A Unified Approach to Statistical Estimation and Model Parameterisation in Mass Calibration

by

Thomas S. Leahy B.Sc.

A Thesis presented to

Dublin City University

For the Degree of Doctor of Philosophy

February 1998

School of Physical Sciences

Dublin City University

Ireland

I hereby certify that this material, which I now submit for assessment on the programme of study leading to the award of Doctor of Philosophy is entirely my own work and has not been taken from the work of others save and to the extent that such work has been cited and acknowledged within the text of my work.

Signed: Thomas S. Leahy ID No.: 93701411

Date: 18/9/98

A Unified Approach to Statistical Estimation and Model Parameterisation in Mass Calibration

"Dishonest scales are an abomination to the LORD
but a just weight is his delight"

Proverbs 11:1

Acknowledgements

There are several people and institutions who have contributed much-valued expertise to this research whose help I gratefully acknowledge.

Dr. Michael Gläser of PTB has provided much help and facilitated two training visits to the mass calibration facilities at PTB, Braunschweig. He has provided valuable insights into several aspects of the experimental work and is responsible for pointing out the correlation between low humidity and calculated mass values discussed in Chapter 10.

Dr. Roman Schwartz, also at PTB, gave valuable assistance at various stages. The data used in Example II of Chapter 10 was obtained at PTB with his help.

I also had a valuable discussion with Dr. Klaus Weise and Dr. Wolfgang Wöger at PTB which greatly increased my understanding of Bayesian probability theory and its applications in scientific analysis which has proved invaluable.

Dr. Walter Bich of IMGC has helped by supplying several publications and useful comments which inspired much of the further development reported in this thesis.

Dr. Richard Davis at the BIPM, Paris was also very helpful in arranging time for me to visit the laboratories and in discussing the work carried out there.

Dr. Brian Lawless, my supervisor at DCU, has been instrumental in encouraging the work at all stages and has provided assistance and advice too numerous to mention! Prof. Eugene Kennedy, Head of the School of Physical Sciences at DCU, has also provided significant help, not least in facilitating me carrying out this research work at DCU, and also in reading and proofing previous drafts of the thesis.

The National Metrology Laboratory in Forbairt, Dublin and its staff have been most helpful in making crucial laboratory time available for experimental research and in providing support for training and travel.

Various post-graduate students past and present deserve to be thanked, in particular mention must be made of Steve, with whom I shared an office for two years and who helped to keep me sane, and Mick who unfailingly solved my various computer problems and rescued me from several potential technological tragedies !!

Finally I must acknowledge my parents whose undying support and encouragement throughout the years has brought me thus far today.

Thomas S. Leahy

January1998

Contents

Abstract

This thesis presents a unified and homogenous system of data analysis and parameter estimation. The process is applied to mass determination but the underlying principles are general and the philosophy applies to any data reduction process.

Two main areas are covered: uncertainty analysis via the recommendations of the ISO Guide and secondly parameter estimation of over-determined measurement systems. Application to mass determination of the ISO-recommended procedures and also parameter estimation in mass calibration have been treated previously. What is done here is an innovative attempt to link these two areas together by focusing on the measurement philosophy underlying each and producing a Unified Approach to parameter estimation in mass determination. A unique feature is the application of the ideas of classical probability theory to uncertainty analysis and mass metrology, particular emphasis being placed on employing a consistent and logically coherent analysis. Criteria of consistency from classical probability theory are used as a basis for much of the work, and some useful definitions with respect to subjective information and unbiased analysis are presented which form a useful contribution to the metrology of uncertainty theory.

With respect to parameter estimation techniques novel methods recently proposed in the literature are investigated on a mathematical level and it is shown that the minimum variance estimator used is in fact an application of Bayesian techniques to parameter estimation. This provides a useful link to the ISO Guide on uncertainty analysis, which is mathematically based on a Bayesian view of probability.

The traditional least squares method of parameter estimation which has been previously shown to be internally inconsistent in its view of the reference information, is shown in this work to be incompatible with the ISO Guidelines and the consistency criteria mentioned above. The benefits of applying the Unified Approach are amply seen in the improved estimates and lower covariances achievable with the Bayesian estimators.

The capabilities of Bayesian estimators are explored in some detail with experimental data. This provides some new insight into the estimation technique and discusses how robustly it can deal with inaccurate data and also attempts to quantify the maximum improvement in uncertainty that is achievable through recalibration and sequential estimation with this method.

The conclusion reached is that a Bayesian view of probability, without the restriction of maintaining a separation between random and systematic uncertainties leads to a much improved system of data analysis.

Introduction

To ensure the accurate transmission of measurement information arising from both the calibration of standards and dissemination of units, and also from experimental research, it is of great importance that there should be an accepted method for describing data and measurement uncertainties. Such agreement is critical among Primary Standards Laboratories which play a pivotal role in establishing measurement links between communities.

However, it is not just sufficient that all involved in the dissemination of measurement data use a common approach. It is equally vital that the method used has a sound mathematical basis which is as objective as possible and which is an accurate description of the physical reality being modelled.

In the past there have been many and various data analysis models in use (Dieck (1997) list a few for example). Currently a consensus has formed in the metrology communities around the International Standards Organisation's Guide to the Expression of Uncertainty in Measurement (ISO, 1993) which lays down extensive guidelines for modelling data and calculating uncertainties.(Arri (1996), Bich (1996), (1997), EAL (1997), Fritz (1995), Orford (1996)). This approach requires an accurate parameterisation of the measurement process in a way which includes all input values needed to obtain the desired quantity. The existence of such a mathematical relationship then allows an uncertainty evaluation to proceed in a uniform manner.

The essential feature of this uncertainty analysis lies in treating all uncertainty components *equally*: the functional relationship allows the evaluation of sensitivity coefficients which dictate the contribution each individual term makes to the overall combined uncertainty. For each individual influence quantity, variance and covariance information is needed which requires distributional information on all the terms. Herein has been much controversy since it has been traditionally felt that distributional information can best be obtained via repeated measurements and an examination of relative frequencies. These lead to random uncertainties in the conventional approach. Such uncertainties can be arbitrarily reduced by taking an ever larger sample of measurements from which to estimate the so-called "true value". A systematic uncertainty cannot be analysed in this way in the conventional approach and must be treated as fixed, since it usually arises in data which cannot be subjected to repeated measurements. It is in the combination of these two that many past difficulties have arisen and it is to overcome this problem that the idea of treating all components equally has been proposed.

In this method, Degrees of Belief about a parameter are more important than information based on an examination of Relative Frequencies, the latter being simply a means of contributing information to the former. Hence, based on whatever is known about a parameter, some probability distribution can be assigned to it and variance information obtained.

In the implementation of any experiment, there are three principal aspects: that of designing the experiment and the necessary equipment, performing the experiment and then finally data reduction to extract the required information. In this thesis, our primary interest lies in the last of these areas. To that end, the concepts outlined in the preceding paragraphs are implemented in the specific example of Mass Determination at the level of the Primary Standards Laboratory. We introduce the idea of a *Unified Approach* to data analysis. By 'Unified' is meant a methodology which is internally consistent and follows a given philosophy at all times. The philosophy is based on a Bayesian view of probability and the ISO recommendations.

In mass determination, there are two principal areas: firstly model parameterisation and uncertainty analysis to correctly describe the experimental information (i.e. mass differences resulting from comparison experiments); and secondly parameter estimation to determine optimum values for the parameters (mass values of the standards) based on the information presented by the set of intercomparisons among the standards. The comparisons are usually carried out in an over-determined design scheme so that there is extra information present among the parameters allowing statistical adjustment to be implemented.

It is the goal of the research presented here to unify both of these aspects of the process, so that an overall package is presented having a common philosophy and an internally consistent methodology. Of course it is also desired that the process should be physically justifiable and a valid representation of all the available information.

The first four chapters deal with uncertainty analysis. The theory of the ISO Guide is presented in Chapter 1, introducing the main mathematical and statistical terms needed, and focusing on the implementation of the Guide's uncertainty analysis. A unique feature of this thesis is the explicit application of probability theory as extended logic (Jaynes 1983, 1996) to uncertainty analysis in general and mass calibration in particular. It is in this context that the assumptions and philosophies of the ISO method are discussed and justified, various objections also being considered. A particularly helpful contribution in this area concerns the themes of *unbiased estimates* versus *subjective assessments.* In this regard, the basic Criteria and Logic of Classical Probability Theory are outlined and shown to be suitable desiderata for any

data analysis system. An unbiased analysis is defined as one where the demands of consistent reasoning of probability theory are implemented while the use of subjective information is simply a realistic reflection of the finite knowledge available in any experimental situation. The ISO approach is shown to meet these essential criteria and the Maximum Entropy theory is explored as a robust and useful extension to the Guide. The Criteria of Consistency highlighted here will be continually mentioned throughout the remainder of the work.

Chapters 2 & 3 should be considered together and present the application of the Unified Approach to the model parameterisation of mass comparison data. The general model of Chapter 1 is made specific here as the experimental system and necessary systematic corrections are described. It should be noted that the equations presented here reflect the system in use in the laboratory where the experimental work was carried out and is thus specific to that situation. In another laboratory, with other instrumentation, the model would perhaps be different, but the method can be adapted to deal with any physical situation by appropriate inclusion of all known influence quantities. Chapter 2 deals with the mass comparison process and develops a scalar version of the Weighing Equation—the fundamental relationship for determining the mass difference terms. Chapter 3 deals with the evaluation of air density. The evaluation of the well-accepted approximate relation for air density in a Standards Laboratory is presented and the general error propagation theory of the ISO Guide is applied to evaluate its standard uncertainty.

This is an example of the consistent approach to data analysis being emphasised in this thesis: the air density equation has of course been tackled many times before, but in the majority of cases the uncertainty analysis is presented with random and systematic components treated differently. Here, however, we maintain a simpler uniform approach and show the power of the general error propagation theory (often called the Gaussian Procedure) in presenting data in a coherent manner. This allows us to combine all the influence quantities into a single relation to produce the overall combined standard uncertainty of the mass difference term.

Chapter 4 tackles the model parameterisation from a multivariate position. This lays the foundation for the parameter estimation techniques discussed in Chapters 5 to 8. We show how the Weighing Equation is developed in matrix notation and how the uncertainty analysis of the Gaussian Procedure is developed in this multidimensional case. We are careful to point out here how the variances and standard uncertainties are assigned to the measurand estimates and not to the unknown errors or contingencies affecting the experiment. The equations presented here provide another vindication of

the Unified Approach owing to their simplicity and conciseness while nevertheless providing a complete treatment of all the relevant data.

One aspect of particular interest is the inclusion of the uncertainty due to the systematic buoyancy correction in the analysis. In many treatments this uncertainty term is either neglected or included after the parameter estimation has been carried out. In the approach presented here this need no longer happen since it is very simple to include all uncertainty information in the analysis. The result is an observation vector and covariance matrix which completely describes all the available information from the comparison experiment.

Chapters 5 to 8 tackle the second major aspect of data analysis in mass determination, that of parameter estimation by statistical adjustment. The Unified Approach developed in the first four chapters is continued. There are two sets of information to be combined: the experimental information determined in the comparison exercise and any previously known parameter values from other calibrations of the standards. We find that the conventionally applied approach is inconsistent in its use of this prior information while the Unified Approach allows extra benefits not otherwise possible.

Chapter 5 considers the Least Squares estimation method, subject to constraints needed to obtain a particular solution (Restrained Least Squares). The inadequacies of the method are highlighted in its use of the constraint information: the previously determined estimates of the constraints are considered as deterministic constants to obtain a solution, while being treated as stochastic quantities to obtain the proper covariance matrix of the parameter estimates. This approach can perhaps be justified in terms of the conventional method of separating random and systematic uncertainties but is not acceptable in a Unified Approach to data analysis. The critique of Restrained Least Squares in this chapter provides a crucial link with earlier chapters where the criteria of consistent analysis are discussed.

