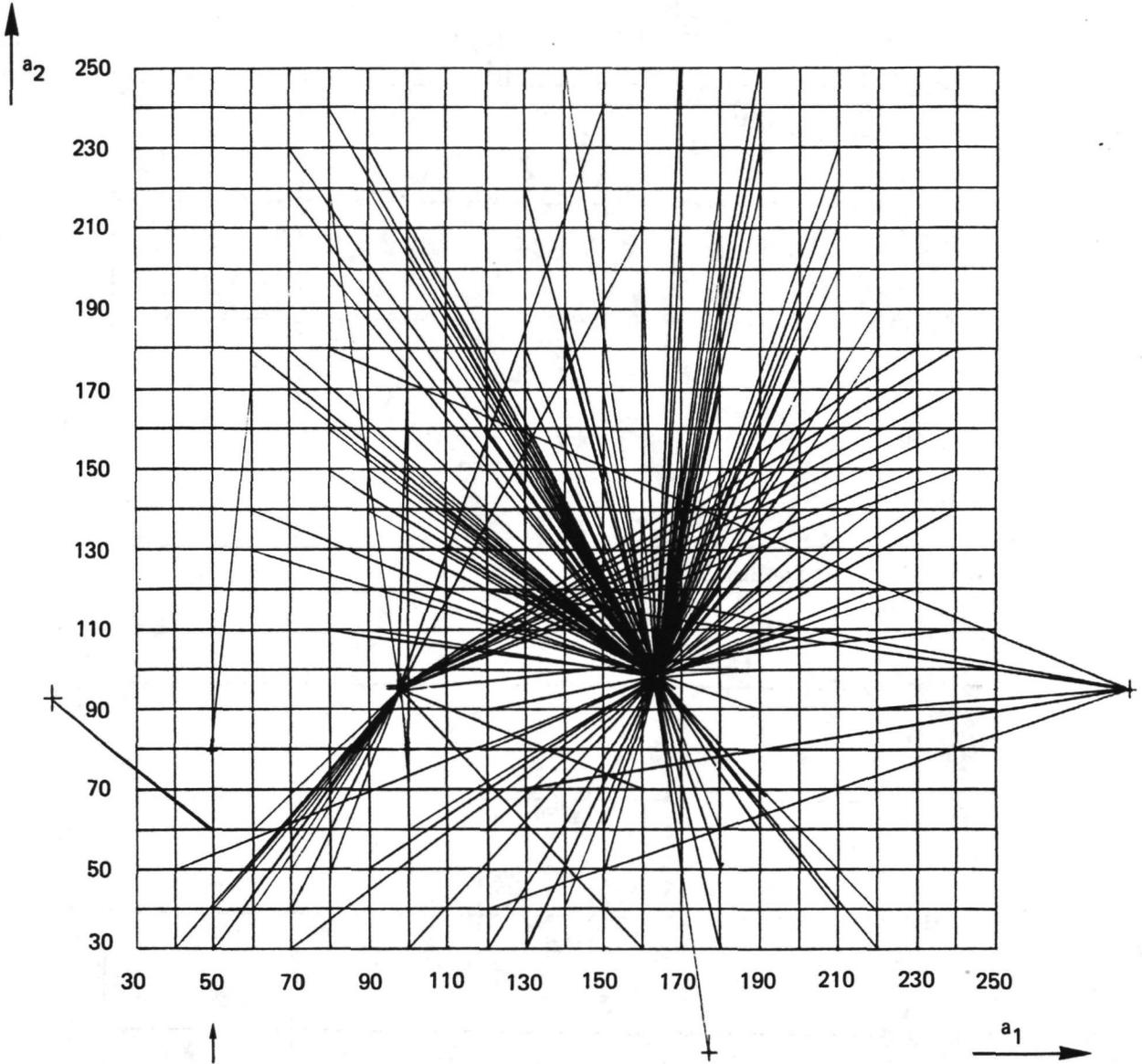
Figure 42 Diagram showing in an (a_1, a_2)-plane which gridpoint (as an initial estimate for the Newton iteration process) converges to which solution of the likelihood equations based on the Gaussian Double Exponential probability density model
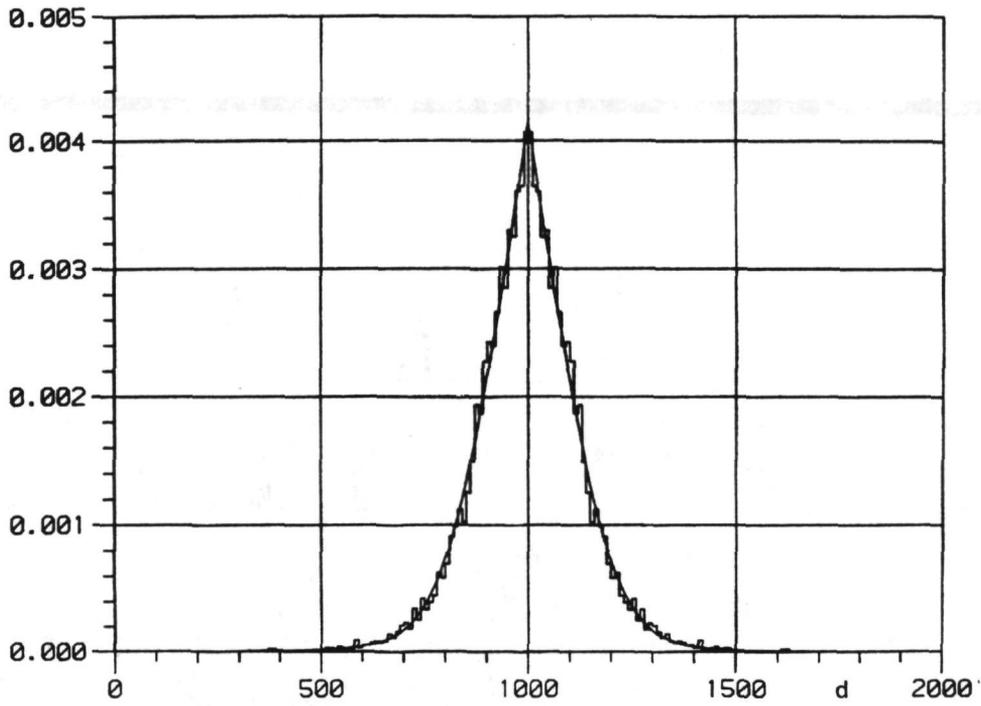
Figure 43   The estimated Gaussian Double Exponential density with the
folded histogram superimposed
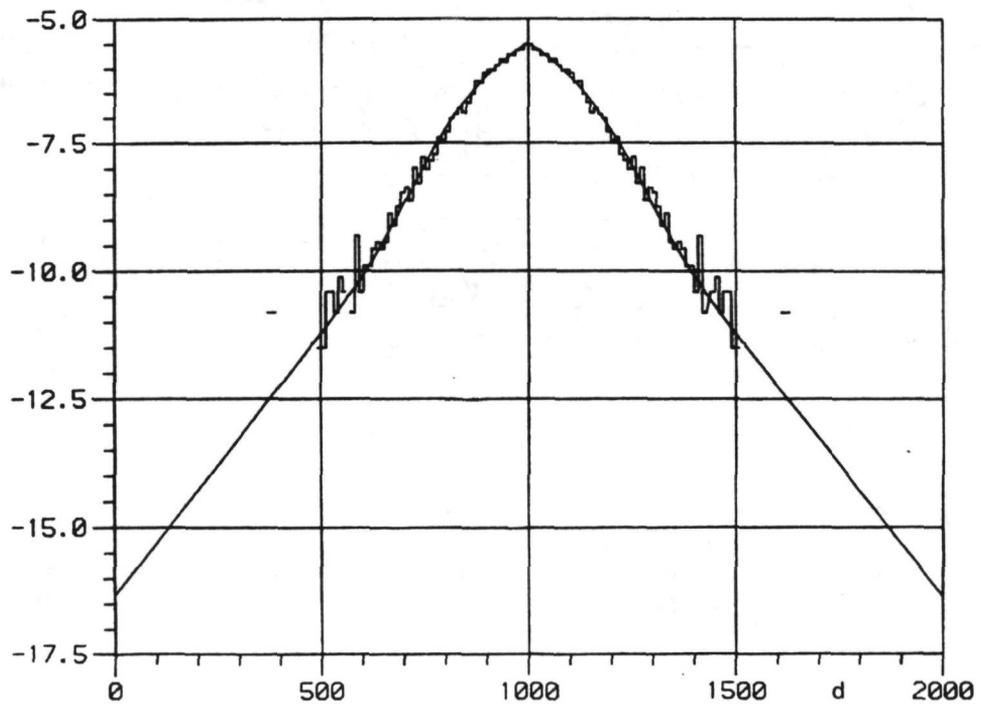
Figure 44   Logarithm (base e) of the estimated Gaussian Double
Exponential density and of the folded histogram
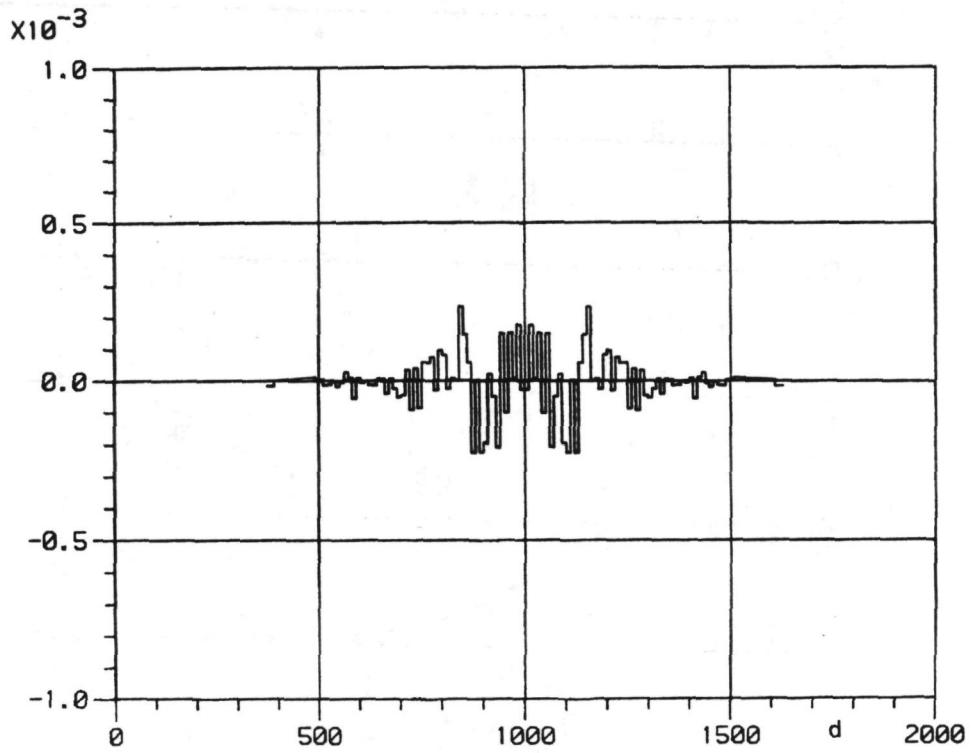
Figure 45   Linear residual plot of the Gaussian Double Exponential density
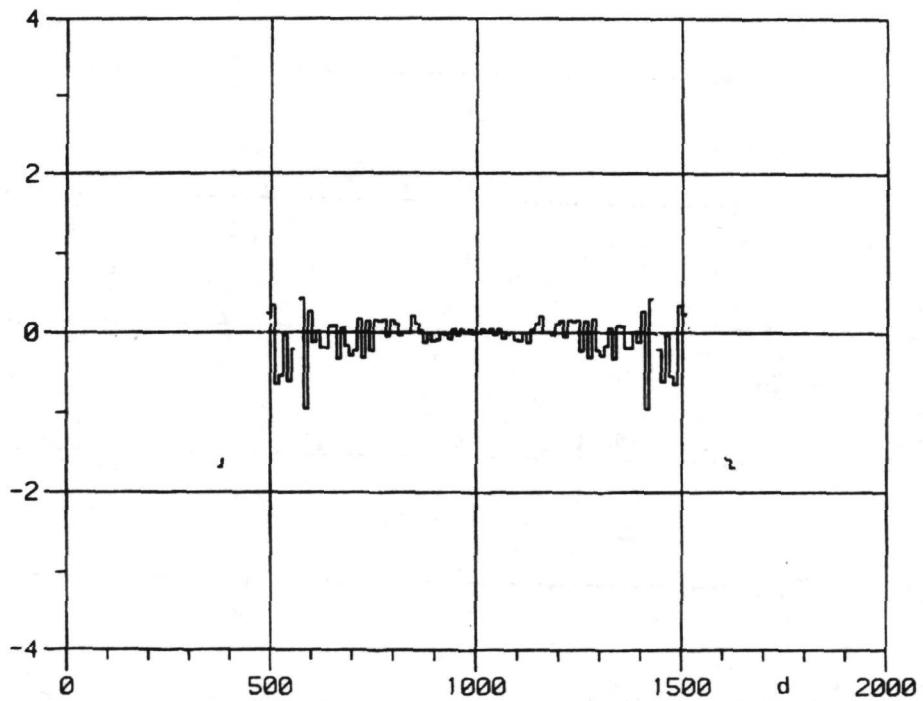
Figure 46   Logarithmic residual plot of the Gaussian Double Exponential density
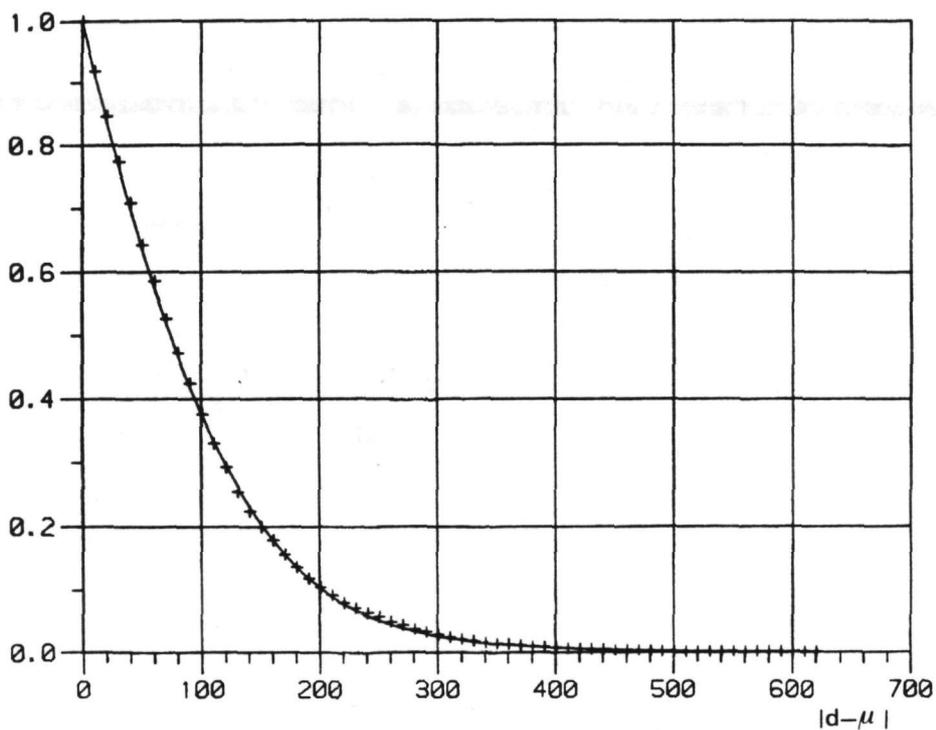
Figure 47   "1-cumulative" curve of the Gaussian Double Exponential
density
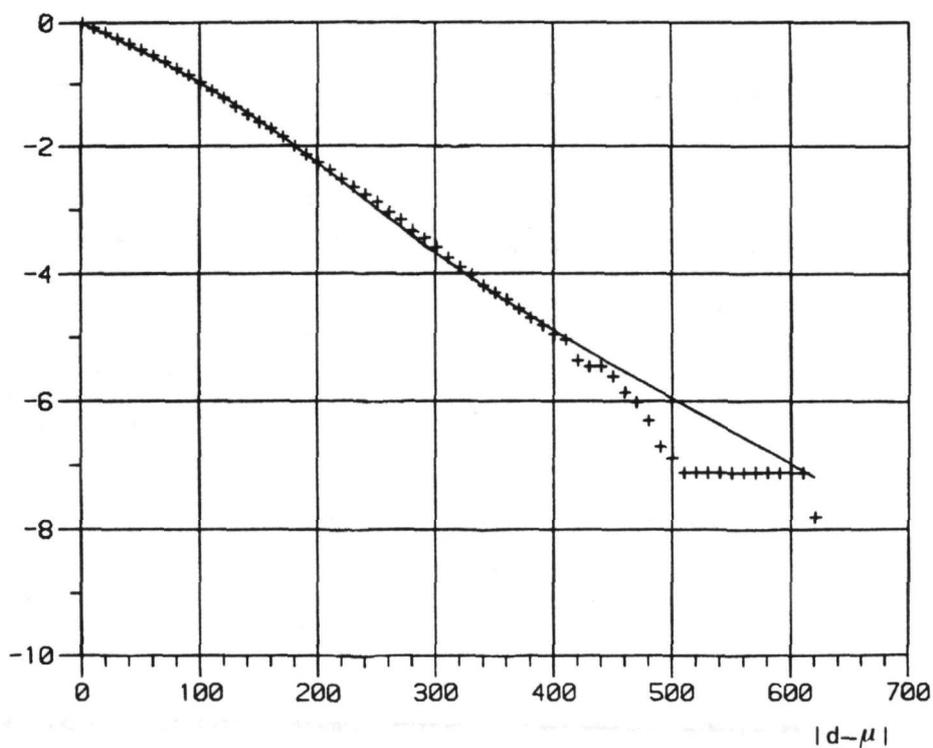
Figure 48   Logarithmic "1-cumulative" curve of the Gaussian Double
Exponential density
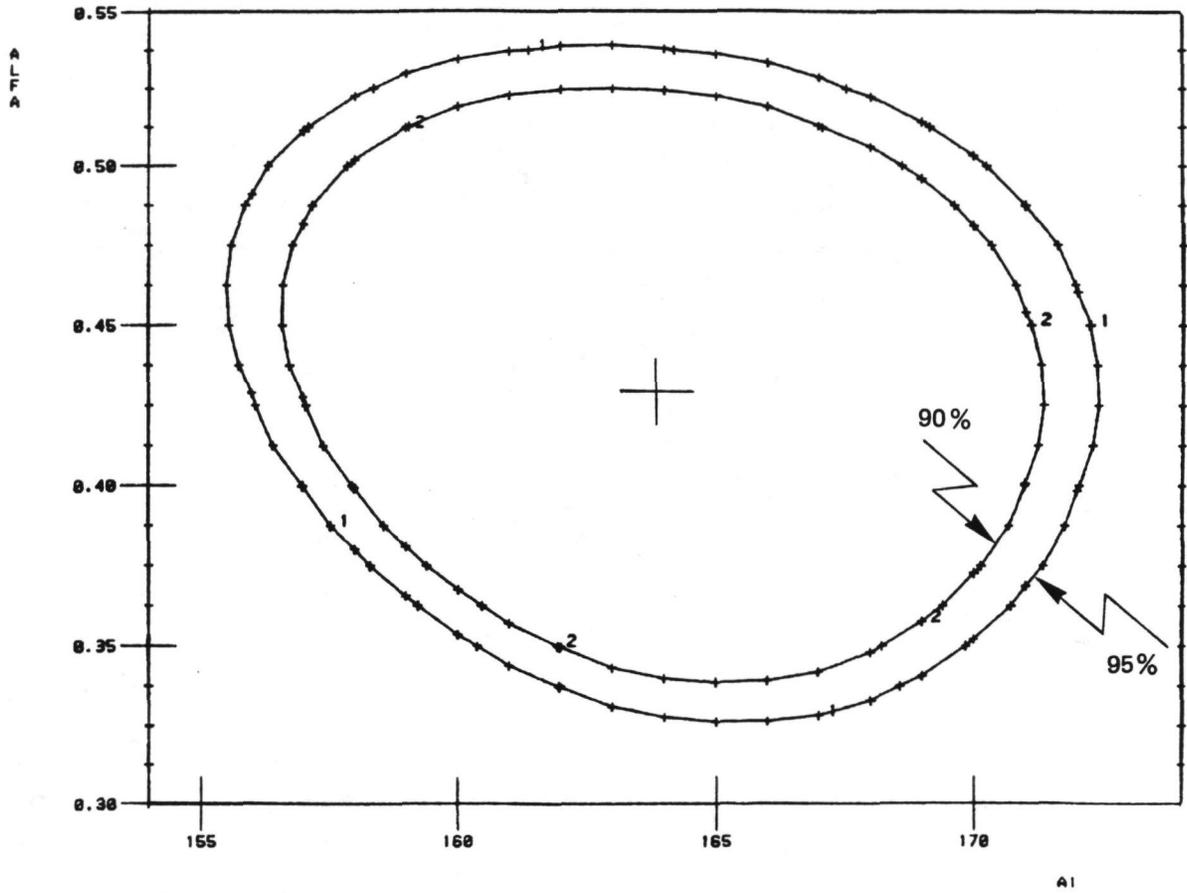
Figure 49 Two isocontours of the relative likelihood function of the
Gaussian Double Exponential probability density model
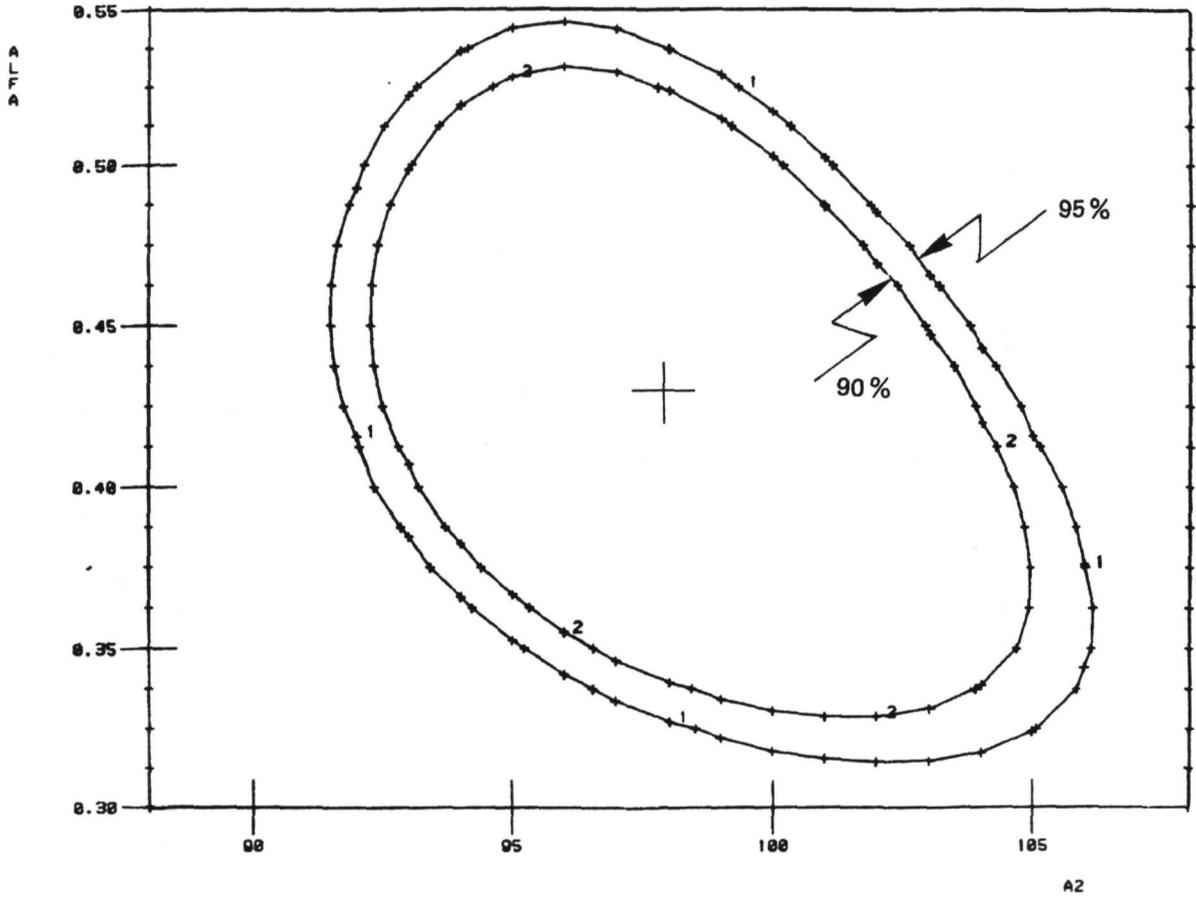
Figure 50   Two isocontours of the relative likelihood function of the
            Gaussian Double Exponential probability density model
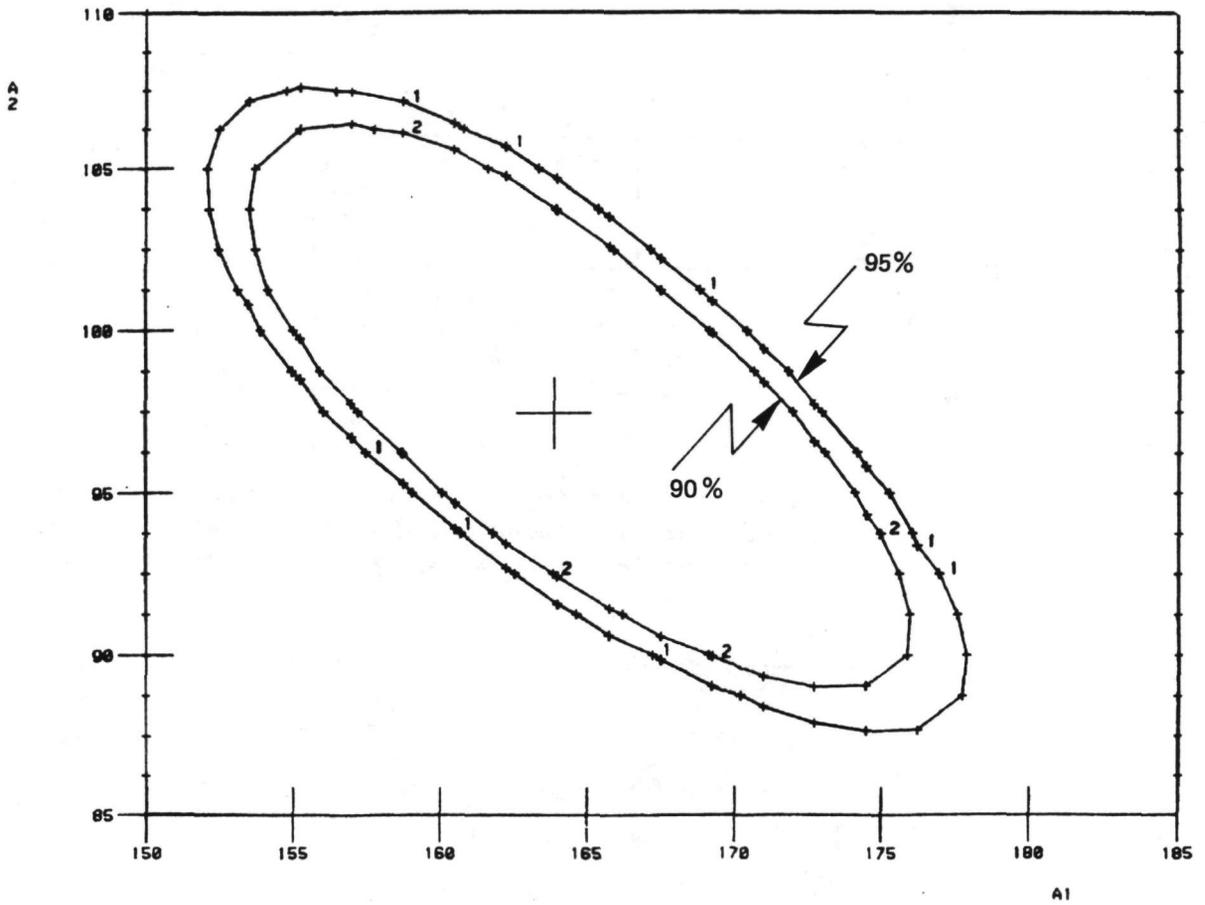
Figure 51 Two isocontours of the relative likelihood function of the Gaussian Double Exponential probability density model
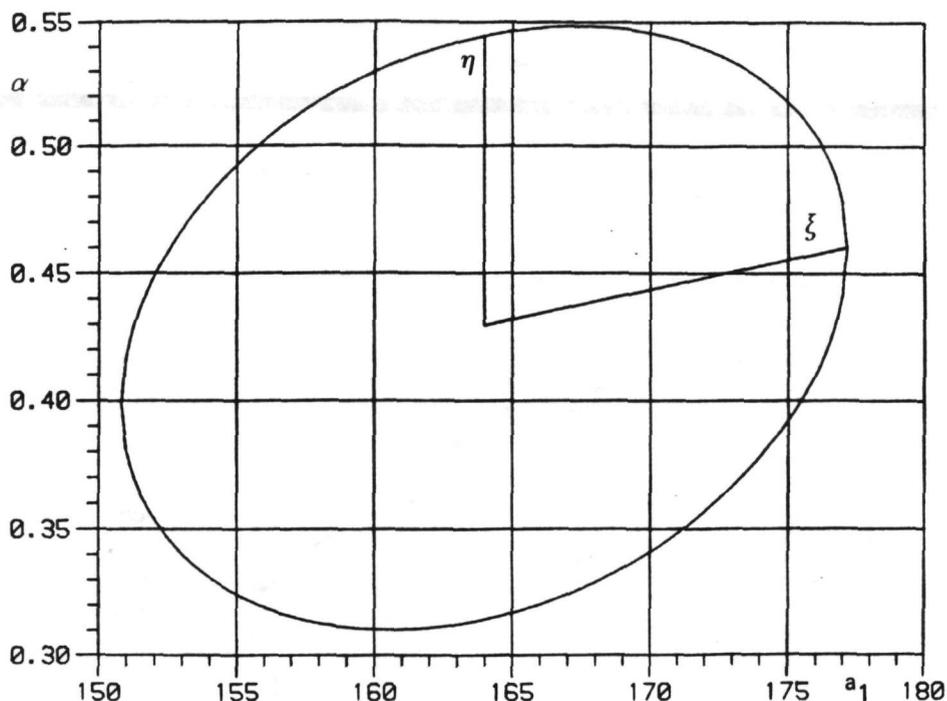