We then note the fundamental distinction that is made by simply treating the constraint information as prior data having its own covariance matrix. This can be augmented with the current information to produce a data set which can be easily estimated using the Least Squares Criterion, or the Gauss-Markov Theorem. Very interesting results are produced by this estimator, wherein a smaller uncertainty is assigned to the prior information and its values are adjusted too. This, while counter-intuitive to the concept of a fixed standard, is entirely justified if the standard value is considered as an estimate with given degrees of belief attached. The comparison

exercise supplies extra information and the minimum variance characteristics of the estimator result in a smaller covariance matrix for the parameter estimates.

Chapters 6 & 7 implement a generalised parameter estimation technique to explore the relationship between deterministic and stochastic constraint information. It is shown how a deterministic view of the constraints leads to the same results as does Least Squares with restraints while a stochastic view leads to an identical solution as the augmented design of Chapter 5. This development is important as it shows from a theoretical basis why the two methods give different results and underlines the importance of properly understanding the nature of all the information used in the data analysis process. The Unified Approach requires all data to be treated equally and we show how advantageous this is in parameter estimation since better estimates can be obtained for the parameters. Of course the approach remains valid if some of the information is deterministic since then it has a null covariance matrix and the generalised estimation technique discussed in these two chapters will deal adequately with it, although adjusted parameter values and smaller variances and covariances for the deterministic information will not be possible in such cases.

Chapter 8 introduces a new perspective by implementing a Maximum a Posteriori estimator which uses the Maximum Likelihood criterion along with Bayes' theorem. This method is ideally suited to parameter estimation in mass calibration since it views the measurements as simply updating the prior knowledge on the parameters. Hence all information is stochastic and once again the posterior estimates show smaller variances and updated parameter values. The technique is very flexible and can easily deal with any situation. For example, new standards with no previous calibration history can be assigned very large or infinite variance in the prior information and the estimation method will then update this information based upon whatever is learned from the comparison experiment. We show how in most cases this estimator will produce the same parameter values as does the augmented design or the generalised estimator. This chapter once again returns to a consideration of the nature of probability as discussed in Chapter 1. We point out that all probabilities are in some way dependent on background information and that realising this permits a more logical analysis of the data. The Bayesian technique is therefore the preferable approach to use in order to illustrate the Unified Approach and so concludes our investigation of parameter estimation techniques.

In Chapters 9 and 10 we present experimental case studies to illustrate the Unified Approach. Three main examples are given and analysis is carried out by both the Classical and Bayesian Unified Approach. While such comparative examples have

been published in the literature before, what is significant here is the detailed examination of the performance of the estimators, in particular the Bayesian one.

We illustrate the inadequacies of the Classical Estimators in dealing with "suspect" or incorrect data and explore in detail how the Bayesian Estimator performs much better under such circumstances. The dependency on the relative accuracy of the prior and current information are among the features highlighted. We find that the estimator behaves as we would expect with more accurate information exerting a greater influence on the result. Thus if incorrect information is assigned a high degree of belief, it will adversely influence the results; however, we find that poor agreement with either prior or current information will highlight the existence of a problem. On the other hand we show how the Bayesian estimator can easily adjust incorrect information if it is assigned a low degree of belief. The robustness of the estimator is thus highlighted in terms of the stability of the output in the face of perturbations in the prior information, or initial conditions.

Having noted the adjustments to the prior data in the posterior estimates, consideration is given to the range of adjustment—in particular to the variances— that is possible for the given prior information (initial conditions). We show that there are theoretical limits, both upper and lower, and that specifically with regard to the lower limit, a numerical technique is required to approach it. Such a technique is adopted by means of scaling the accuracy of the current information over a wide range, and graphically presenting the results for the parameters of interest.

This shows how a lower limit *is* approached for the posterior variances, the ultimate improvement corresponding to the current information having zero uncertainty in a single trial. This of course does not occur in practice but gives us an estimate of what will be the best improvement possible and we can compare any given improvement in accuracy with this benchmark. In sequential estimation, the posterior estimate forms the new prior data for another calibration, later in time. We point out how, with the same design scheme and standards, the subsequent posterior variance estimates will tend to converge to a lower limit, below which no further improvement in accuracy would be possible without introducing additional external information.

We also find that the upper limit or 'worst case' corresponds to no change to the prior covariance matrix and would occur in the limit of 'infinitely *inaccurate*' current information. Thus we see that the estimator has the capability to add new stochastic information, learned in the comparison exercise, without adding "noise" or uncertainty to the final outcome.

A lot of effort is expended in these two chapters discussing the influence of the prior information. This is quite in order since in the Unified Approach the prior information plays an important role, moreso than in the Least Squares methods. We consider the problem of drift on mass standards in Chapter 10, which would have a significant effect upon the prior information, showing how the estimator can deal with this providing there is some accurate information present in the experiment. In other words, if all of the standards have been subject to drift it is not possible to rescue the situation. In this way we are being physically realistic about the capabilities and limitations of the Bayesian estimator, reminding us that any mathematical method is always limited by the information supplied by the analyst. A solution may be mathematically possible but may not be physically meaningful.

In this regard we once again invoke the criteria of consistency in analysis, pointing out that all the relevant information with respect to the standards involved in a comparison exercise must be considered in order to accurately model any potential drift. In this was we provide a useful link with the starting position of Chapter 1. We also emphasise the importance of relative accuracy among the various sets of information: 'suspect' prior information can be assigned a low degree of belief and then in the posterior estimate its variance will be greatly reduced, while its assigned value will only be adjusted if the available evidence demands it.

Thus a useful quantity of new information about the performance of the Bayesian estimator is presented in these two chapters, helping to confirm that it is indeed a robust and reliable means of treating over-determined calibration problems of this nature. The philosophy of including all known information in the analysis, and not just some of it, is vindicated, in agreement with what we would expect on the basis that better conclusions and decisions can be made with full information rather than partial information.

Chapter 11 is the final chapter and gives a short description of the experimental system used to obtain the data discussed in Chapters 9 & 10. A computerised measurement system was implemented to gather data from the automated mass comparators. The procedure used is described and some of the software is discussed. The model parameterisation of Chapters 2 & 3 was developed for the system described here. Some example data graphs are included to illustrate the kind of analysis that was carried out.

****

1. Modeling & Parameterising Experimental Data

1.0 Summary

This opening chapter develops a methodology for analysing experimental data in order to present results and uncertainties in a coherent manner. The philosophy of the International Standards Organisation's "Guide to the Expression of Uncertainty in Measurement" (ISO, 1993) is followed, taking an essentially Bayesian view of probability and implementing the General Law of Error Propagation, often called the 'Gaussian Procedure'.

One of the key points that will be noticed is the uniform manner in which all influence quantities are processed: there is no mathematical distinction allowed between "random" and "systematic" uncertainties. The justification for doing this is presented, by showing how we seek to implement the criteria of consistency underlying classical probability theory and pointing out the importance of viewing probability in terms of Degrees of Belief about an event, or parameter, rather than basing it on Relative Frequencies observed in some experiment or trial. However, such experimental information is nevertheless often an important means of gaining additional knowledge about the parameters. As a result of this, it is necessary, in this method, to establish a distribution function for each influence quantity based on whatever information is available at the time. The Principle of Maximum Entropy is discussed in this context, showing how it allows an unbiased estimate to be obtained from subjective information. By this we mean simply using all the given information, without assuming anything else, in a consistent and logical manner. We point out how the Maximum Entropy theory predicts the two most commonly used distributions in uncertainty analysis : the Uniform and Normal Distributions.

With this it is possible to obtain the variance/covariance information about the influence quantities needed for uncertainty analysis. It will be noted that the distribution information is considered in relation to the parameters themselves and not the unknown random errors responsible for creating the distribution of values observed. From the mathematical functional relationship among the input quantities which generates the parameter of ultimate interest, it is then possible to establish the contribution of each term to the overall uncertainty of the final parameter value. Fig. (1.0.1) overleaf gives a 'flow-diagram' for the analysis process that is used. The various terms are explained in detail within the chapter.

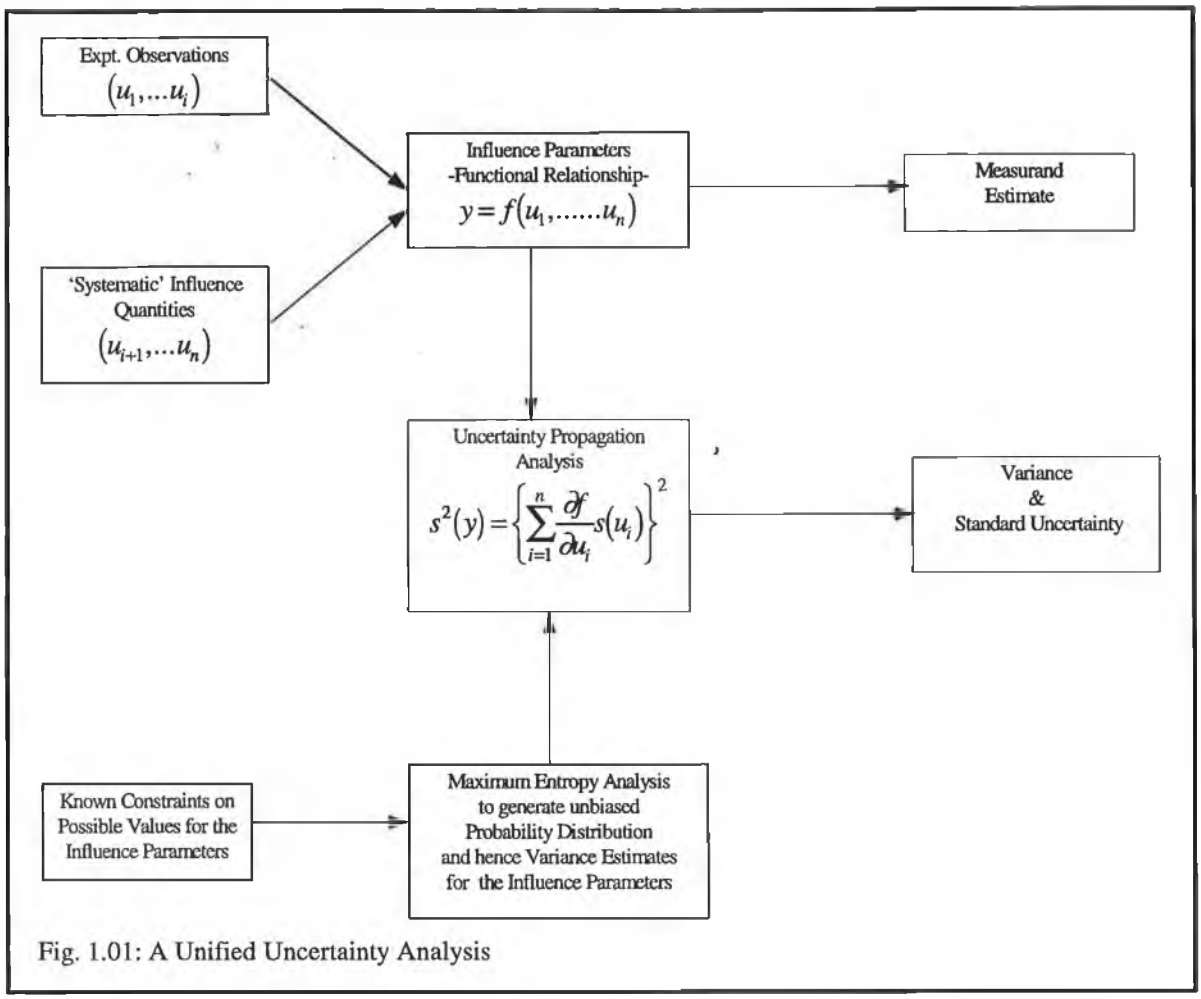


Fig. 1.01: A Unified Uncertainty Analysis

1.1 Introduction

When we wish to investigate a physical process/phenomenon by experimental means it is essential to adequately describe the situation under investigation. In other words, a mathematical model is needed which includes *all* influence quantities affecting the output or result. This should comprise both directly measured quantities, indirect quantities such as data from tables/published information etc. and also any systematic effects which must be included. To take a simple example, the measured length of an object at temperature t is related to a standard length at temperature t_{std} and the thermal expansion coefficient α by:

$$L_{meas} = L_{std}(1+\alpha(t-t_{std}))$$

or, for example, $P_t = I^2 R_0(1+\alpha(t-t_0))$for the power dissipated at temperature t by a current I flowing in a resistor whose resistance is known to be R_0 at temperature t_0. In essence, what is required is a *Functional Relationship* among the influence quantities, which generates the required output quantity. That is

$$y = f(x_1, x_2, ..., x_n)$$

1.2 Terminology

At this point we should pause to consider what we really mean by "quantities", "values", "measurements" etc., since if we do not define our terms properly it will be difficult to proceed.(See Mari (1996)) Following the spirit of the ISO Guide (ISO, 1993), we interpret the *measurand* to be a "specific quantity subject to measurement", or, about which quantitative information is required. A *measurement* on the other hand is a logical procedure, having as its aim the determination of the measurand. *Influence Quantities* are those quantities, secondary to the measurement, but nevertheless affecting its result, whose effects must be considered in order to properly arrive at the measurand. Indeed, a full statement of the problem will indicate the quantity to be obtained [the measurand] and under what conditions [e.g. temperature, barometric pressure etc.] this is to be done. Note that there are many influence quantities of a short term nature, of which the experimenter is not aware, which as a result of 'lack of knowledge' are interpreted as "random" fluctuations. As a result, a full statement completely describing the measurand is impossible without an infinite amount of information; and hence, apart from intrinsic constants of nature, the measurand remains potentially unknown and unknowable. This idea leads us naturally to concepts of "uncertainty" and "accuracy".