Figure 52 A 95 per cent isocontour based on the asymptotic distribution of the maximum likelihood parameter estimator for the Gaussian Double Exponential probability density model
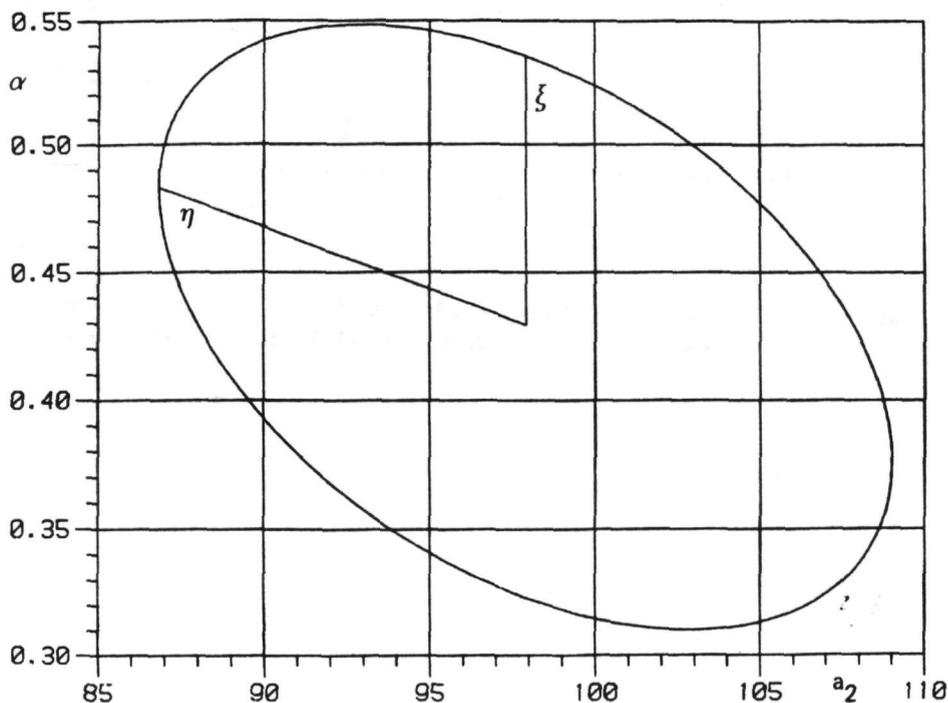
Figure 53 A 95 per cent isocontour based on the asymptotic distribution of the maximum likelihood parameter estimator for the Gaussian Double Exponential probability density model
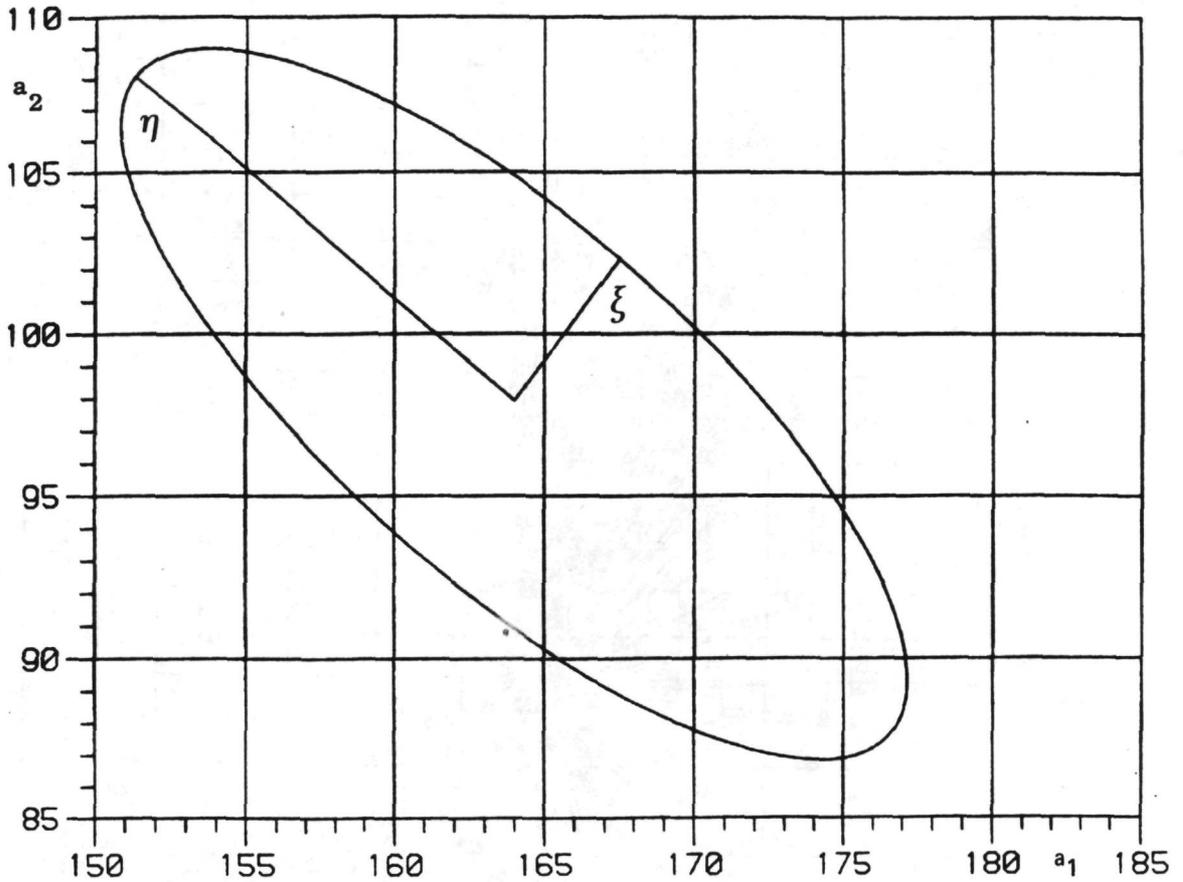
Figure 54   A 95 per cent isocontour based on the asymptotic distribution
of the maximum likelihood parameter estimator for the Gaussian
Double Exponential probability density model
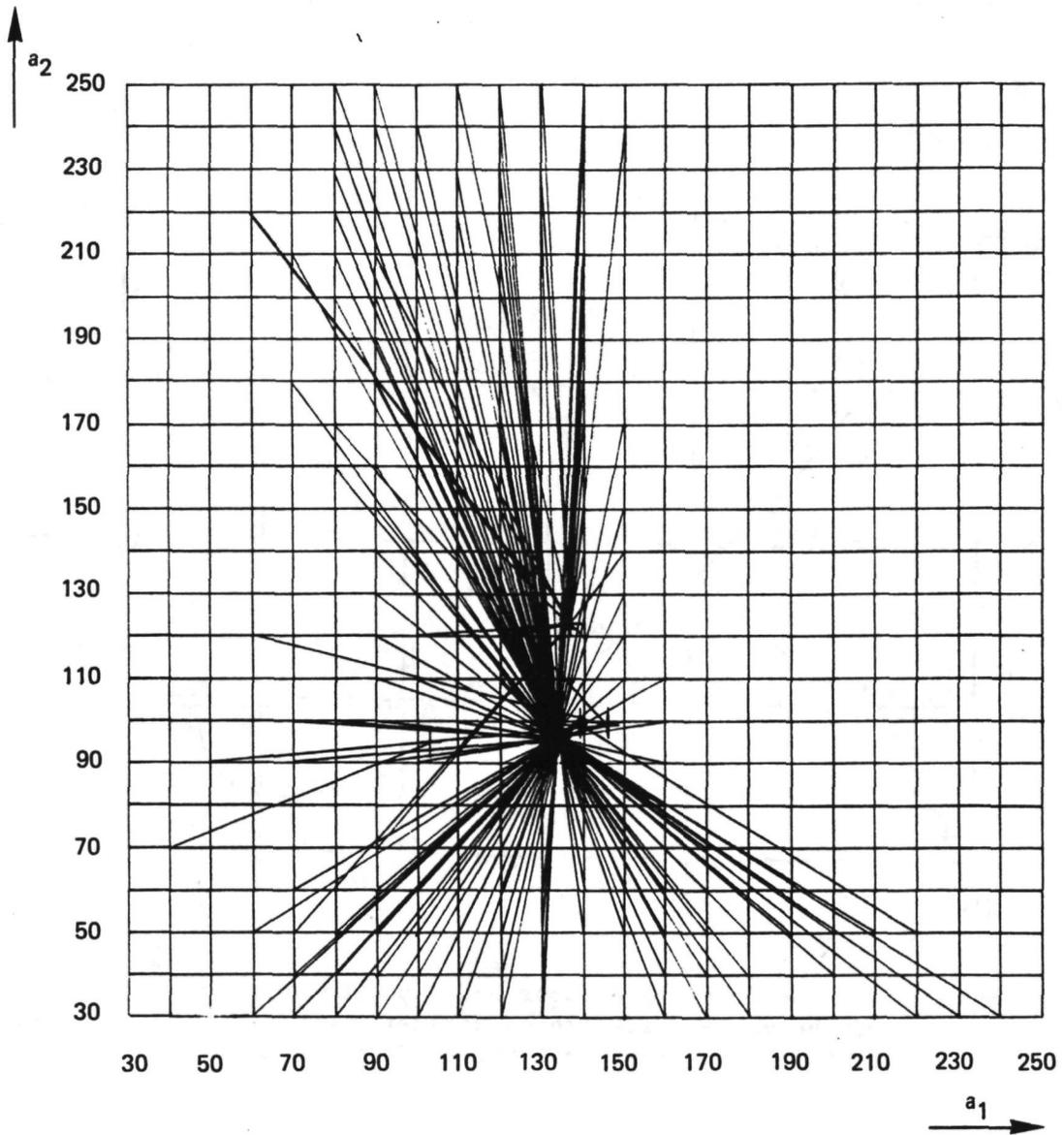
Figure 55   Diagram showing in an (a_1, a_2)-plane which gridpoint (as an initial estimate for the Newton iteration process) converges to which solution of the likelihood equations based on the Generalized Laplace Double Exponential probability density model