To the extent that we cannot fully model all influence parameters effecting our determination of the measurand, we must introduce a quantitative measure of the resulting "uncertainty" in our effort. The measurand itself is deterministic but it is also indeterminate—the distinction between these two being important. We mean by this that the measurand has a real, definite, value at the instant of measurement, which can never be determined with infinite accuracy.

In other words, because our model parameterisation is imperfect, we must refer to the resulting 'corrected measured value' [which is our measurement subject to whatever systematic corrections we know about] as an ESTIMATE, and as such, we need to establish *Degrees of Belief,* or *Plausibility* for it. We need *Dispersion Characteristics* for the estimate, in order to give an indication of the range of values it could reasonably adopt—any one of which, based on the information we have, could be the measurand under scrutiny. We are interested, in this chapter, in seeing how we can arrive at such a measure, basing our investigations on the totality of the information available to us. We term the difference between such a value and the measurand, as the *error*. Clearly the real value of this is also unknowable, as a result of the measurand being indeterminate. It should be an aim of any experiment to ensure this error is small. Just because the Dispersion measure is small, does not mean the

error is small—it simply means that the evidence from existing knowledge is accurate to within tight boundaries while saying nothing about other possible [systematic] information which may have been unrecognized. It may be possible to double-check for this type of situation by performing another experiment based on different physical principles and re-determining an estimate of the measurand. *Reproducibility* is the degree to which these two estimates are in agreement, good reproducibility suggesting the errors are small. 'Good agreement' would be defined as agreement to within the combined dispersion characteristics of the two estimates.

1.3 Basic Statistical Terms

We must now turn our attention to the identification and quantification of the dispersion characteristics of our estimate of the measurand. As we would expect, these will depend upon the dispersion characteristics of the various individual quantities involved in generating the corrected realised quantity (our estimate). Following the treatment of e.g. Beck & Arnold (1977), Eadie (1971) or Ross (1972), we use the probability density function $p_Z(z)$ for a parameter Z (usually called a continuous random variable in this context) to describe the range of possible values the parameter could adopt. This function is normalized such that

$$\int_{-\infty}^{\infty} p_Z(z)dz = 1 \tag{1.3.1}$$

The probability distribution function $P_Z(z)$ gives the probability that the random variable Z is less than some value z. Thus

$$P_Z(z) = \Pr(Z \le z) = \int_{-\infty}^{z} p_Z(u)du \tag{1.3.2}$$

where $\Pr(x)$ is the probability associated with the value x, expressed as a fraction or percentage. There are some important 'statistics' associated with a probability distribution which we can now define.

Expectation Value: For a continuous random variable, z, we have

$$\mu_z = E[z] = \int z\, p_Z(z)dz \tag{1.3.3}$$

Note that $E[z]$ is a linear operator, i.e.:

$$E\left[\sum_{i=1}^{n} a_i x_i\right] = \sum_{i=1}^{n} a_i E[x_i] \tag{1.3.4}$$

$$\&\quad E[ax + by] = a\,E[x] + b\,E[y] \tag{1.3.5}$$

The expectation value can usually be estimated by the arithmetic mean

$$\bar{z} = \frac{1}{n}\sum_{i=1}^{n} z_i \tag{1.3.6}$$

That this is an unbiased estimator can be seen by considering that, if $E[z_i] = \mu_z \;\forall\; i$ holds, then:

$$E[\bar{z}] = \sum_{i=1}^{n} \frac{1}{n} E[z_i] = \mu_z \tag{1.3.7}$$

That an estimator is unbiased is a particularly desirable feature and one to which we will return many times in the future when looking at more complex parameter estimation techniques. A biased estimator is really not an estimator of the desired measurand at all.

Variance: This is the second principal term of interest. The variance of a random variable is defined as :

$$V[z] = \sigma^2(z) = E\left[(z - E[z])^2\right] \tag{1.3.8}$$

$$= E\left[(z - \mu_z)^2\right] \quad ,$$

$$= \int (z - \mu_z)^2 p_Z(z) dz \tag{1.3.9}$$

This indicates that the Variance is the Expectation Value of the squared centered random variable $(z - \mu_z)$. We must be aware that variance is *not* a linear operator! For example, for a constant a we find:

$$V[ax] = E[ax - E[ax]]^2$$

$$= a^2 E[x - E[x]]^2 = a^2 V[x]$$

also, for constants a & k we find $V[ax + k] = a^2\, V[x]$. For $\bar{z}$ as defined in (1.3.6), we have

$$V[\bar{z}] = \frac{\sigma^2(z)}{n} \tag{1.3.10}$$

where $V[z_i] = \sigma^2(z)\ \forall\ i \in (1,....n)$. A valid estimate of $\sigma^2(z)$, obtained from n observations of z is:

$$s^2(z) = \left(\frac{1}{n-1}\right) \sum_{i=1}^{n} (z_i - \bar{z})^2 \tag{1.3.11}$$

in which $s(z)$, the positive square root of (1.3.11), is usually referred to as the 'standard deviation' of the random variable z. This statistic is used to quote an uncertainty for an estimate of μ_Z. Essentially, $E[z]$ is a location parameter, giving the position of a distribution of values while $V[z]$ is a scale parameter for the dispersion characteristics associated with the particular distribution. If μ_Z is known, rather than estimated by $\bar{z}$, (1.3.11) becomes:

$$\sigma^2(z) = \left(\frac{1}{n}\right) \sum_{i=1}^{n} (z_i - \mu_z)^2 \tag{1.3.12}$$

We can see immediately how this kind of information should be very useful in establishing accuracy criteria/uncertainty limits on an experimental measurement, should we be able to compute it for a given experimental situation. When we can carry out repeat measurements in circumstances where uncorrectable random fluctuation occurs, it is then possible to compute (1.3.6) & (1.3.11). (For another view on uncertainty measures, see Allan, (1987)). In order to establish a probability

distribution from this information in a consistent manner, as advocated by the "unified approach" of this thesis, we must turn to the Maximum Entropy Principle as discussed in Section 1.5.

Covariance: If there are two random variables defined on the same sample space, their covariance is defined as:

$$Cov[y,z] = E\left[\left(y-\mu_y\right)\left(z-\mu_z\right)\right] = v(y,z) \tag{1.3.13}$$

$$= \iint \left(y-\mu_y\right)\left(z-\mu_z\right)p(y,z)\,dydz \tag{1.3.14}$$

where $p(y,z)$ is the joint probability density function. The covariance can be estimated from n simultaneous observations of y & z by:

$$s(y,z) = \left(\frac{1}{n-1}\right)\sum_{i=1}^{n}\left(y_i-\bar{y}\right)\left(z_i-\bar{z}\right) \tag{1.3.15}$$

$\bar{y}$ & $\bar{z}$ being the respective arithmetic means.

Correlation: The correlation coefficient is defined as:

$$\rho(y,z) = \frac{Cov[y,z]}{\sqrt{V[y]}\sqrt{V[z]}} = \frac{v(y,z)}{\sigma(y)\sigma(z)} \tag{1.3.16}$$

Using estimates, (1.3.16) becomes:

$$r(y,z) = r(z,y) = \frac{s(y,z)}{s(y)s(z)} \tag{1.3.17}$$

$\forall\, y,z$ as sample elements from the space of y , z values. Note that $-1 \le r(y,z) \le +1$, as the correlation coefficient is a pure number, indicative of the relative mutual dependence of the two variables y & z. Thus it gives the estimated change in one variable likely to result from a given change in the other. Also, with respect to covariance:

$$V[ax+by] = a^2V[x] + b^2V[y] + 2abCov[x,y] \tag{1.3.18}$$

Eqs. (1.3.6), (1.3.11), (1.3.15) & (1.3.17) allow the evaluation of the center and spread of a distribution which is thought to characterise the measurements made to determine the parameter of interest, along with the interactions between any pair of similar parameters. In many experimental cases the distribution so described may very well be Normal, or Gaussian, but this should not *a priori* be assumed. However, depending upon our knowledge, it can be satisfactorily confirmed using the Maximum Entropy Principle. In situations where repeated data is obtained with the measurement instrumentation, an examination of relative frequencies in the results allows both mean values and variances to be estimated and as we show in Sec. 1.5, maximum entropy does predict a Gaussian distribution in these circumstances. There are, though, many cases where such data is not available, and one must assign a distribution ***first*** in order to estimate a variance. However, before looking at these situations in more detail, we must consider the propagation of measurement uncertainty via the functional relationship for the measurand estimate. Since the desired parameter

depends on several, if not many, input quantities, it is necessary to investigate how the individual variances contribute to the final one. It will be in this context that we make further remarks about the determination of variance and the assignment of confidence intervals.

1.4 Uncertainty Propagation (Following the ISO Guide's treatment (ISO, 1993))

Assume that a parameter z is determined from a functional relationship given by:

$$z = f(x_1, x_2, ..., x_N) \tag{1.4.1}$$

i.e. z depends upon N influence quantities, each of which is either determined in the current measurement procedure, or is known initially from another source. We further assume that dispersion information is available on all the N influence quantities—we will comment later on how this might be obtained.

A Taylor series expansion, to Order 1, of (1.4.1) will yield:

$$(z + \delta z) = z + \sum_{i=1}^{N} \frac{\partial f}{\partial x_i} \delta x_i \tag{1.4.2}$$

for δz & δx_i small deviations from z & x_i respectively. So an evaluation of (1.4.2) will give an estimate of the deviation of our estimate z from the measurand value, providing we know deviation estimates for the individual influence quantities. For each influence quantity we can assign a probability distribution describing the range of possible values it could adopt. Therefore, from Section 1.3, we will have an expectation value and variance to describe the quantity.

Hence if we write

$$\delta z \equiv (z - E[z]) \tag{1.4.3a}$$

$$\& \qquad \delta x_i \equiv (x_i - E[x_i]) \tag{1.4.3b}$$

We can then re-write Eq. (1.4.2) as:

$$(z - E[z]) = \sum_{i=1}^{N} \frac{\partial f}{\partial x_i} (x_i - E[x_i]) \tag{1.4.4}$$

Squaring both sides gives:

$$\begin{aligned} \{z - E[z]\}^2 &= \left\{ \sum_{i=1}^{N} \frac{\partial f}{\partial x_i} (x_i - E[x_i]) \right\}^2 \\ &= \sum_{i=1}^{N} \left(\frac{\partial f}{\partial x_i} \right)^2 (x_i - E[x_i])^2 + 2 \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \frac{\partial f}{\partial x_i} \frac{\partial f}{\partial x_j} (x_i - E[x_i])(x_j - E[x_j]) \end{aligned} \tag{1.4.5}$$

Now taking expectation values of (1.4.5):

$$E[z - E[z]]^2 = \sum_{i=1}^{N} \left(\frac{\partial f}{\partial x_i} \right)^2 E[x_i - E[x_i]]^2 + 2 \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \frac{\partial f}{\partial x_i} \frac{\partial f}{\partial x_j} E[(x_i - E[x_i])(x_j - E[x_j])] \tag{1.4.6}$$

But since from Eq. (1.3.8), $E[z - E[z]]^2 = \sigma_z^2$, the variance of z, we can write (1.4.6) as:

$$\sigma^2(z)=\sum_{i=1}^{N}\left(\frac{\partial f}{\partial x_i}\right)^2\sigma^2(x_i)+2\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\frac{\partial f}{\partial x_i}\frac{\partial f}{\partial x_j}v(x_i,x_j) \qquad (1.4.7)$$

(1.4.7) can be further expressed as:

$$\sigma^2(z)=\sum_{i=1}^{N}\left(\frac{\partial f}{\partial x_i}\right)^2\sigma^2(x_i)+2\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\frac{\partial f}{\partial x_i}\frac{\partial f}{\partial x_j}\sigma(x_i)\sigma(x_j)\rho(x_i,x_j) \qquad (1.4.8)$$

with $\rho(x_i,x_j)=\dfrac{v(x_i,x_j)}{\sigma(x_i).\sigma(x_j)}$ being the correlation coefficient

In commenting on this development, we can point out that Eq. (1.4.8) is a *complete* description of the dispersion characteristics of z as specified in (1.4.1); based on a knowledge of the dispersion characteristics of the influence parameters and also on their correlations, if any. The positive square root of (1.4.8) gives the Combined Standard Uncertainty and this serves as an adequate measure of the uncertainty in the measurement/analysis process—*based on the currently available knowledge* and is accurate as an estimate of the possible variation between our knowledge and the value of the measurand in question, assuming no known systematic effects have been inadvertently omitted. Expectation values and Variances have been used in the analysis and it is entirely general. Eq (1.4.8) is recognized as the Law of Propagation of Uncertainty (ISO, 1993) otherwise known as the Gaussian Procedure for Error Propagation (see Weise, 1985 & 1987). It is illustrated in Fig. (1.0.1), page 10.

One of the key aspects of this procedure is the uniform manner in which it treats all variance components. Traditionally uncertainty components are divided among so-called "random" and "systematic" components. The former are those which can be estimated by examination of relative frequencies in a set of data, while the latter are of a more constant nature—they do not decrease with increasing sample size for example. However, that which is systematic in one experiment may very well be random in another, so the distinction can be confusing.

In the language of the ISO Guide (ISO, 1993) the positive square root of a variance component is referred to as a *Standard Uncertainty*. It is the basic building block/component for error propagation theory and since it is a mathematical term there is no need for further divisions into random/systematic sections.

In understanding the ISO Guide approach to uncertainty analysis it is helpful to look at the basis of Classical Probability Theory, which is essentially Inductive Logic—that is, given effects, we want to decide among several possible causes what is the most likely candidate to explain the observed phenomena.(see, e.g., Buck & Macaulay, (1991) and Garett, Chapter 6 in *ibid*).

This framework is ideal for analysing experimental data in which we have effects (data or observations) for which we need to know causes (hypotheses, models, parameter values, etc.). In stating that we desire to find "the most likely" cause, we are immediately invoking the idea of 'plausibility' or 'likelihood' or 'truth value' or some such term indicating the extent to which we believe a given hypothesis or proposition is the best explanation of the observed effects. Thus we can say that the probability of the proposition being true is the same as the *Degree of Rational Belief* about the possibility of the proposition being true, and it is this interpretation of probability which is crucial to a correct understanding of the treatment of uncertainty developed in this thesis. (See, e.g., Jaynes, 1957, Jaynes, 1982, Cox, 1946)

Referring to a probability as a Degree of Belief reminds us that probabilities should not be considered absolute! A degree of belief will always be tempered by the totality of knowledge we have about a subject. Thus probabilities are always conditional on other, background (prior) information. It is important to realise here that conditional probabilities, which assume prior information, represent *Logical* connections, rather than *Causal* ones. For example, if our available knowledge leads us to be sure proposition A is true, then what does this logically imply for some other proposition B? In this way probabilities represent epistemological knowledge rather than ontological information. They describe what we know about the event or proposition, based on the evidence at hand, and allow us a framework for reasoning in the absence of certainty. (See Jaynes, 1983, Levine & Tribius, 1979).

It is clear from the foregoing paragraphs that we have carefully avoided any reference to the idea of Relative Frequencies in repeated trials of an experiment in coming to our definitions of probability. It is precisely this which has led to the charge of 'subjectivism' being leveled against this approach. It is thought that a list of relative frequencies provide a definite measure of objective reality which can be completely relied upon. However, this can be objected to on several grounds, not least of these being the problems attached to ensuring repeated trials are indeed reproducing "random" errors. A more fundamental problem is to assume that this "randomness" is a property of natural systems, existing in an ontological manner. The view of probability being highlighted here has no such requirement. Rather, randomness simply is an explanation of our lack of knowledge—we do not know all the forces and influences affecting our system and therefore cannot predict with 100% certainty exactly what will happen. Thus we say there are "random errors" affecting the system. Other problems arise in the frequentist approach where we must imagine a large set of possible outcomes which *could* have happened (but didn't!) of which our small set of observations is a member. (see, e.g. Jaynes, 1996)

If this approach should seem excessively subjective and thus dependent upon the individual's analysis, it should be realised that Degrees of Belief are based on all the relevant information that is available. Thus two people presented with the same information should assign the same probability; (Cox, 1946) another person with different information or observational experience may well make a different diagnosis—hence we see again that probabilities are *always* conditional on the available background information. Incidentally, we can further point out that we always speak of assigning probabilities, not determining them. This is because probability is a mathematical description of what we know about the system and not some inherent property of the system.

In implementing a probability analysis based on this philosophy there are several Criteria we need to enumerate which will form the basis for ensuring that everything we do maintains a coherent approach. Following Jaynes (1996), we can state these as follows:

1). Degrees of Plausibility should be represented by real numbers. This we have already hinted at by noting that probability theory simply gives a mathematical statement of what we know.

2). Secondly there should be Qualitative Agreement with Common Sense, which we would expect on the basis of Logical Reasoning.

Finally, and most importantly, we have several Desiderata, or **Criteria of Consistency**. These are as follows:

3a). If a conclusion can be reached in more than one way, then every possible avenue of logic should lead to the same result.

3b). All available evidence relevant to a question must be considered. Portions cannot be arbitrarily left out, conclusions being based only on what remains.

3c). Equivalent states of knowledge must be represented by equivalent plausibilities. Hence in two problems, if the same state of knowledge exists, the same plausibility must be assigned to each.

These criteria underpin the basic rules of probability theory, from which can be established a consistent and logical data analysis. We will look at these rules in more detail later when we consider Bayesian Parameter Estimation in Chapter 8. Here we just want to point out the essential features of probability analysis and show that these criteria exist as desirable goals to be aimed at in any analysis. With that in mind, we

can consider again the uniform manner in which the Gaussian Error propagation procedure deals with all the influence quantities in a functional relationship. This Unified Approach certainly makes analysis easy, but we want to ensure that it is justified in the light of the criteria of consistent reasoning underlying probability logic.

Initially we point out that it is widely accepted that some errors are random in nature—that is they can be arbitrarily reduced in size by increasing the number of samples taken—while others will remain constant and are not affected by repeated measurements. (Bohm (1984) usefully defines errors as being 'due to contingencies outside the context of the experiment').We reiterate that these "errors"–of whatever nature–are deviations between the measurand and either the realised or corrected realised quantity. As such, they are of course unknowable (otherwise we could correct for them and they would no longer be errors) and that is why we represent the measurand estimate by a probability distribution, which is a statement either negatively of *lack* of knowledge, or positively, of the extent of our knowledge/our degree of belief about the measurement just made (Bölöni, 1997). Note that all of these distributions are selected on the basis of available information—we cannot postulate data that does not exist. Any new data can be compared with the previously accepted distributional information and a posterior estimate constructed accordingly, reflecting any changes the new information might imply.

So we can see that this uniform approach to probability assignment is a valid and acceptable method, in view of the type of (incomplete) information available. The fact that we assign distributional properties/dispersion characteristics to a constant systematic error is simply a numerical description of what we know about the systematic error and not an infallible statement about its true numerical value. Of course, to the extent that a systematic error is known to exist, a correction should be made for it on the basis of whatever information led to the conclusion that it existed and could be quantified (See Weise & Woeger, 1992 and also 1994 for an application of Bayesian uncertainty analysis).

It is important to realise in the Unified/Gaussian theory that we are not dealing with distributions of errors, but with distributions attached to the parameter estimates. This is because we know nothing about the errors so rather we consider the distribution as a statement of our degrees of belief in our parameter estimate. This is quite a different approach philosophically (which may well lead to the same numerical results in many cases) which shows clearly why the random/systematic distinction is redundant—we're dealing with parameters, not errors

The Guide does, however, permit the classification of the methods of obtaining standard uncertainties into Type A (those based on statistical methods—nearly always involving dealing with repeated data/an examination of relative frequencies) and Type B (those based on "other" methods—i.e. non-statistical methods). The need for Type B methods arises as a result of the requirement to provide a Standard Uncertainty in all cases: if an analysis of repeat measurements is not possible, a *distribution* must be estimated in some way first and then a variance appropriate to this distribution can be obtained. It may well be that little information is available about the dispersion of the estimate, but remembering to interpret probability as Degrees of Belief, then whatever distribution can be decided upon is simply a reflection of what knowledge exists at the time about the estimate, be that more or less {See Annex E in ISO, 1993}.

The overall goal of this Uncertainty Propagation Procedure is to generate a simple Variance/Standard Uncertainty for the output quantity, which can then be easily incorporated into other analyses, and is clearly understood. Many problems can arise if an estimate is claimed to have an uncertainty of x with a $y\%$ confidence level. Unless something is known about the distribution of x, it is impossible to properly understand the quoted uncertainty. (See also Mueller, 1984).

However, if a Combined Standard Uncertainty is given (i.e. positive square root of (1.4.8)), then a variance is immediately available for inclusion in subsequent work. It is not necessary to give confidence levels for the result, but this is often considered to be useful information. To do this though, requires knowledge about the shape and type of the distribution of the estimate, and this will depend on a convolution of all the probability distributions of the influence quantities involved in generating the estimate. This convolution can be difficult to evaluate if there are many different types of distributions assigned to the various parameters. However, in many cases the output distribution will be approximately Normal, even if the input distributions are not exactly Normal. This is a consequence of the central Limit Theorem (Beck & Arnold, 1977), (Eadie, 1971). If there is a large non-Normal distribution element present in the input, the conditions become less favorable to the Central Limit Theorem and its validity cannot be guaranteed. However it is common practice to include a so-called "coverage factor", usually of $k = 2$, to give an Expanded Uncertainty which often has a confidence level of the order of 95% if the distribution is approximately Normal. The use of higher coverage factors—e.g. $k = 3$ for ~99% confidence is hard to justify since this results in the widening of the uncertainty bracket to include possible values to which a vanishingly small probability of occurrence has already been assigned.

Aside (Correlation): We have seen in (1.4.8) how to evaluate $\sigma^2(z_1)$ where $z_1 = f(x_1,\ldots x_N)$, i.e. z_1 is a function of (some of) the N parameters $x_1 \ldots x_N$. Now

if we have another measurand estimated by $z_2 = g(x_1,...x_N)$, we can of course also evaluate $\sigma^2(z_2)$. However, any terms common to both z_1 and z_2 will result in the possibility of a correlation between the two functions, which could be significant if they were to be later combined in a third evaluation. In general, this covariance is evaluated by:

$$v(z_1,z_2) = \sum_{i=1}^{N} \frac{\partial f}{\partial x_i} \frac{\partial g}{\partial x_i} \sigma^2(x_i) \qquad (1.4.9)$$

Thus for any x_i not common to both z_1 & z_2, one or other of the partial derivatives will be zero, and if no terms are common, there will be no correlation.