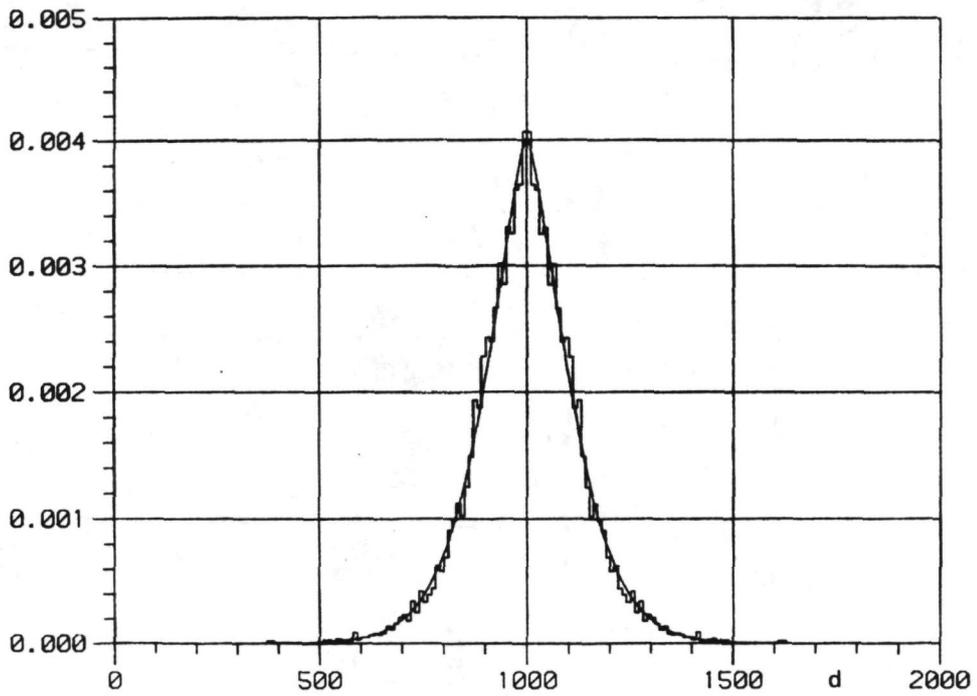
Figure 56   The estimated Generalized Laplace Double Exponential density
with the folded histogram superimposed
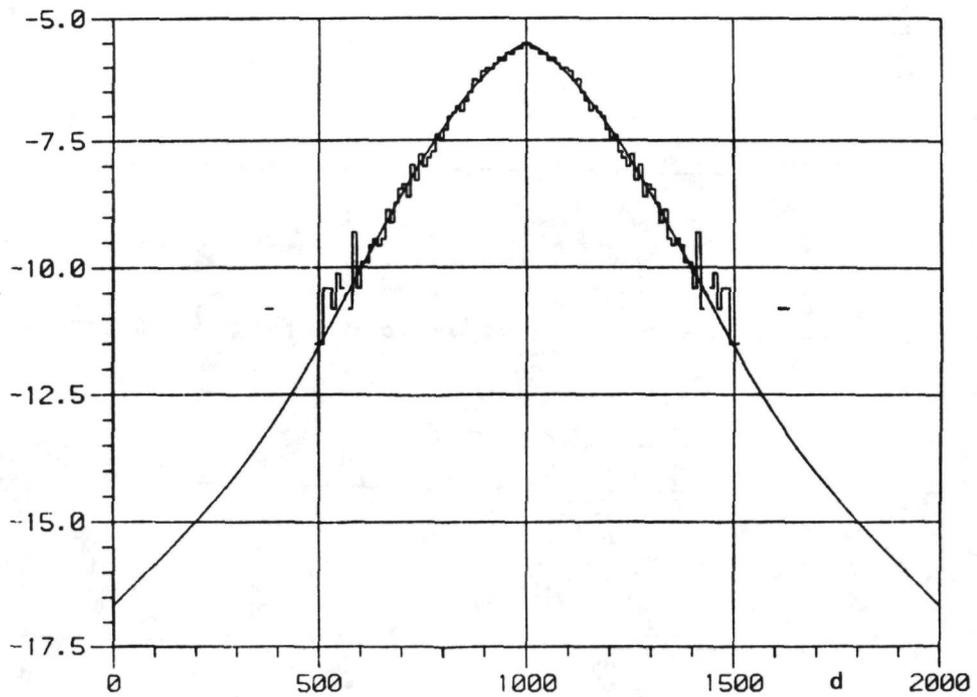
Figure 57   Logarithm (base e) of the estimated Generalized Laplace Double
Exponential density and of the folded histogram
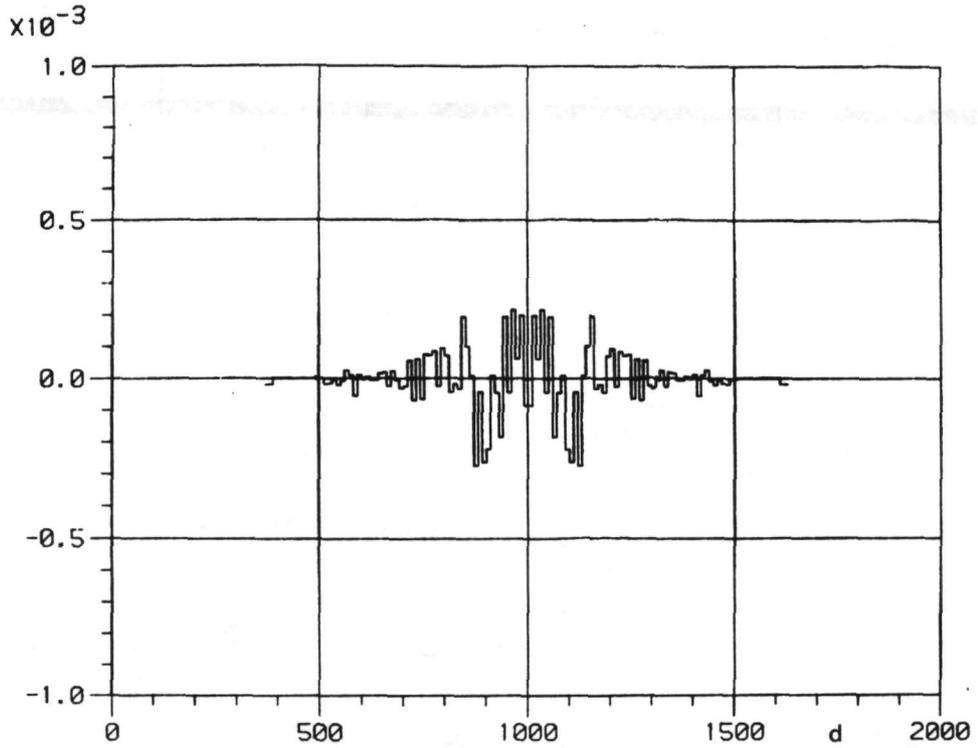
Figure 58 Linear residual plot of the Generalized Laplace Double
Exponential density
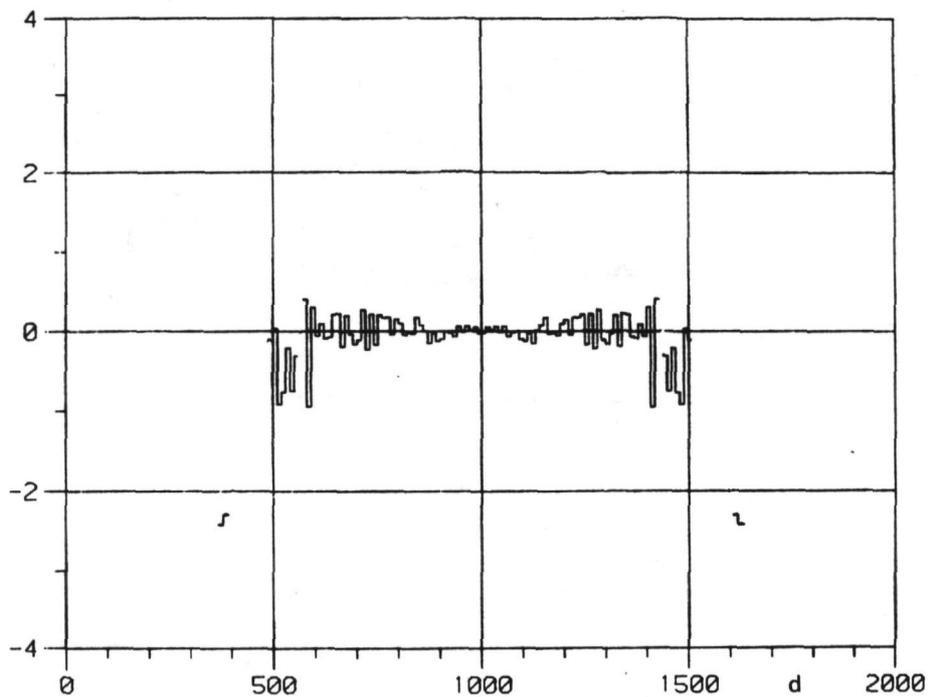
Figure 59 Logarithmic residual plot of the Generalized Laplace Double
Exponential density
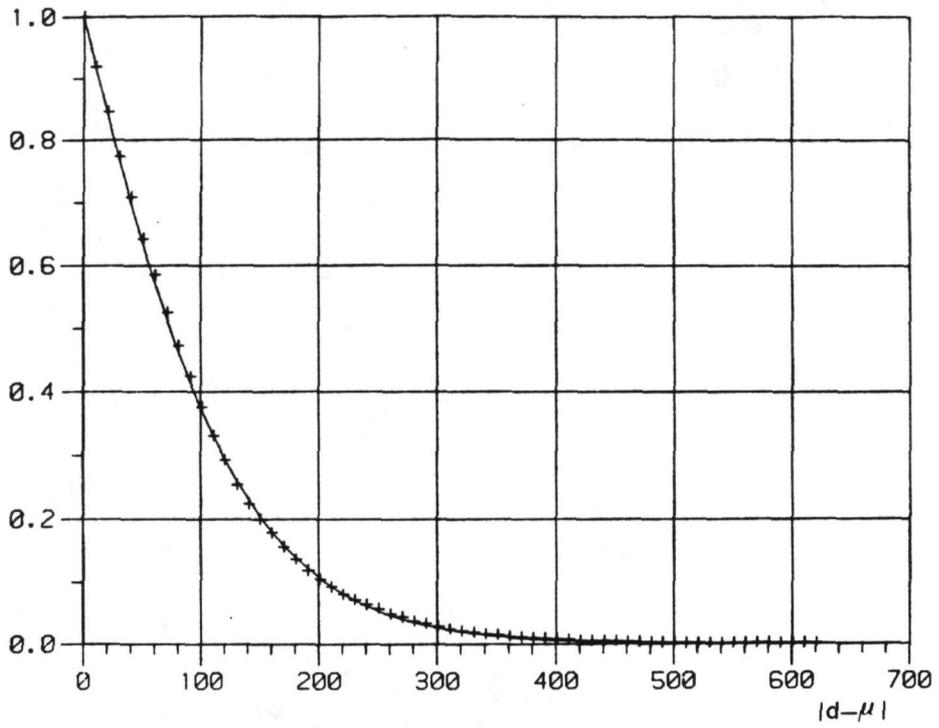
Figure 60   "1-cumulative" curve of the Generalized Laplace Double Exponential
density
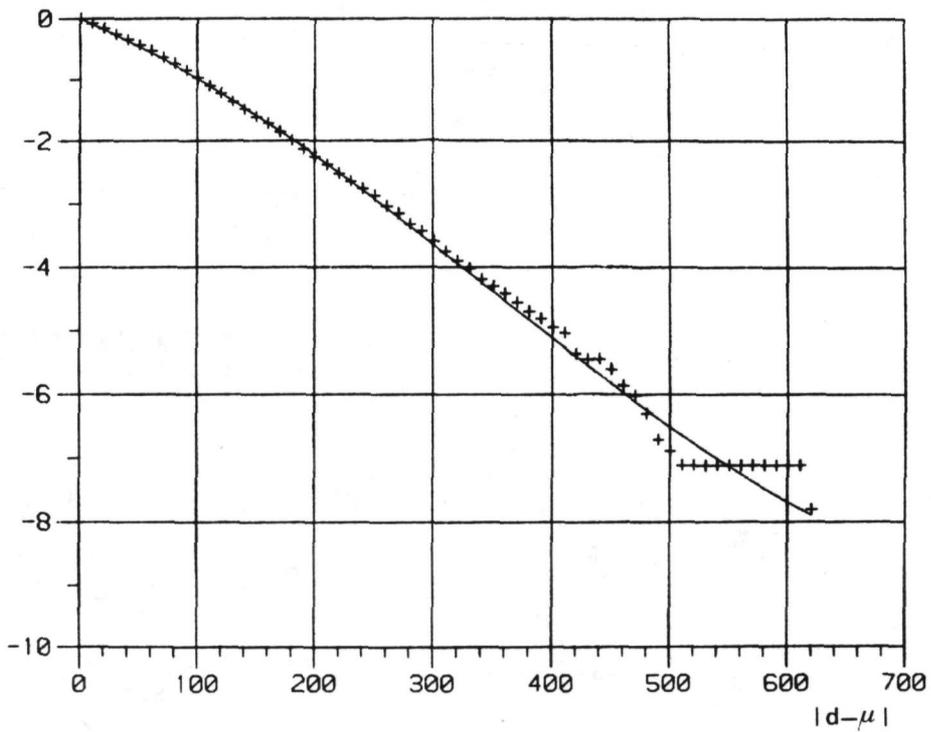
Figure 61   Logarithmic "1-cumulative" curve of the Generalized Laplace Double
Exponential density
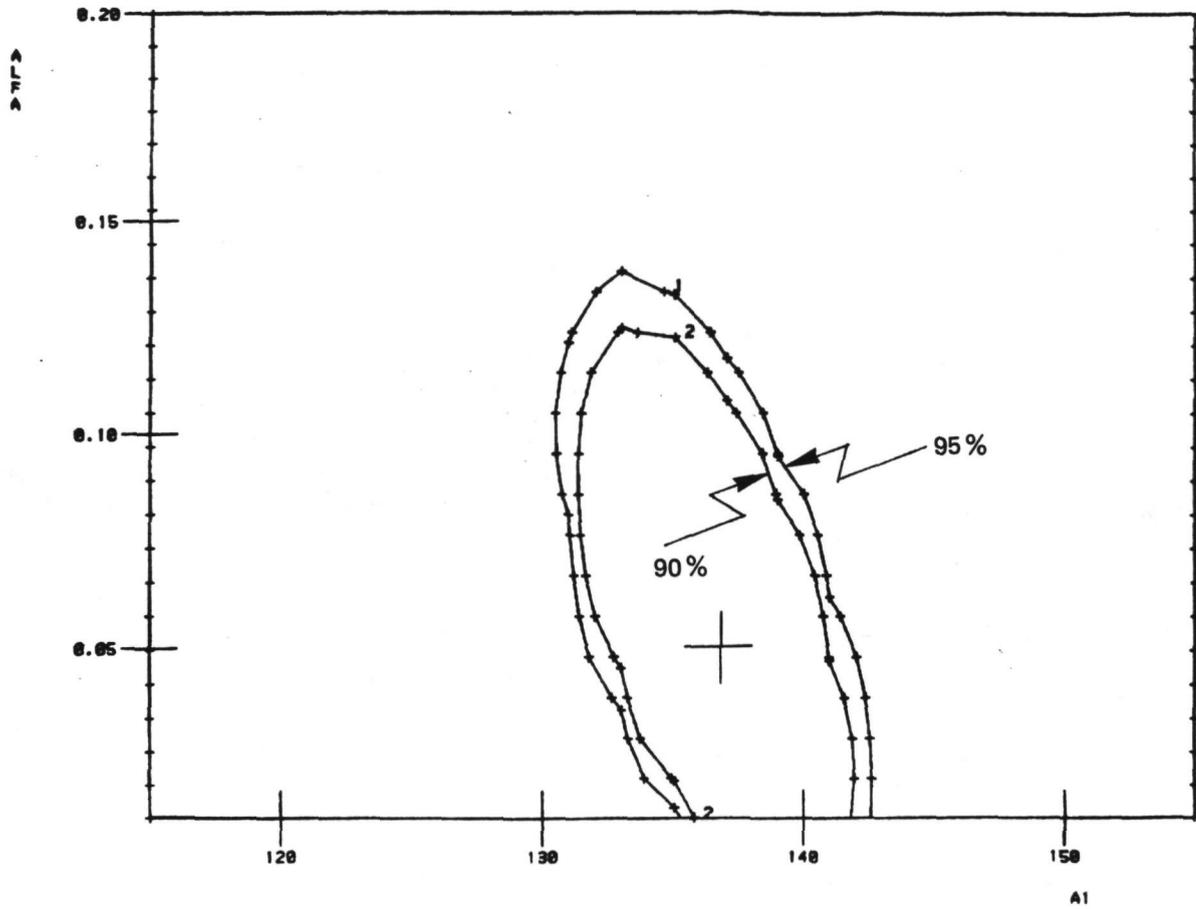
Figure 62   Two isocontours of the relative likelihood function of the
            Generalized Laplace Double Exponential probability density model
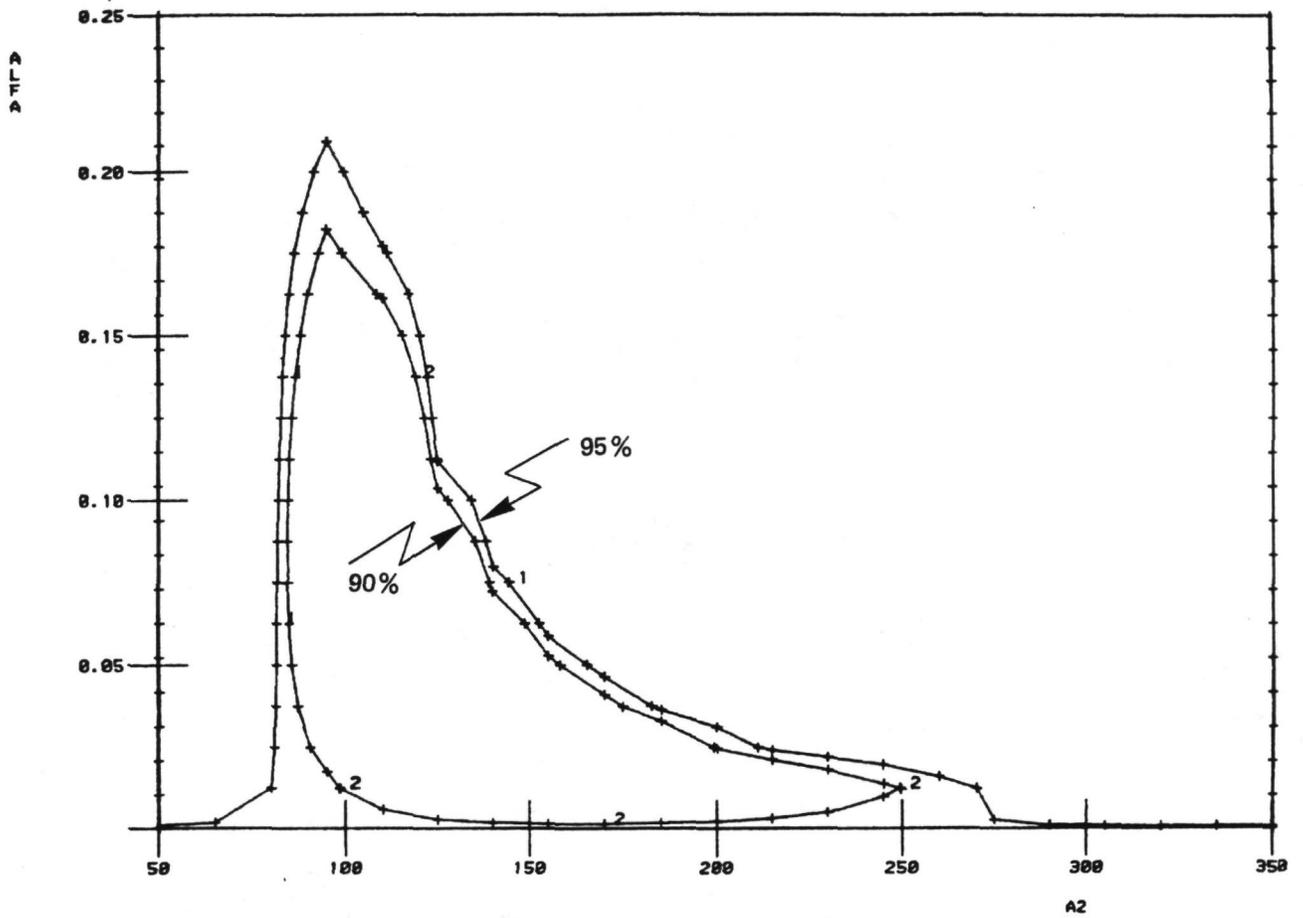
Figure 63   Two isocontours of the relative likelihood function of the
Generalized Laplace Double Exponential probability density model
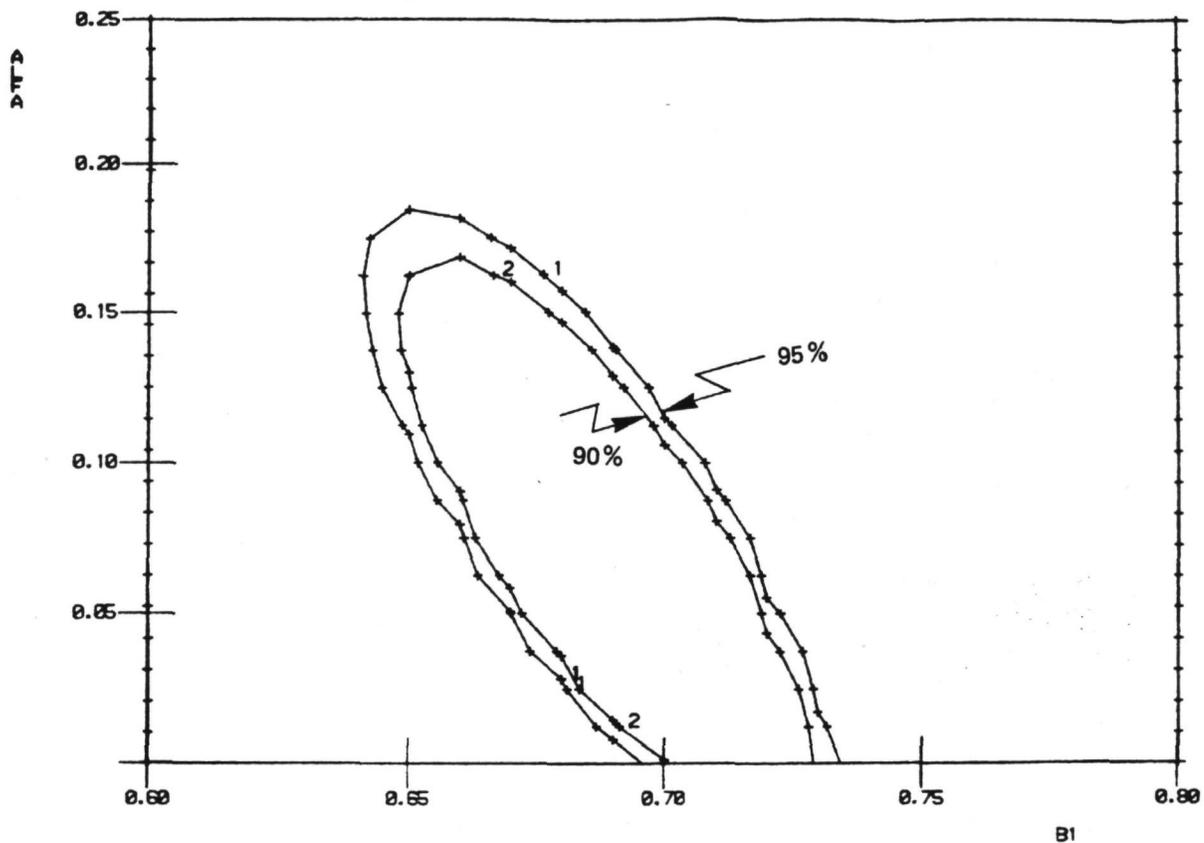
Figure 64   Two isocontours of the relative likelihood function of the
Generalized Laplace Double Exponential probability density model
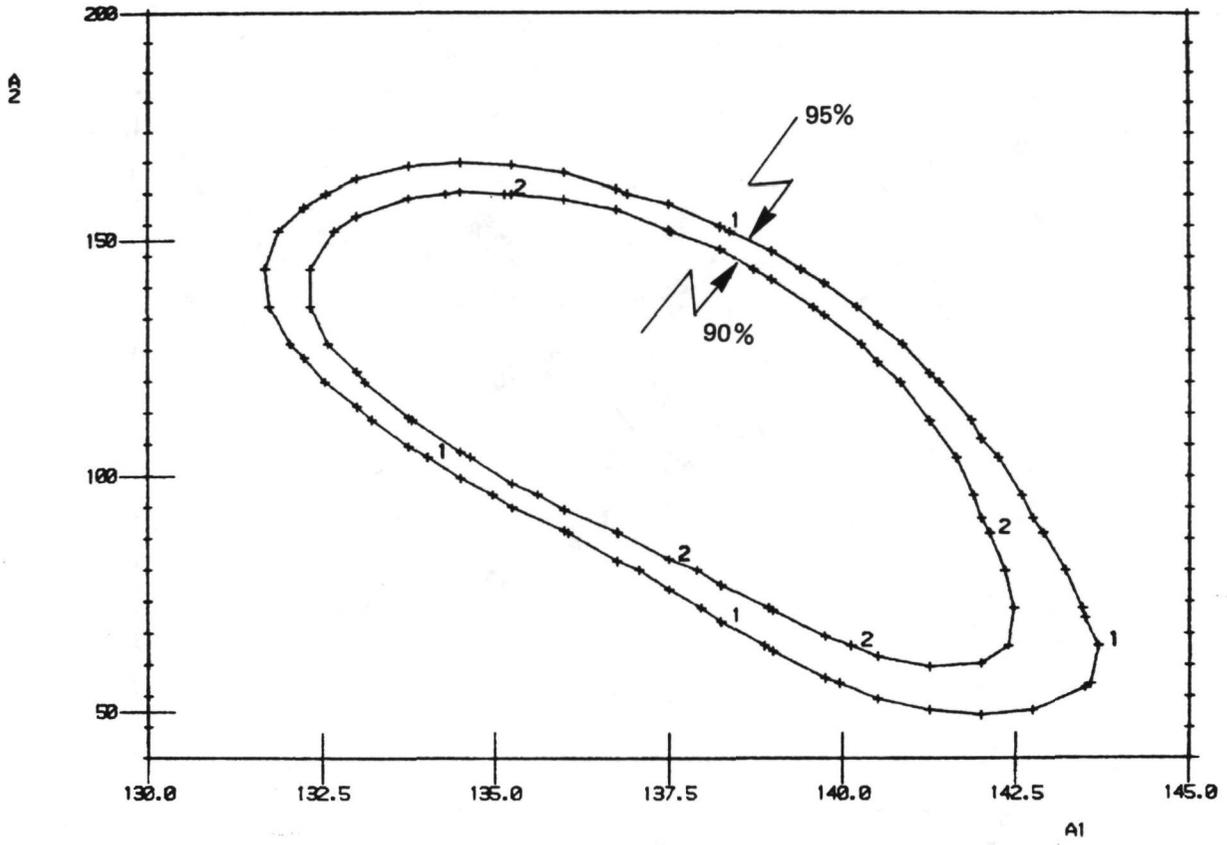
Figure 65   Two isocontours of the relative likelihood function of the
Generalized Laplace Double Exponential probability density model
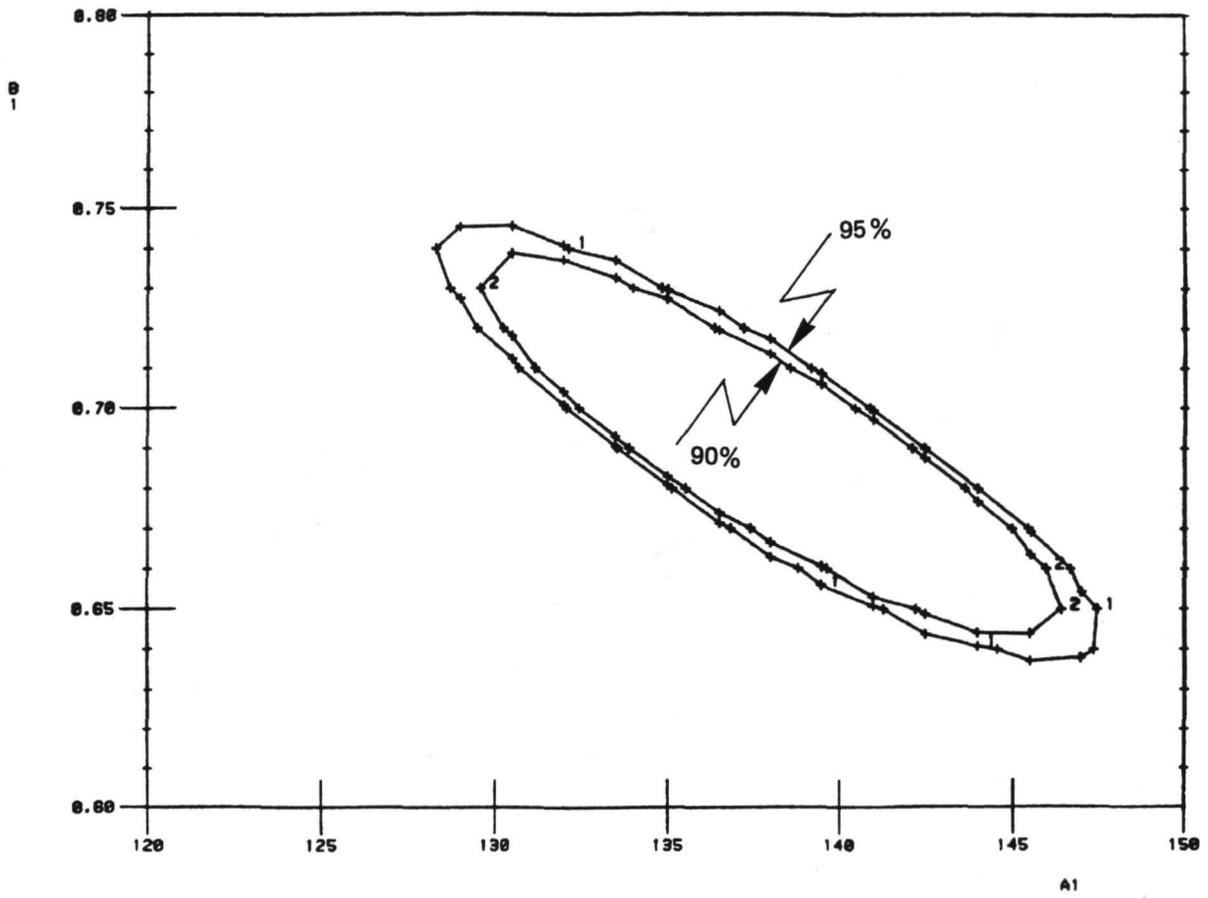