To conclude this section we highlight the principal points of this Unified Method of model parameterisation and uncertainty analysis. The measurand must be clearly defined: the circumstances of measurement, physical conditions etc. must all be stated. The functional form of the realised quantity from the experiment should be set down and all known corrections for various systematic biases must be included to obtain the corrected realised quantity—this is the estimate of the measurand desired. Then an investigation of the dispersion characteristics of the various influence quantities must be carried out in order to obtain the dispersion characteristics of the output quantity.

The key to the approach is to establish a variance for all the parameters, the positive square root of which is termed a Standard Uncertainty. The variances, and any known covariances, are treated by the Gaussian Procedure to obtain the Combined Standard Uncertainty of the estimate of the measurand. (See Arri, 1996 & Cox, 1996 for comments on measurement procedure and model parameterisation)

Coverage factors may be included at this stage, to give an *expanded uncertainty* as it is termed in the ISO Guide, but care should be exercised in doing this and the way in which it is done should be transparent to ensure that no ambiguity arises. Coverage factors can tend to have a somewhat "sledgehammer" effect on the whole process, which emphasizes *realistic* estimation of uncertainties, rather than so-called "safe" estimates. (Other thoughts on coverage factors are given in Godec (1997)). In other words our uncertainty measures should only be based on the extent of our available knowledge and not on "guesstimates" with no justification from the current data. The reason for this can be further reinforced by considering that whatever distribution the estimate takes, including a coverage factor implies pushing the boundaries of possible values out into the tails of the distribution, which by their very nature are considered highly improbable.

In the Unified Approach it is permissible to point out that the combined standard uncertainty contains components from various sources and to itemize these. Such information could be useful to an end-user of the calibration information in comparing the quoted uncertainty from two different establishments, as it is not always easy to identify how the original calibrator arrived at the stated uncertainty. However, note that the functional relationship will dictate the *Sensitivity Coefficients* (partial derivatives of the functional form) and these multiplicative factors will affect the overall contribution of each term to the final combined standard uncertainty. Perhaps a good method would be to state the functional relationship and then to tabulate the standard uncertainties of each component and their contribution to the overall result. Such a suggestion is indeed but forward in EAL (1997).

1.5 Subjective Probabilities and Maximum Entropy

The Unified Approach to data analysis in mass metrology developed in this thesis includes an approach to uncertainty analysis, aspects of which have been the cause of much controversy in the metrological community. This in particular concerns the manner in which it removes the distinctions between random and systematic uncertainty components. We have shown above how the method is justified by a correct understanding of probability theory and of the nature of the information available to the analyser of experimental data. The points we have raised will be relevant to the remainder of this thesis and we will see the emphasis on a Unified Approach and a Consistent Analysis many times in the succeeding chapters, particularly in discussions on Parameter Estimation methods and data analysis of mass calibration experiments. (See Bretthorst (1989) for a helpful tutorial on applications of Bayesian probability theory to parameter estimation. Fröhner (1997) has further useful information).

However, the above considerations of Probability Logic notwithstanding, the major controversy has centered on the assignment of Probability Distributions to experimental data, the classical analysts holding rigidly to a separation of systematic and random variables, claiming that the unified approach is too subjective to be realistic (e.g. Colclough, 1987). From our earlier comments of course, we can claim that all probabilities are subjective, depending as they do on available information in reaching a decision. It can indeed be argued that the traditional approach, too, must use subjective assessments in establishing uncertainties for suspected systematic errors, about which very little prior knowledge may be available.

Hence out task now is to ensure that the probability assignments we do reach are completely unbiased with respect to the available information. Being 'unbiased' is quite different to making a 'subjective' assessment: the latter is simply a reflection of our finite knowledge; while the former is an assurance that our demands of consistency are met (Cox, 1946): all available knowledge is used, we do not arbitrarily ignore some of it, and neither do we assume other information not warranted by the given data and known history.

In many cases it is possible to establish a distribution, or at least a variance estimate, empirically, by an examination of relative frequencies from repeated experimental data. Where such information is available it is of course valid, but in other cases an estimate of dispersion characteristics must be established by invoking such criteria as "experience", "available information" or some similar idea. So how can we be sure that such an estimate is the best we can do? We need some procedure which can guarantee we have done the best, most optimum, analysis with the supplied information. Remember that we very much desire "optimal estimation" of uncertainties—not a 'sledgehammer' approach of coarse limits which must contain the measurand, and not either, excessively optimistic narrow limits.

The easiest way to assign probabilities is by invoking the "Principle of Indifference": if it is possible to break the problem up into a set of mutually exclusive and exhaustive possibilities, then there is essentially no reason to assign any one of them a higher probability than any other and we arrive at a Uniform Distribution, which is intuitively the simplest we can imagine. However, often we do have information to suggest that some propositions (or data) *are* more likely than others. What do we do then? How do we proceed in a manner which takes account of this more specific information while at the same time remaining unbiased and not making unwarranted assumptions about the other unknown information? (see Jaynes, 1985)

We know the Uniform Distribution to be the most non-committal with regard to all possibilities, while a perfectly sharp function (e.g. Delta Function) is absolutely definite as to the parameter's value. We need some method which can reproduce both of these situations whilst also spanning the continuum between them in a manner which is maximally unbiased. Again let us re-iterate what is meant by 'unbiased': we mean adherence to the rules of consistency such that no attempt is made to assume knowledge we do not have. Thus any distribution should be as vague as possible while taking account of any known data (testable information).

Every probability assignment can be looked upon as expressing how much uncertainty we have about the proposition, or parameter. This is not to be interpreted

negatively, but rather as a fair expression of the limitation in our knowledge. What we need is a numerical measure of the "amount of uncertainty" represented by a distribution which tells us how little we know. The most consistent probability assignment will then be the one which maximises this, subject to the constraint of whatever we *do* know—i.e. what testable information do we have? The uncertainty content is largest for a Uniform measure and zero for a Delta function.

Such a measure does exist, and is termed the **Entropy** of a probability distribution. (See, e.g., Woeger, 1987, Sivia, 1996, Lieu, 1987). It is a measure of the uncertainty or alternatively, the Information Content of a distribution and can be given by:

$$S = -\int p(x) \log\left(\frac{p(x)}{m(x)}\right) dx \tag{1.5.1}$$

for $p(x)$ the distribution in question and $m(x)$ a function representing particular prior information available about the problem. Maximising S subject to the constraints of any known testable information will yield the best probability assignment that can be made. It tells us how much we don't know about the parameter, or how Uniform is the probability distribution. Note that it does not tell us which distribution is absolutely right, but simply is a means of inductive reasoning in the absence of certainty which tells us what conclusions are the most plausible in the context of the currently available information. Some useful information with respect to the relationship between thermodynamics and data analysis / information processing is found in Trebbia (1996).

The general approach to evaluating S_{max} is as follows: (Jaynes, 1996) we take a discreet case where we have a set of possible values $(x_1, x_2, \ldots\ldots x_n)$ for a parameter x and a corresponding probability distribution $(p_1, p_2, \ldots\ldots p_n)$. There may be a set of m functions of the data whose mean values we know. This is our testable information. Thus we have a set of $f_k(x)$ for $1 \le k \le m$ and also

$$F_k = E[f_k(x)] = \sum_{i=1}^{n} p_i f_k(x_i) \tag{1.5.2}$$

So our constraints include, firstly normalisation of the distribution, $\sum p_i = 1$ and also a set of F_k given by Eq. (1.5.2). If we define

$$Q = S + \lambda_0\left(1 - \sum p_i\right) + \lambda_1\left(F_1 - \sum p_i f_1(x_i)\right) + \ldots\ldots + \lambda_m\left(F_m - \sum p_i f_m(x_i)\right) \tag{1.5.3}$$

where S is now given in the discreet case by $S = -\sum p_i \log\frac{p_i}{m_i}$, then in the Lagrange Variational method we require $\frac{\partial Q}{\partial p_i} = 0$ in order to maximise S subject to the $(m+1)$ constraints. That is, we need:

$$\frac{\partial}{\partial p_i}\left\{-\sum_{i=1}^{n} p_i \log\frac{p_i}{m_i}+\lambda_0\left(1-\sum p_i\right)+\sum_{j=1}^{m}\lambda_j\left(F_j-\sum_{i=1}^{n} p_i f_j(x_i)\right)\right\}=0 \qquad (1.5.4)$$

$$=\sum_{i=1}^{n}\left\{-1-\log\frac{p_i}{m_i}-\lambda_0-\sum_{j=1}^{m}\lambda_j f_j(x_i)\right\}=0 \qquad (1.5.5)$$

Hence for each p_i we have:

$$\left(\frac{m_i}{p_i}\right)\exp\left(-\lambda_0-\sum_{j=1}^{m}\lambda_j f_j(x_i)\right)=e^1 \qquad (1.5.6)$$

$$\boxed{\Rightarrow p_i = m_i e^{-(1+\lambda_0)} e^{-\sum\lambda_j f_j(x_i)}} \qquad (1.5.7)$$

Or, in continuous form:

$$p(x)=m(x)\,C\,\exp\left[-\sum_j \lambda_j f_j(x)\right] \qquad (1.5.8)$$

for $C=e^{-(1+\lambda_0)}$. Thus knowing the Lagrange parameters λ_0 and $\lambda_1.....\lambda_m$ allows the probability distribution to be determined. This involves solving *(m+1)* simultaneous equations which may require numerical methods.

The function *p(x)* given by Eq. (1.5.8) generates the most consistent probability assignment for the known information. It is 'subjective' to the extent that it is a measure of our 'uncertainty' or lack of knowledge, but it is a completely objective use of the available data on the problem.

We will look now at just two situations which we will find are sufficient for the subsequent analysis in this thesis (Sivia, 1996). In the first case we consider, there are no known constraints and *m(x)* is a Uniform Measure—that is, we are completely noncommittal about the possible parameter values, only knowing that $\sum_{i=1}^{M} m_i = 1 \quad \Rightarrow \quad m_i = 1/M$ by the Indifference Principle. Thus in Eq. (1.5.7), the second exponential reduces to unity and we have:

$$p_i = m_i e^{-(1+\lambda_0)} \qquad (1.5.9)$$

We know that $\sum_{i=1}^{n} p_i = 1$ by normalisation so therefore we find:

$$\lambda_0\left(1-\sum p_i\right)=0$$
$$\Rightarrow \lambda_0\left(1-\sum m_i e^{-(1+\lambda_0)}\right)=0$$

But given that $\sum_{i=1}^{M} m_i = 1$ we find that $\lambda_0 = -1$ for a non-trivial solution and thus from Eq. (1.5.9), $p_i = m_i$ or, in the continuous case, $p(x)=m(x)$. This is of course intuitive and tells us that, in the absence of any other information, except normalisation on *p(x)* and an uninformative prior *m(x)*, the best we can expect is just such a uniform distribution for which :

$$m(x) = \begin{cases} \dfrac{1}{x_{\max} - x_{\min}} & \text{for } x_{\min} \le x \le x_{\max} \\ 0 & \text{otherwise} \end{cases} \qquad (1.5.10a)$$

with a variance of

$$s^2(x) = \frac{(x_{\max} - x_{\min})^2}{12} \qquad (1.5.10b)$$

Thus when we know nothing except the boundaries for x, maximising the Entropy predicts a Uniform Distribution, which we would in any event expect, both intuitively and by the Principle of Indifference.