Figure 66   Two isocontours of the relative likelihood function of the
Generalized Laplace Double Exponential probability density model
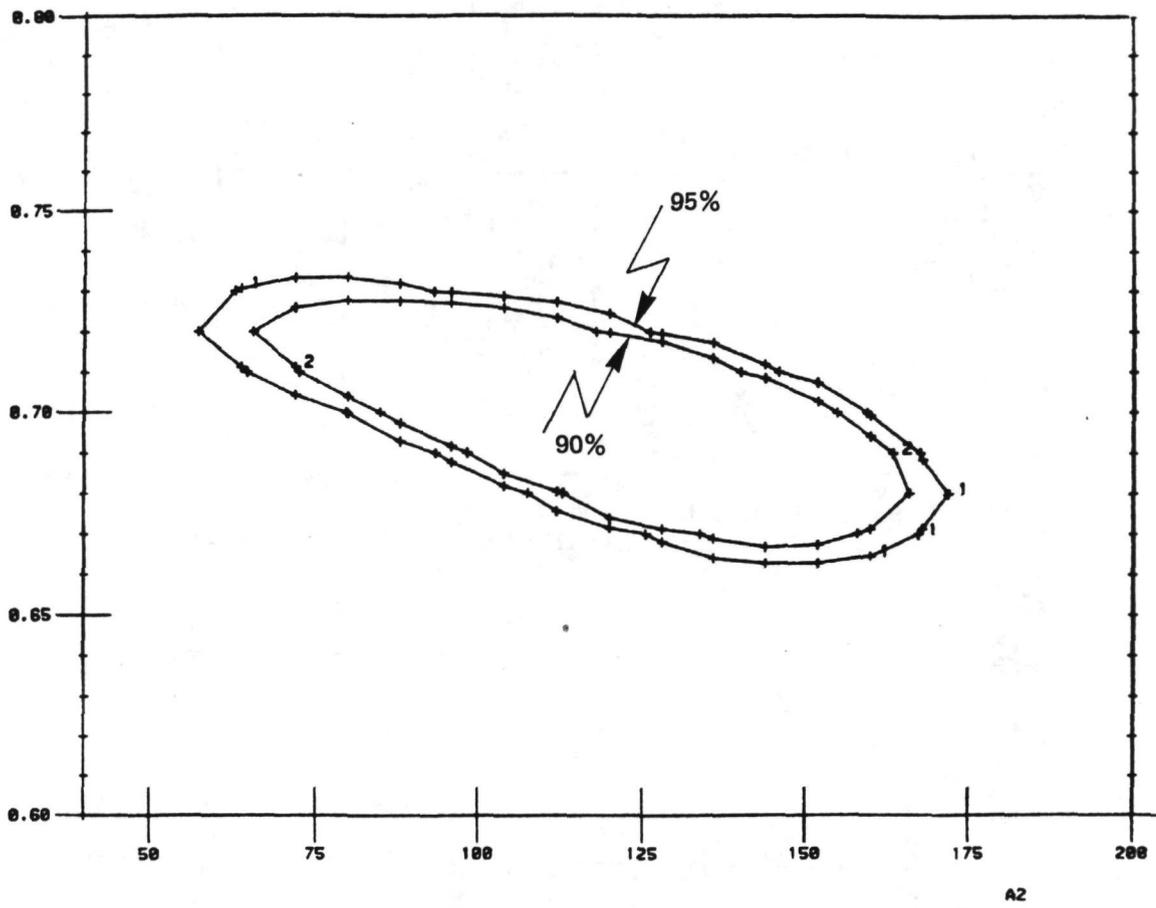
Figure 67 Two isocontours of the relative likelihood function of the Generalized Laplace Double Exponential probability density model
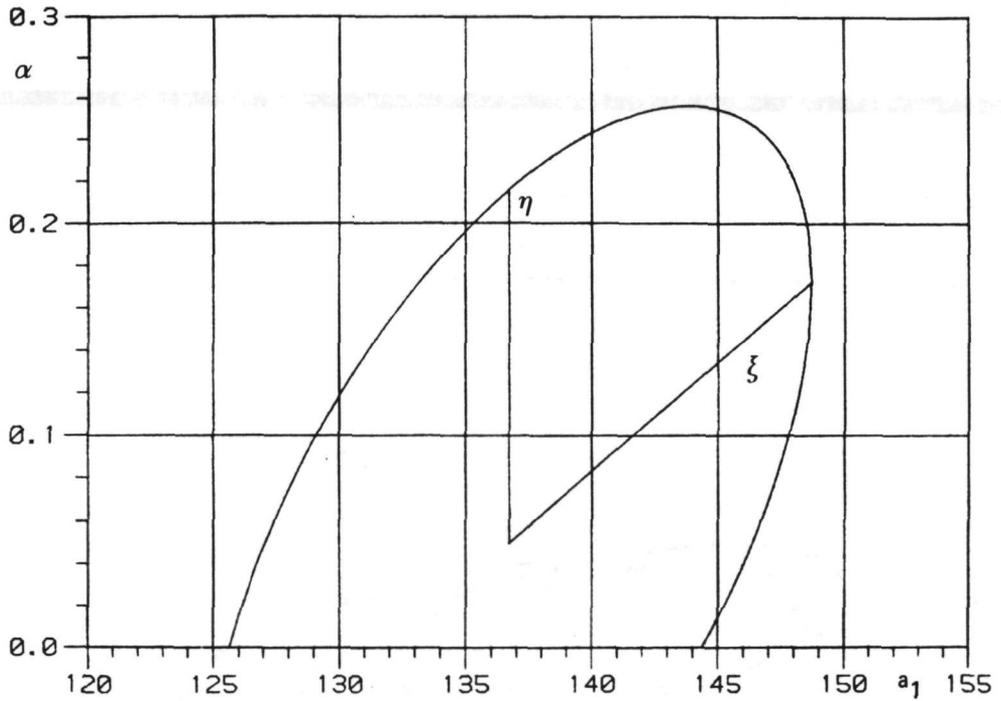
Figure 68   A 95 per cent isocontour based on the asymptotic distribution
of the maximum likelihood parameter estimator for the Generalized
Laplace Double Exponential probability density model
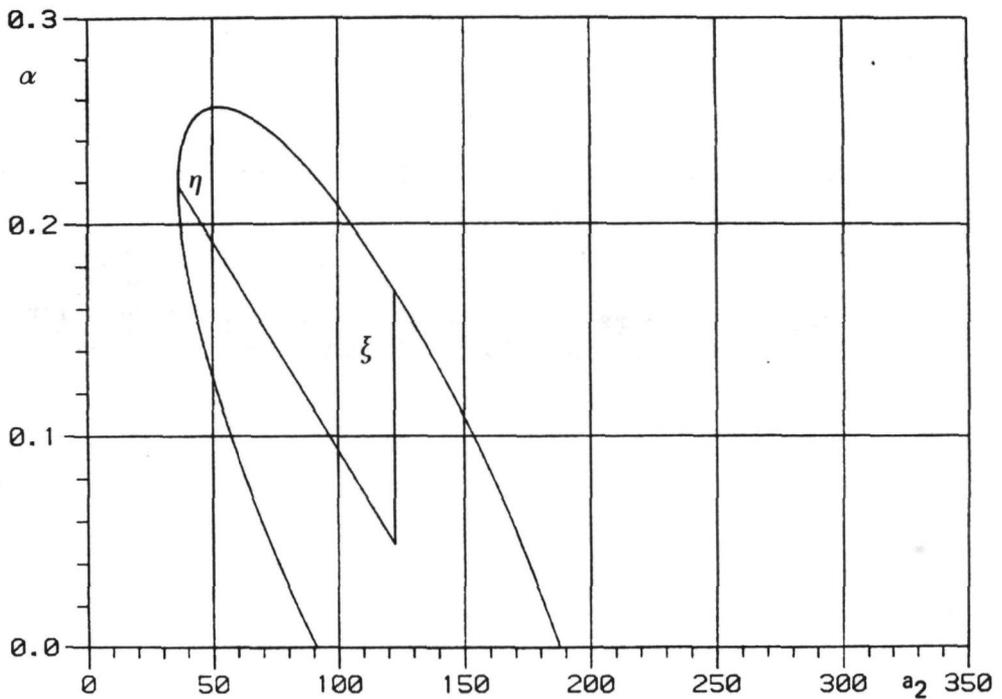
Figure 69   A 95 per cent isocontour based on the asymptotic distribution
of the maximum likelihood parameter estimator for the Generalized
Laplace Double Exponential probability density model
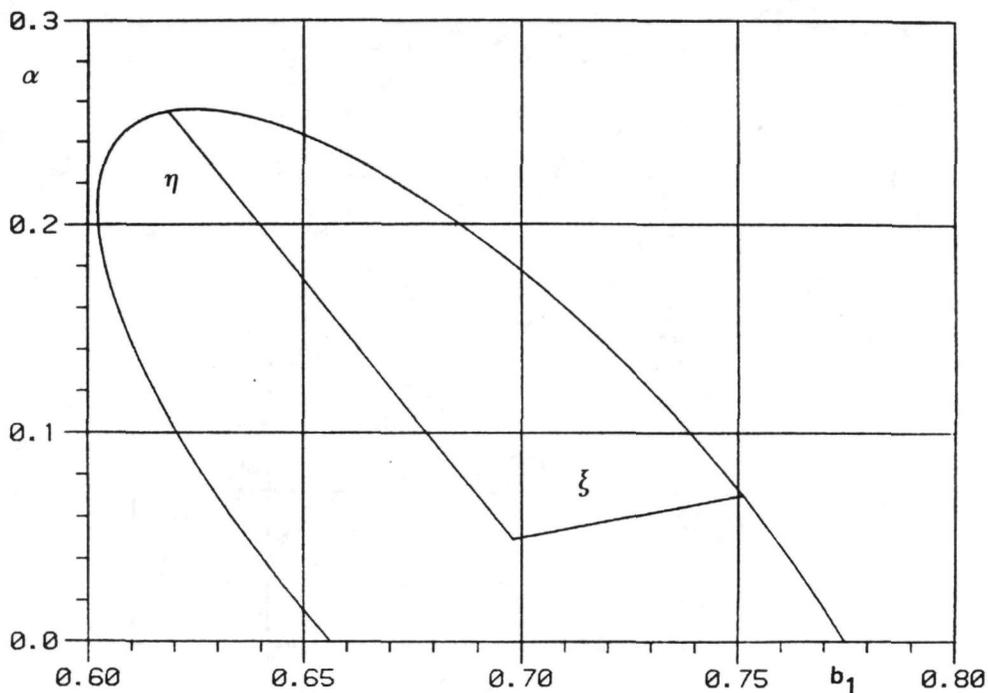