A second common situation arises when we know a mean value $E[x] = \mu$ and a variance of

$$E\left[(x-\mu)^2\right] = \sigma^2 = \int (x-\mu)^2 p(x) dx \qquad (1.5.11)$$

Once again we need not have any prior reason to select one dispersion characteristic over another so we can have *m(x)* as a uniform distribution. Thus Eq. (1.5.8) gives, for the maximum entropy distribution

$$p(x) = m(x) e^{-(1+\lambda_0)} e^{-\lambda_1 (x-\mu)^2} \qquad (1.5.12)$$

since we have just one constraint $f_1(x) = (x-\mu)^2$ with expectation value $F_1 = E[f_1(x)] = \int f_1(x) p(x) dx$. We can write Eq. (1.5.12) as

$$p(x) = C\, e^{-\lambda_1 (x-\mu)^2} \qquad (1.5.13)$$

since *m(x)* is a constant(uniform) distribution. With this value for *p(x)* we can go back to the constraint equation (1.5.11) to find:

$$\sigma^2 = \int_{-\infty}^{\infty} (x-\mu)^2 C\, e^{-\lambda_1 (x-\mu)^2} dx \qquad (1.5.14)$$

where the $(-\infty, \infty)$ limits can be set unless we have further constraints to the contrary. Evaluating this standard integral yields:

$$\sigma^2 = \frac{C\sqrt{\pi}}{2\lambda_1^{3/2}} \qquad (1.5.15)$$

while the normalisation constraint gives another standard integral:

$$1 = \int_{-\infty}^{\infty} C e^{-\lambda_1 (x-\mu)^2} dx \qquad (1.5.16)$$

$$\Rightarrow 1 = \frac{C\sqrt{\pi}}{\lambda_1^{1/2}} \qquad (1.5.17)$$

Solving Eqs. (1.5.15) & (1.5.17) as simultaneous equations yields values for λ_1 & C:

$$\lambda_1 = \frac{1}{2\sigma^2} \qquad (1.5.18a)$$

$$C = \frac{1}{\sigma\sqrt{2\pi}} \qquad (1.5.18b)$$

Thus we finally have for Eq. (1.5.13):

$$\boxed{p(x)=\frac{1}{\sigma\sqrt{2\pi}}\exp\left\{\frac{-(x-\mu)^2}{2\sigma^2}\right\}} \qquad (1.5.19)$$

which we can recognise as a Gaussian Distribution. This shows us that the Gaussian Distribution is the most honest description when we know nothing except a mean value and a variance estimate.

This development is significant since we said earlier that even though a mean value and variance estimate could be computed from an empirical examination of relative frequencies, we could not automatically assume a Normal Distribution applied. Further information would be needed we thought; however, here we see that in fact a Normal Distribution is the best we can do when presented with such information.

1.6 Conclusion

This concludes our study of the fundamentals of probability theory and its application to uncertainty analysis. The concepts we have developed here are crucial to what follows and we will see how the general philosophy of a unified approach and a consistent analysis is applied to all subsequent calculations.

We have seen in this chapter how the Law of Error Propagation provides a convenient and mathematically concise (as well as accurate!) representation of the uncertainty influence of all involved parameters on the final outcome. Criteria of Consistency underlying probability theory provide a firm justification for this approach and show that a reliance on relative frequencies is unnecessary as these are only a subset of the extant information on the subject.

A crucial point for the ISO procedure is the establishment of variance components for all terms in the functional relationship. The Maximum Entropy Formalism indicates a Uniform measure when only upper and lower limits for a parameter are known, while a Gaussian Distribution best describes parameter estimates for which means and variances are available. These two cases adequately describe the information presented in succeeding chapters, and so we can proceed, confident that the Unified Approach can be maintained.

****

2. Parameterising Mass Calibration Experiments

2.0 Summary

In this chapter and the next we apply the principles of consistent reasoning developed in Chapter 1 to an analysis of the issues arising in mass calibration by comparison experiments. The basic measurand to be determined is the mass *difference* of two standards or ensembles of standards. The data is obtained from 'in-air' comparisons with automated mass comparators, resulting in the need for various corrections (influence quantities) to be incorporated in the overall functional relationship. We illustrate the development of a scalar version of the 'Weighing Equation', the key functional relationship in mass metrology. Incorporated in this are well-known corrections for buoyancy, center-of-gravity differences and volume expansion coefficients.

We then illustrate how the uncertainty analysis is carried out according to the principles of the ISO Guide. This is a new development where we adopt the unified approach and do not separate the components into random and systematic terms as is traditionally done. The expressions for the standard uncertainty thus developed can then easily be incorporated into other work as required.

2.1 Introduction

We wish to apply the theory developed in Chapter 1 to the analysis of comparison experiments in mass calibration. The experimental procedure is well known and will not be touched on here, although some details are given in Chapter 11. For our purposes now it will be sufficient to note that comparisons are carried out between nominally equal mass standards, or ensembles of mass standards, the residual mass difference between the pair being the measurand of interest. Data is obtained with Electromagnetic Force Compensation comparators and various corrections may be required to the resulting data. For example, operating in air implies a correction for the difference in buoyant force, should the two standards have different densities. There may also be corrections due to center of gravity differences and perhaps different volume expansion coefficients. There can also be issues arising from magnetic properties of standards (see, e.g., Davis, 1992b, 1993, 1995a, Myklebust, 1995, Ballantine, 1996).

2.2 System Modeling—The 'Weighing Equation'

When all necessary influence quantities have thus been identified, the functional relationship can be established, allowing an estimate of the measurand and associated standard uncertainty to be calculated from the appropriate experimental measurements. For the case of the one-pan electronic comparator we can write

$$(M_s - \rho_a V_s)g - (M_x - \rho_a V_x)g = F_s \qquad (2.2.1)$$

where M_s is the physical mass of the standard weight;

V_s is its volume;

M_x & V_x are the (unknown) mass and volume of the comparator's internal tare weight (only residual differences in apparent[1] mass are measured.)

ρ_a is the air density at the time of measurement;

F_s is the electromagnetic restoring force exerted by the comparator to compensate for the residual in-air mass difference (apparent mass difference) between the internal tare weight and the externally applied weight.

(See Jaeger & Davis (1984) for example). If the comparator indication for this measurement is W_s, it will be related to the force F_s by

$$W_s = kF_s \qquad (2.2.2)$$

where k is an instrument constant, fixed when the comparator is calibrated. We can interpret W_s as the apparent mass difference (or "weight-in-air" difference) between M_s and M_x. In doing this we can neglect a small correction factor, depending on the air density when the comparator was calibrated and also on the density of the calibration weight used (Schwartz, 1995), particularly since W_s is a residual mass difference and thus small. Any correction would then be less than the comparator's resolution. Repeating the measurement process with a test weight we have

$$(M_t - \rho_a V_t)g - (M_x - \rho_a V_x)g = F_t \qquad (2.2.3)$$

(For simplicity we assume air density remains unchanged between these two measurements). The comparator indication in this case would be W_t. If we now evaluate the difference between (2.2.1) & (2.2.3) we can eliminate the unknown tare weight term to obtain:

$$(M_s - \rho_a V_s)g - (M_t - \rho_a V_t)g = F_s - F_t \qquad (2.2.4)$$

We can define

$$\Delta W = W_s - W_t \qquad (2.2.5)$$

as the apparent mass difference between the standard and test weights. Usually this term is evaluated by a double-substitution comparison.

---

[1] Note: In this thesis, when we use the term 'apparent mass' we mean the resulting measured mass from a measurement in air of a particular density, before any buoyancy corrections are made. We will sometimes refer to 'apparent mass difference' as 'Weight-in-Air' difference.

Now it may occur that the centers of gravity of the standard and test weights do not coincide in which case we would have to write (2.2.4) as

$$(M_s - \rho_a V_s)g_s - (M_t - \rho_a V_t)g_t = \Delta W g_t \qquad (2.2.6)$$

where we have considered ΔW as a mass (in effect of density 8000 kgm^{-3} since comparators are usually calibrated for *conventional mass* (OIML IR 33 & Schwartz (1995)) which would 'balance' the force equation of Eq. (2.2.4). For convenience we have considered its centre of gravity to coincide with that of the test weight. Now we can note, following the treatment of gravitational effects in Almer & Swift (1975), that the gravitational force experienced by two masses, M_1 & M_2 is:

$$F_1 = \frac{GM_e M_1}{r_e^2} = M_1 g_1 \qquad (2.2.7a)$$

$$F_2 = \frac{GM_e M_2}{(r_e + d)^2} = M_2 g_2 \qquad (2.2.7b)$$

For M_e & r_e the mass and radius of the earth and d the distance:

$$(\text{center of mass})_1 \text{ - } (\text{center of mass})_2$$

Now if these forces are equal we find:

$$F_1 = F_2 \Rightarrow M_2 = M_1 + \frac{2dM_1}{r_e} + \frac{d^2 M_1}{r_e^2} \qquad (2.2.8a)$$

$$\& \quad \frac{M_2}{M_1} = \frac{g_1}{g_2} \qquad (2.2.8b)$$

$$\Rightarrow \quad \frac{g_1}{g_2} = 1 + \frac{2d}{r_e} + \frac{d^2}{r_e^2} \qquad (2.2.9)$$

but since $d \sim 1cm$ usually and $r_e \sim 6.4\times10^6 m$, the third term on the rhs of (2.2.9) is vanishingly small and can safely be ignored. Following (2.2.9), (2.2.6) can be re-expressed as:

$$(M_s - \rho_a V_s)\left(\frac{g_s}{g_t}\right) - (M_t - \rho_a V_t) = \Delta W \qquad (2.2.10)$$

Substituting (2.2.9) into (2.2.10) and re-arranging gives:

$$\begin{aligned} &(M_s - M_t) - \rho_a(V_s - V_t) + \frac{2dM_s}{r_e} - \frac{2d\rho_a V_s}{r_e} = \Delta W \\ &\Rightarrow (M_s - M_t) = \Delta W + \rho_a(V_s - V_t) - \frac{2dM_s}{r_e} + \frac{2d\rho_a V_s}{r_e} \end{aligned} \qquad (2.2.11)$$

The last term on the r.h.s. of (2.2.11) is some 4 to 6 orders of magnitude less than the others and thus can be neglected so that the functional form for our influence parameters can be expressed as:

$$\Delta M = \Delta W + \rho_a \Delta V - \frac{2dM_s}{r_e} \qquad (2.2.12)$$

where ΔM is the physical mass difference of the standards.

Equation (2.2.12) is the well-known, "weighing equation", although the extra term $\frac{2dM_s}{r_e}$, the correction due to different centers of gravity is not always included; however it can amount to some *3* or *4* μg which is easily measurable with modern comparators and so does need to be included for the most accurate work. M_s will be taken as the nominal mass of the standards for this calculations. In most cases the standards are either cylindrical, or very close to cylindrical so that *d*, the center of gravity difference, will be half the difference in height of the two cylinders. However, Davis (1995) has described a device allowing the centre of mass of a standard OIML mass (OIML, 1994) to be determined quite easily. Note that while it is not uncommon to employ sensitivity weights in the experimental process (e.g. Davis, 1987) leading to extra terms in Eq. (2.2.12), this has not been done here since they are not employed in the experimental system used in this work, to be described later.

One further influence we can add to (2.2.12) is that due to the temperature dependence of volumes of the standards (See Schwartz, 1991):

$$V(t) = V(20)\{1 + \alpha_t(t-20)\} \tag{2.2.13}$$

where $\alpha_t \propto t$ also, but can be taken to be constant over the narrow range of temperatures which will be encountered in the calibration laboratory. Also α_t is material dependent, but since we will only encounter stainless steel weights in this work, we can use a fixed value of α, and write (2.2.12) as:

$$\Delta M = \Delta W + \rho_a \Delta V(1 + \alpha(t-20)) - \frac{2dM_s}{r_e} \tag{2.2.14}$$

ΔM is the measurand of final interest for our purposes. For basic calibration, this is not strictly true since $\Delta M = M_s - M_t$ and m_t is the unknown to be estimated. In that case, what is required is $M_t = M_s - \Delta M$. However, what we are interested in is just *mass differences* which will be used in the Estimation Procedures to be described later, in order to estimate *mass values*.

2.3 Uncertainty Propagation

The analysis of Section 2.2 has described one form of the well-established Weighing Equation. Now we apply the unified approach to evaluating the standard uncertainty of (2.2.14). Note in particular that we will not split the influence quantities into those contributing randomly and systematically to the overall uncertainty. We must evaluate Eq. (1.4.7), restated here for convenience as:

$$s^2(\Delta M) = \sum_{i=1}^{N}\left(\frac{\partial f}{\partial x_i}\right)^2 s^2(x_i) + 2\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\frac{\partial f}{\partial x_i}\frac{\partial f}{\partial x_j}s(x_i, x_j) \tag{2.3.1}$$

where the symbols have their usual meanings and $\Delta M = f(x_1,...x_N)$. From (2.2.14), this can be written as $\Delta M = f(\Delta W, \rho_a, V_s, V_t, t, d)$. We must write $\Delta V = (V_s - V_t)$ since there will be a standard uncertainty associated with each volume but Δw is just an instrument indication. For M_s in Eq.(2.2.14) we can use the nominal value of the standard since typical deviations of a few *mg* from this nominal value will make insignificant differences of the order of $\sim 10^{-5}$ μg to the Gravitational Correction. Further, r_e, the earth radius can be treated as a constant here for similar reasons.

Before proceeding to the evaluation of (2.3.1), we need to consider covariance elements which will result. On a mathematical level, the possible covariances which can arise among the influence quantities in the functional form of Δm are:

$s(\Delta W, \rho_a)$, $s(\Delta W, V_s)$, $s(\Delta W, V_t)$, $s(\Delta W, t)$, $s(\Delta W, d)$;
$s(\rho_a, V_s)$, $s(\rho_a, V_t)$, $s(\rho_a, t)$, $s(\rho_a, d)$;
$s(V_s, V_t)$, $s(V_s, t)$, $s(V_s, d)$;
$s(V_t, t)$, $s(V_t, d)$;
$s(t, d)$.

It is very important when considering correlations to consider whether or not they are physically meaningful: it is always possible to mathematically evaluate a correlation coefficient or covariance among sets of data, but one must always check whether such a relationship can be physically justified. It is easy to make pronouncements, based on such statistical analyses, which do not have any foundation in physical reality.

We can immediately say that all correlations involving the volumes are zero since the volumes are determined independently by a hydrostatic weighing experiment at another time and another place. (See, e.g., Bowman (1967), Spieweck & Bettin (1992), Heierli (1997)) The possibility of correlations *between* the two volumes may be speculated, but as a rule, information about this is never available to the experimenter doing mass calibrations and so cannot be included. Similarly, the covariance $s(\rho_a, d)$ can be dismissed, as can $s(\Delta w, d)$ and $s(t, d)$. Since $\rho_a \propto t$–among other variables–a covariance between these two does not arise, but rather, there is a direct contribution to the combined variance of ρ_a from the variance of t, of which more will be said in Chapter 3. Assuming steady state conditions during the measurement period–which is reasonable in a high accuracy laboratory – there will be no correlation between the weight difference, Δw, and t or ρ_a, which are measured/calculated by other instrumentation.

The important point of the foregoing considerations is that we have taken on board all information that is to hand about the experimental process. In the absence of any information to the contrary we have no reason to assume that a covariance exists between the volumes, for example. If, however, the volume calibration data indicated

such a covariance, then it would of course have to be included. The same applies to the other variables in Eq. (2.2.14). The result is that, having considered the physical situation, in the absence of any information to the contrary, we can assign a zero value to the covariances among the input parameters. Again this is an example of implementing a philosophy of consistent reasoning in the analysis. So now (2.3.1) can be evaluated to give:

$$s^2(\Delta M)=\left(\frac{\partial f}{\partial \Delta W}\right)^2 s^2(\Delta W)+\left(\frac{\partial f}{\partial \rho_a}\right)^2 s^2(\rho_a)+\left(\frac{\partial f}{\partial V_s}\right)^2 s^2(V_s) \\ +\left(\frac{\partial f}{\partial V_t}\right)^2 s^2(V_t)+\left(\frac{\partial f}{\partial t}\right)^2 s^2(t)+\left(\frac{\partial f}{\partial d}\right)^2 s^2(d) \tag{2.3.2}$$

and using (2.2.14) for the functional form of ΔM we obtain the expression in (2.3.3).

$$\begin{aligned} s^2(\Delta M)= \quad & s^2(\Delta W) \\ & +s^2(\rho_a)(V_s-V_t)^2(1+\alpha(t-20))^2 \\ & +\left(s^2(V_s)+s^2(V_t)\right)\rho_a^2(1+\alpha(t-20))^2 \\ & +s^2(t)\rho_a^2(V_s-V_t)^2\alpha^2 \\ & +s^2(d)\left(\frac{-2M_s}{r_e}\right)^2 \end{aligned} \tag{2.3.3}$$

The positive square root of (2.3.3) is then the combined standard uncertainty of Δm. In practice, the last two terms, due to the systematic corrections, will evaluate to 5 or 6 orders of magnitude less than the others and thus contribute negligibly to the overall uncertainty term. For example, in Davis (1995b), it is shown that the centre of mass can be determined to perhaps 3 μm which would lead to an uncertainty contribution of around $1\times10^{-3}\ \mu g$ for a *1 kg* standard.

We now need to consider the possibility of correlations *between* two separate evaluations ΔM_i and ΔM_j, which could arise as a result of common influence quantities in each. Recall that the correlation can be evaluated from:

$$\begin{aligned} & s(y_j,y_k)=\sum_{i=1}^{N}\frac{\partial f}{\partial x_i}\frac{\partial g}{\partial x_i}s^2(x_i) \\ & \text{for} \quad y_j=f(x_1,\ldots.x_N) \\ & \& \quad y_k=g(x_1,\ldots x_N) \end{aligned} \tag{2.3.4}$$

where not all the influence quantities may necessarily occur in both f & g. Covariance terms will arise only as a result of those terms which are common to both. In the present case, from Eq. (2.2.14), we may have, for example:

$$\Delta M_i = \Delta W_i + \rho_{a_i}\left(V_l - V_m\right)\left(1+\alpha\left(t_i - 20\right)\right) - \frac{2d_i M_{s_i}}{r_e}$$
$$\Delta M_j = \Delta W_j + \rho_{a_j}\left(V_n - V_p\right)\left(1+\alpha\left(t_j - 20\right)\right) - \frac{2d_j M_{s_j}}{r_e} \qquad (2.3.5)$$

Clearly, ΔW_i & ΔW_j never both occur in the same function, the same applying to ρ_i & ρ_j and also t_i & t_j. α is considered constant since its variability will have negligible influence so it cannot contribute any covariance either. Thus the only possibility for covariances between ΔM_i & ΔM_j arises as a result of weights being common to both comparisons. For example, if $\Delta M_i = M_1 - M_2$ & $\Delta M_j = M_1 - M_3$,we can see that M_1 is used in both comparisons and so V_1 will appear in both equations in (2.3.5) above. Thus we find that the covariance term can be expressed as:

$$s\left(\Delta M_i, \Delta M_j\right) = \sum_{k=1}^{P}\left(\frac{\partial \Delta M_i}{\partial V_k}\right)\left(\frac{\partial \Delta M_j}{\partial V_k}\right)s^2\left(V_k\right) \qquad (2.3.6)$$

'P' in (2.3.6) is the total number of weights used in the two comparisons and the derivative product is only non-zero in cases where a weight is used in both comparisons.

So, by applying the ISO Procedure to the Weighing Equation, we have been able to generate two equations {(2.3.3) & (2.3.6)} which provide all the necessary information to evaluate the various uncertainties resulting from the comparison experiments. This has been done in a unified manner with respect to the treatment of the various influence terms and we have sought to uphold demands on consistent reasoning in the analysis. With the functional relationship established as shown in Eq. (2.2.14), we only need to find variance components for each influence quantity in order to complete the error propagation analysis. We have seen in Chapter 1 how Maximum Entropy considerations allow this to be done in a "maximally unbiased" manner. For example, quantities like Δw which are estimated by repeated measurements can be considered to have a Normal Distribution, since a mean value and variance are available. Quantities from other calibrations (such as volume determinations for example) should have been processed according to ISO principles, in which case a standard uncertainty will be available which we can be confident in inserting into subsequent calculations. However, if it is not clear how the quoted uncertainty was determined, we must use Maximum Entropy considerations to assign the least committal distribution that is compatible with the supplied data. Very often this will simply be a Uniform distribution.

Before proceeding to look at combinations of comparisons, as used in disseminating the mass scale, we must first consider the uncertainty propagation analysis for the air density, ρ_a.

3. The Evaluation of Air Density

3.0 Summary

Here the model parameterisation for mass calibration of Chapter 2 is continued. The calculations relating to air density are included here and are kept separate from the considerations of Chapter 2 for clarity. However, the two components form an integrated process, as can be seen in Fig. (3.0.1) below which illustrates the method. The development of the air density equation outlined in this chapter is the accepted approximate relationship for use in a typical Standards Laboratory, sometimes referred to as the 'BIPM Formula'. What has been done here however, is to ensure that it is treated according to the Unified Approach to uncertainty analysis as outlined in Chapter 1. The functional relationship given is for direct measurements of temperature, barometric pressure, relative humidity and possibly CO_2 level; as well as imported values for the Gas Constant R, and the molar masses of moist and dry air.

The Gaussian procedure is applied to this functional form in order to obtain the standard uncertainty of the air density estimate. To evaluate the partial derivatives (sensitivity coefficients) of the functional form is a little difficult since within it there are several polynomial terms with implicit dependencies on the influence quantities and thus the sensitivity coefficients are themselves functions of the influence quantities to be measured. In Appendix 1 details are shown of computer simulations by means of which the 'typical' values of the sensitivity coefficients used in this chapter were evaluated.

In conclusion, a generalised expression for the variance of the air density estimate is given from which the actual uncertainty can be evaluated if the details of the equipment used are available. An example is given for the equipment used in the experimental parts of this research.

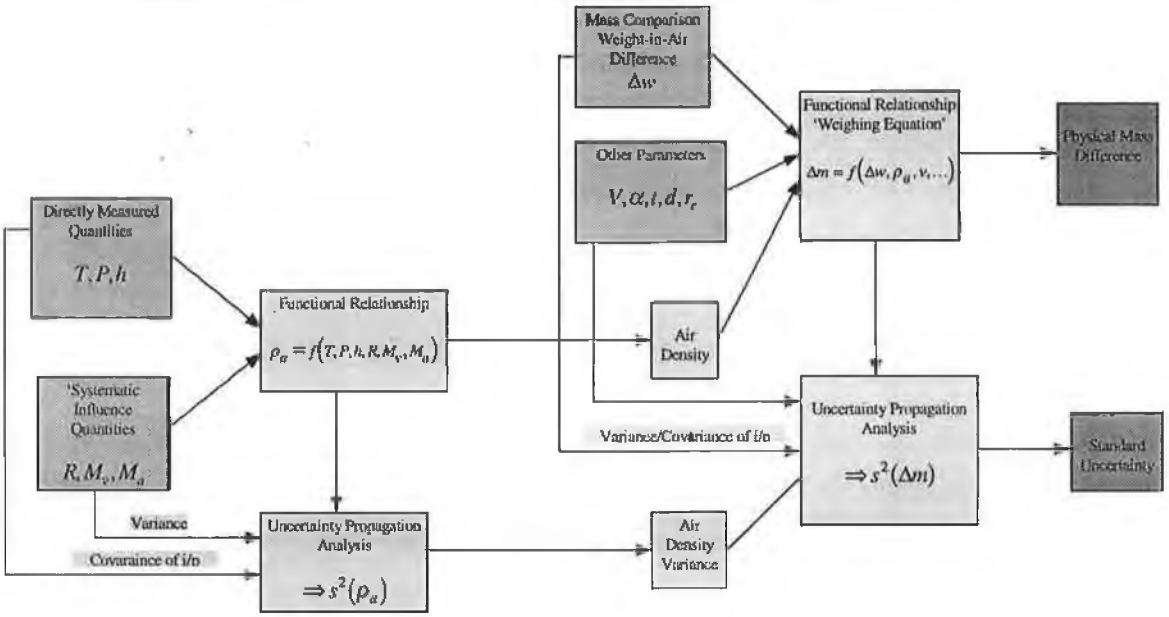


Fig. 3.0.1: Schematic of complete data processing for mass comparison calibrations

3.1 The Functional Relationship

The procedure for evaluating air density has been described in the literature e.g. Giacomo (1981), Davis (1992a), Jones (1978). A brief summary of the calculation follows. Starting from the ideal gas law:

$$PV = nRT \tag{3.1.1}$$

for a gas of volume V at pressure P and thermodynamic temperature T, containing n moles. R is the molar gas constant. For a real gas one has:

$$PV = nZRT \tag{3.1.2}$$

Z being the compressibility factor. Since the gas density is $\rho = m/V$ if its mass is m, we can say $\rho = \frac{nM}{V}$ since $nM = m$ for M the molar mass. Then, from (3.1.2):

$$\rho = \frac{PM}{ZRT} \tag{3.1.3}$$

However, air is composed of both dry air and water vapour. If the mole fraction of water vapour is x_v and its molar mass is M_v, while the molar mass of dry air is M_a, we have:

$$M = M_a\left(1 - x_v\left(1 - M_v/M_a\right)\right) \equiv (1 - x_v)M_a + x_v M_v \tag{3.1.4}$$