Figure 70   A 95 per cent isocontour based on the asymptotic distribution
of the maximum likelihood parameter estimator for the Generalized
Laplace Double Exponential probability density model
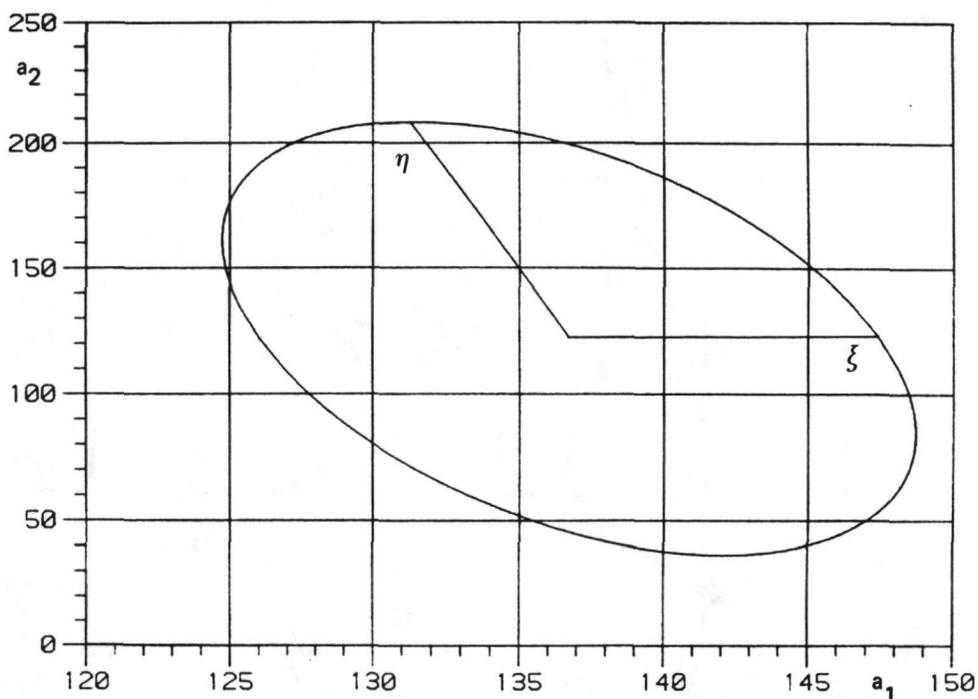
Figure 71   A 95 per cent isocontour based on the asymptotic distribution
of the maximum likelihood parameter estimator for the Generalized
Laplace Double Exponential probability density model
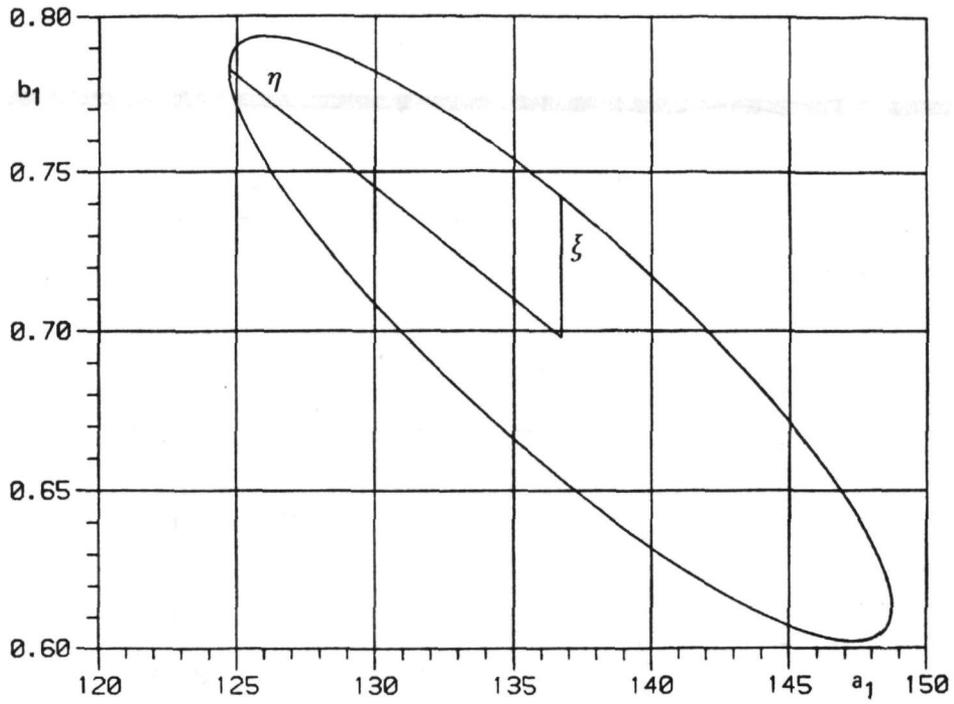
Figure 72 A 95 per cent isocontour based on the asymptotic distribution
of the maximum likelihood parameter estimator for the Generalized
Laplace Double Exponential probability density model
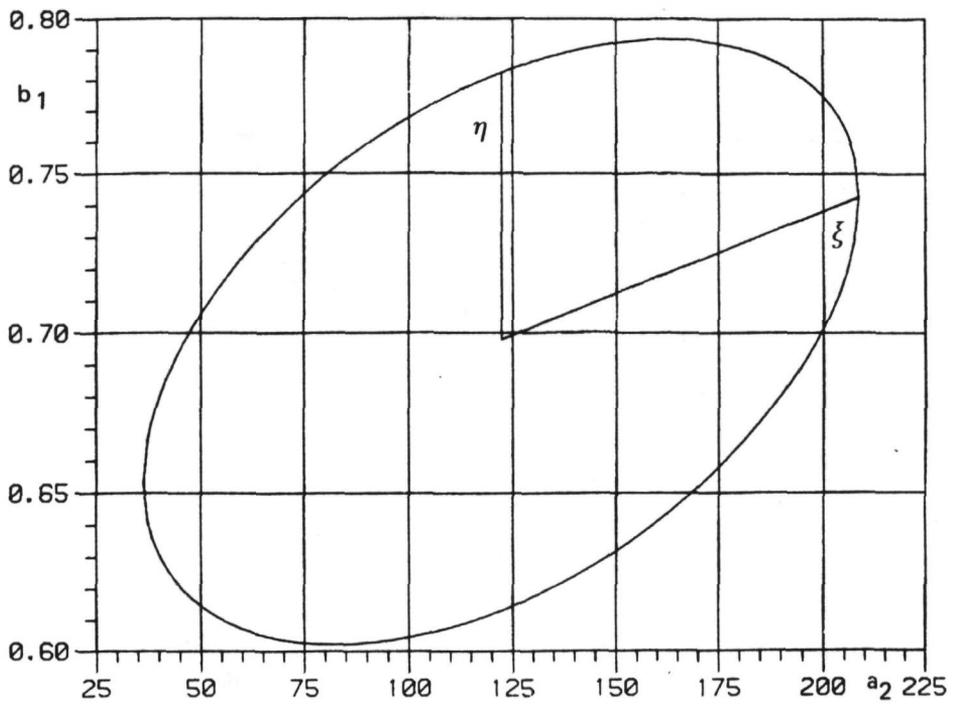
Figure 73 A 95 per cent isocontour based on the asymptotic distribution
of the maximum likelihood parameter estimator for the Generalized
Laplace Double Exponential probability density model
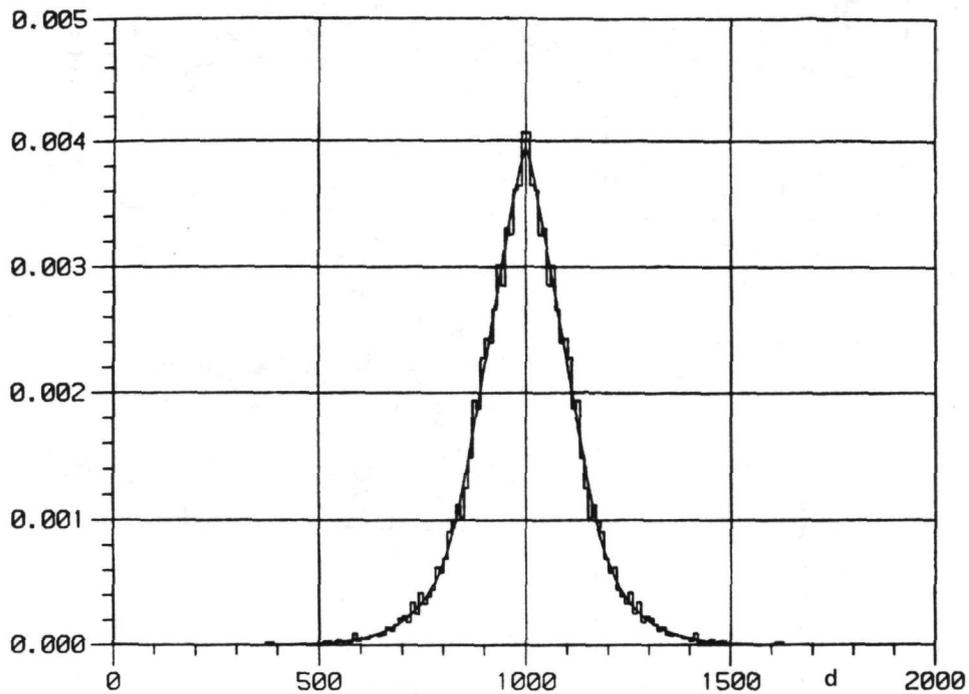
Figure 74   The estimated Double Generalized Laplace density with the folded histogram superimposed
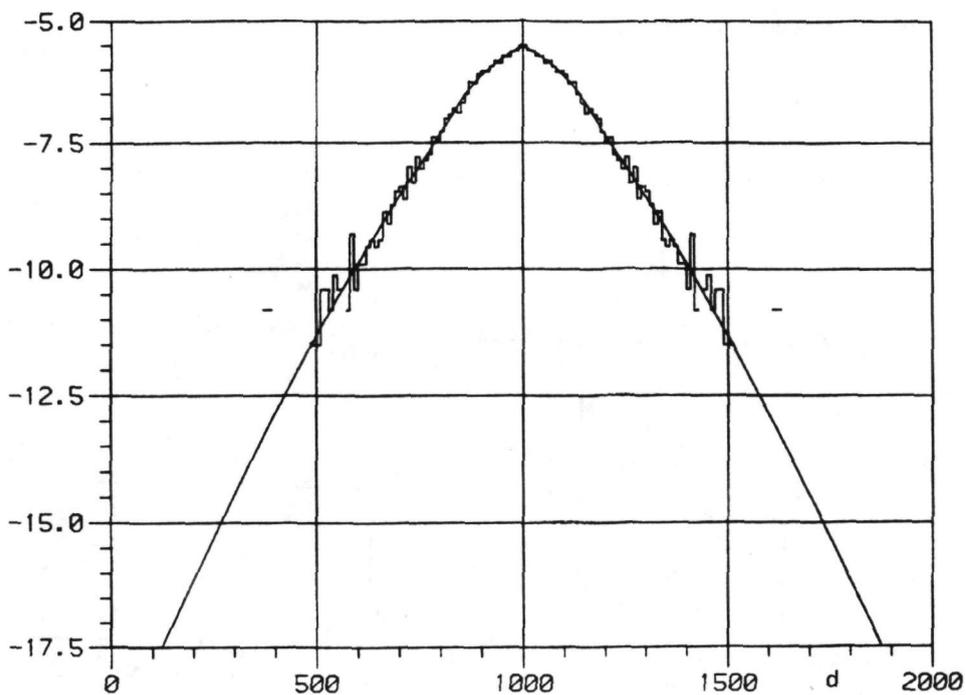
Figure 75   Logarithm (base e) of the estimated Double Generalized Laplace density and of the histogram
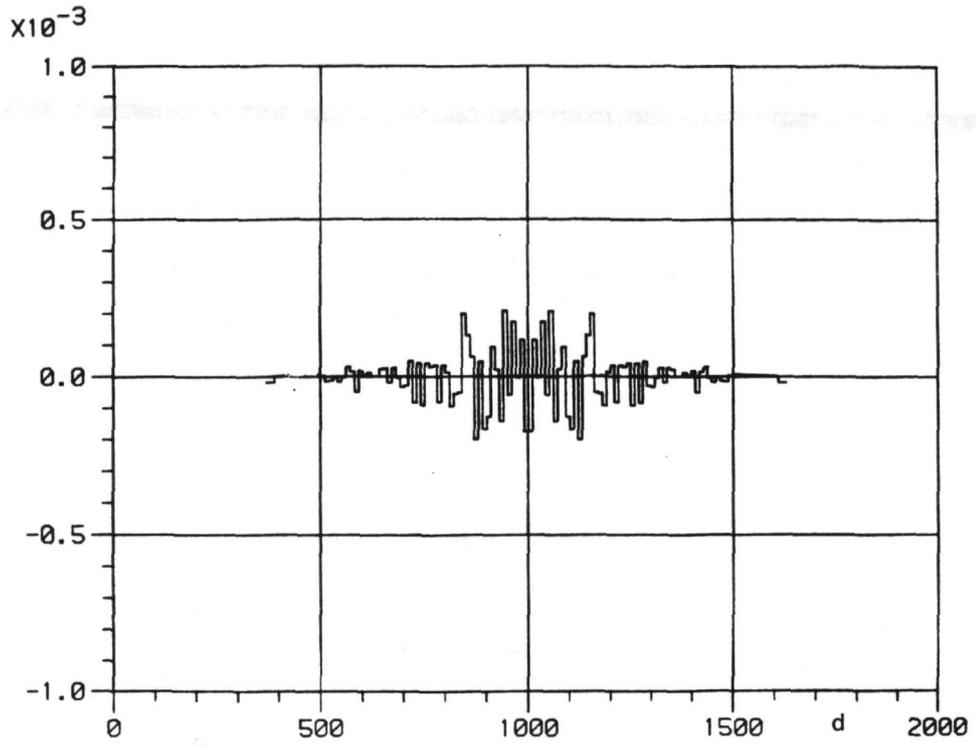
Figure 76 Linear residual plot of the Double Generalized Laplace density
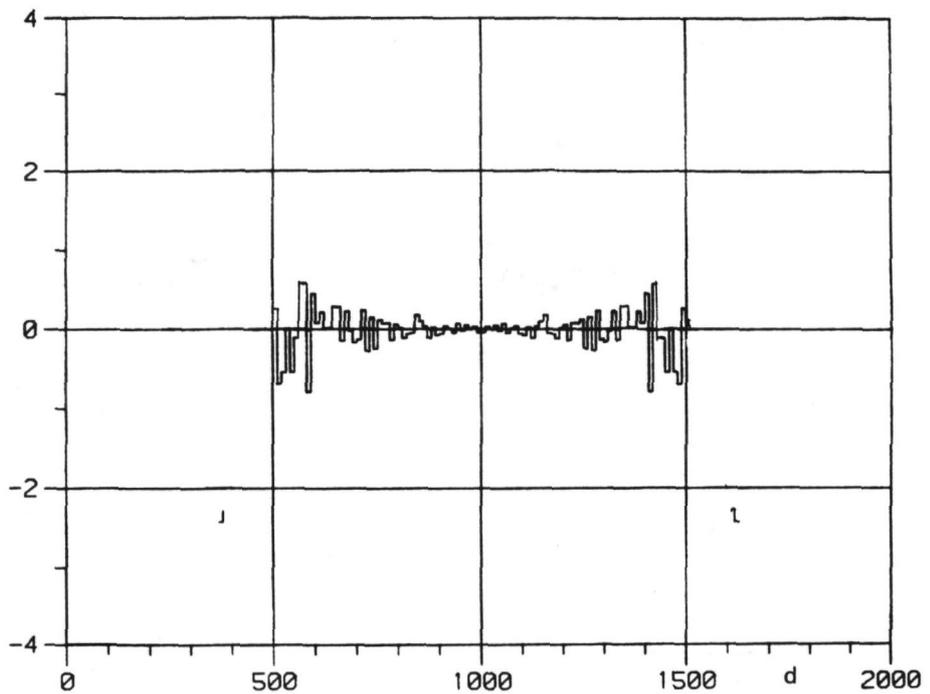
Figure 77 Logarithmic residual plot of the Double Generalized
Laplace density
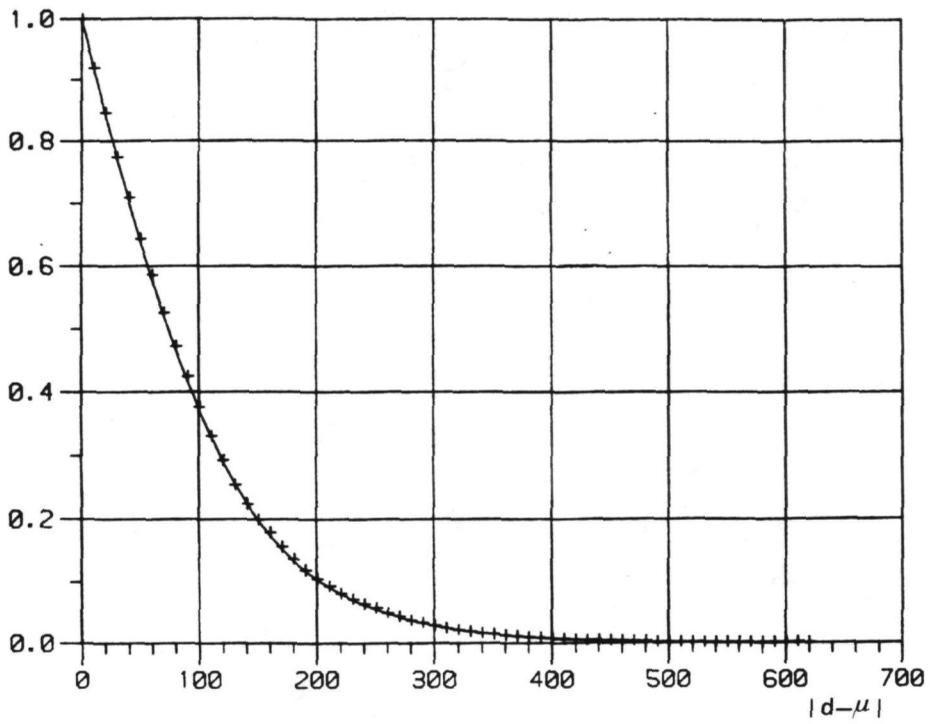
Figure 78 "1-cumulative" curve of the Double Generalized Laplace density
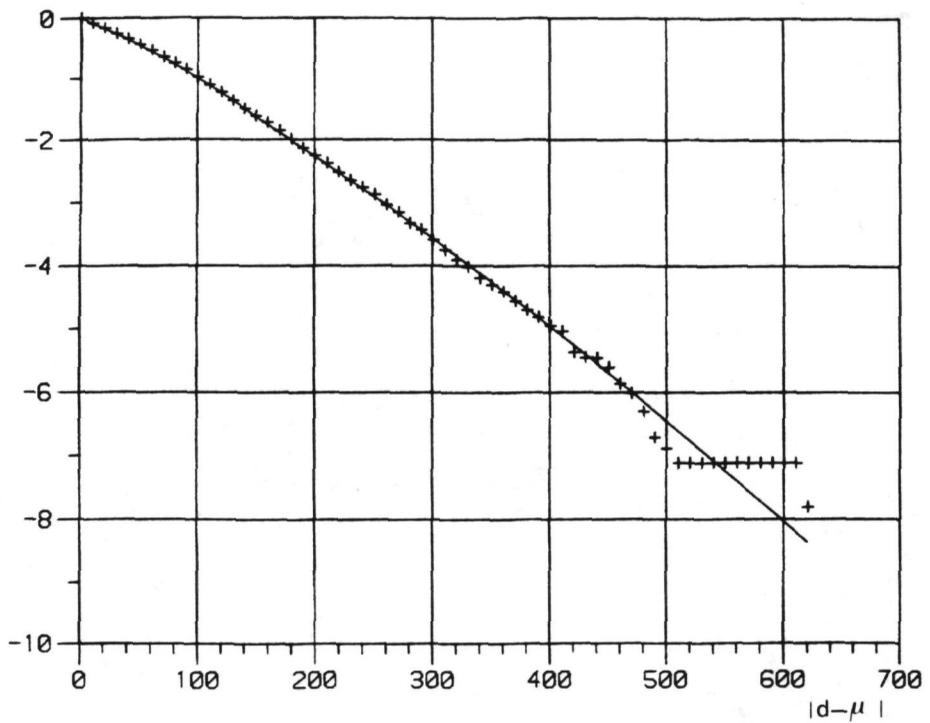
Figure 79 Logarithmic "1-cumulative" curve of the Double Generalized
Laplace density