In this case the density can be expressed as:

$$\boxed{\rho_a = \frac{PM_a}{ZRT}\left(1 - x_v\left(1 - \frac{M_v}{M_a}\right)\right)} \tag{3.1.5}$$

Eq. (3.1.5) is the essential equation for calculating the density of moist air from measurements of P, T and a knowledge of M_v, M_a Z & x_v. M_a is obtained from a knowledge of the constituents of dry air and their relative abundances:

$$M_a = \frac{\sum_{i=1}^{N} x_i m_i}{\sum x_i} \tag{3.1.6}$$

where x_i is the mole fraction of the i^{th} gas molecule having molar mass m_i. Tabulated values {e.g. Davis (1992a), Jones (1978)} lead to (3.1.6) having the value

$$M_a = 28.9635 \times 10^{-3} \text{ kg / mol.} \tag{3.1.7}$$

assuming, however, that the level of CO_2 is constant, and indeed that $x_{CO_2} = 0.0004$. This of course may not be so, and x_{CO_2} might be measured in the lab, in which case it is possible to provide an adjustment to M_a to account for measured departures of x_{CO_2} from the assumed reference level. This can be achieved using the working approximation that the sum of O_2 and CO_2 in the air remains constant, (Giacomo (1981), Jones (1978)) that is more CO_2 implies less O_2 so that:

$$x_{CO_2} + x_{O_2} = \text{constant} = 0.20979 \tag{3.1.8}$$

where known tabulated values for the abundance of each have been used. Then:

$$M_{O_2} x_{O_2} + M_{CO_2} x_{CO_2} = 31.9988 x_{O_2} + 44.0098 x_{CO_2} \tag{3.1.9}$$

can be written, from (3.1.8), as:

$$M_{O_2} x_{O_2} + M_{CO_2} x_{CO_2} = 12.011 x_{CO_2} + 6.7130 \tag{3.1.10}$$

so:

$$\delta(M_a) = \delta\left[M_{O_2} x_{O_2} + M_{CO_2} x_{CO_2}\right] = 12.011 \delta\left(x_{CO_2}\right) \tag{3.1.11}$$

—i.e. assuming all the other constituents remain constant, or at least do not change by anything other than infinitesimal amounts, then the variations in M_a will be due to CO_2 variation, and thus will be given by 12.011 times the variation in CO_2 abundance. 12.011 being, of course, the atomic weight of carbon. We can define

$$\delta\left(x_{CO_2}\right) \equiv \left(x'_{CO_2} - 0.0004\right) \tag{3.1.12}$$

where $0.0004 = x^{ref}_{CO_2}$, a reference CO_2 level, and x'_{CO_2} is a measured CO_2 level. In that case we have $M'_a = M^{ref}_a + \delta(M_a)$ where M^{ref}_a is the dry air molar mass for reference CO_2 level. Thus:

$$\boxed{M'_a = 28.9635 \times 10^{-3} \text{ kg / mol. } + 12.011\left(x'_{CO_2} - 0.0004\right) \times 10^{-3} \text{ kg / mol.}} \tag{3.1.13}$$

However, in our case, we will use (3.1.7) rather than (3.1.13) since facilities for measuring CO_2 content of the air are not available in the later experimental work.

M_v is simply the mass of water vapour in the air, i.e.:

$$M_v = 2 \times M_H + M_o \tag{3.1.14}$$

So, with (3.1.7) & (3.1.14), (3.1.5) becomes:

$$\rho_a = \frac{PM_a}{ZRT}(1 - 0.378x_v) \tag{3.1.15}$$

The evaluation of x_v, the mole fraction of water vapour, proceeds from a measurement of either relative humidity, or dew point temperature (Giacomo, (1981),Wexler & Wildhack (1965)) In our case, the % Relative Humidity is measured, and we note relative humidity is defined as the ratio:

$$h = {}^{x_v}\!/_{x_{sv}} \tag{3.1.16}$$

where x_{sv} is the mole fraction of water vapour in saturated moist air at the same temperature and pressure. Now x_{sv} can be calculated from:

$$x_{sv} = \frac{f(p,t)P_{sv}(t)}{P} \tag{3.1.17}$$

for $P_{sv}(t)$ the saturation vapour pressure (Wexler, 1976) and $f(p,t)$ a correction called the 'enhancement factor'. (Greenspan, 1976, Hyland, 1975) Thus $x_v = hx_{sv}$ from (3.1.16) and from (3.1.17):

$$x_v = \frac{hfP_{sv}}{P} \tag{3.1.18}$$

where h is expressed as a fraction, rather than a percentage.

Finally, approximate polynomials have been developed (Giacomo (1981), Davis (1992a)) for P_{sv}, f and Z, which are valid over the narrow range of standard conditions encountered in the calibration laboratory:

$$\begin{aligned} f &= \alpha + \beta P + \gamma t^2 \\ P_{sv} &= 1Pa \times \exp\left(AT^2 + BT + C + DT^{-1}\right) \\ Z &= 1 - \frac{P}{T}\left\{a_o + a_1 t + a_2 t^2 + (b_0 + b_1 t)x_v + (c_0 + c_1)x_v^2\right\} + \left(\frac{P}{T}\right)^2 \left(d + ex_v^2\right) \end{aligned} \tag{3.1.19}$$

Equations (3.1.19), (3.1.18) & (3.1.15) allow for an evaluation of air density. We use $T=(273.15 + t)$ K when t is measured in $°C$. Eq. (3.1.7) provides the value of M_a and R is taken from Cohen & Taylor (1987) as $R = 8.314510$ $J/Mol.k$.

3.2 Uncertainty Propagation

We must now evaluate a Standard Uncertainty for ρ_a. From an analysis of the above expressions, it can be seen that:

$$\rho_a = f(T, P, h, R, M_v, M_a) \tag{3.2.1}$$

We do not have any information to suggest possible covariances/correlations between these influence quantities so we can neglect them in the analysis, in particular since separate instrumentation is used to measure each of T, P and h, and R, M_v and M_a come from entirely separate analyses. Thus we can say:

$$s^2(\rho_a) = \left(\frac{\partial \rho_a}{\partial T}\right)^2 s^2(T) + \left(\frac{\partial \rho_a}{\partial P}\right)^2 s^2(P) + \left(\frac{\partial \rho_a}{\partial h}\right)^2 s^2(h) + \left(\frac{\partial \rho_a}{\partial R}\right)^2 s^2(R) + \left(\frac{\partial \rho_a}{\partial M_v}\right)^2 s^2(M_v) + \left(\frac{\partial \rho_a}{\partial M_a}\right)^2 s^2(M_a) \qquad (3.2.2)$$

The partial derivatives can be evaluated if the various constants are inserted and appropriate values of T, P & h chosen. While the partial derivatives are still functions of T, P and h, it was found that for a wide range of values, covering the range of interest which could possibly be encountered in the laboratory, constant values could be taken (see Appendix 1). The values in (3.2.3a), below are adopted. $s(M_a)/M_a$ is given as comprising the terms 4×10^{-5} random & 3×10^{-5} systematic; while $s(M_v)/M_v$ is declared to be systematic with value 5×10^{-4} (See Jones (1978) for details) In the Unified Approach being developed in this thesis we do not need such distinctions, so with no further information available and seeking to maintain a consistent analysis, we can only consider the systematic components to represent limits of possible parameter values which leads to a Uniform distribution from the Maximum Entropy considerations of Chapter 1. Thus we obtain the figures quoted in (3.2.3b) below.

$$\begin{array}{ll}
\frac{\partial \rho_a}{\partial t} = -4.4\times10^{-3} & \mathrm{K^{-1}.mg.cm^{-3}} \\
\frac{\partial \rho_a}{\partial P} = +11.2\times10^{-6} & \mathrm{Pa^{-1}.mg.cm^{-3}} \\
\frac{\partial \rho_a}{\partial h} = -10.5\times10^{-3} & \mathrm{mg.cm^{-3}} \\
\frac{\partial \rho_a}{\partial R} = -0.144 & \mathrm{J^{-1}.mol.K.mg.cm^{-3}} \\
\frac{\partial \rho_a}{\partial M_v} = +0.5 & \mathrm{kg^{-1}.mol.mg.cm^{-3}} \\
\frac{\partial \rho_a}{\partial M_a} = +41.0 & \mathrm{kg^{-1}.mol.mg.cm^{-3}}
\end{array} \qquad (3.2.3a)$$

$$\begin{array}{ll}
s(R) = 7\times10^{-5} & \mathrm{J.K^{-1}.mol^{-1}} \\
\frac{s(M_a)}{M_a} = 4.36\times10^{-5} & \\
\Rightarrow s(M_a) = 1.26\times10^{-6} & \mathrm{kg.mol^{-1}} \\
\frac{s(M_v)}{M_v} = 2.9\times10^{-4} & \\
\Rightarrow s(\mathrm{M_v}) = 5.22\times10^{-6} & \mathrm{kg.mol^{-1}}
\end{array} \qquad (3.2.3b)$$

With the data in (3.2.3), (3.2.2) becomes:

$$s^2(\rho_a)=\left(\frac{\partial \rho_a}{\partial t}\right)^2 s^2(t)+\left(\frac{\partial \rho_a}{\partial P}\right)^2 s^2(P)+\left(\frac{\partial \rho_a}{\partial h}\right)^2 s^2(h)+2.76\times10^{-9}\left[\text{mg.cm}^{-3}\right]^2 \quad (3.2.4)$$

For the equipment used in this research, described in Chapter 11 the following data is available from calibration certificates:

$$s(t)=0.06 \quad \text{K}$$
$$\text{s(P)}=15.0 \quad \text{Pa}$$
$$\text{s(h)}=2\%=0.02$$

Thus:

$$s^2(\rho_a)=\left(142*10^{-9}+2.76\times10^{-9}\right)\left[\text{mg.cm}^{-3}\right]^2 \quad (3.2.5a)$$
$$\cong 145\times10^{-9}\ \left[\text{mg.cm}^{-3}\right]^2$$

or:

$$s(\rho_a)=3.8\times10^{-4}\ \ \text{mg.cm}^{-3} \quad (3.2.5b)$$

Eq. (3.2.5a) indicates that the components due to the measurements of T, P & h during the calibration experiment supply a much greater amount of the final standard uncertainty than do those due to the imported data from other evaluations—i.e. R, M_v, and M_a.

One final point we must check is the possibility of correlations *between* one evaluation of air density and another, since the three terms R, M_v & M_a, are common to each. Although we have earlier stated that such correlations do not exist (Sec. 2.3), we wish here to verify that they are negligible. The covariances can be calculated from:

$$s(\rho_{a_1},\rho_{a_2})=\left(\frac{\partial \rho_{a_1}}{\partial R}\right)\left(\frac{\partial \rho_{a_2}}{\partial R}\right)s^2(R)+\left(\frac{\partial \rho_{a_1}}{\partial M_v}\right)\left(\frac{\partial \rho_{a_2}}{\partial M_v}\right)s^2(M_v)+\left(\frac{\partial \rho_{a_1}}{\partial M_a}\right)\left(\frac{\partial \rho_{a_2}}{\partial M_a}\right)s^2(M_a) \quad (3.2.6)$$

by Eq. (2.3.4) earlier. But of course the expression for ρ_{a_1} is the same as that for ρ_{a_2}, so in fact we have for the covariance:

$$s(\rho_{a_1},\rho_{a_2})=\left(\frac{\partial \rho_a}{\partial R}\right)^2 s^2(R)+\left(\frac{\partial \rho_a}{\partial M_v}\right)^2 s^2(M_v)+\left(\frac{\partial \rho_a}{\partial M_a}\right)^2 s^2(M_a) \quad (3.2.7)$$

which from our data in (3.2.3) may be evaluated as $\cong 3\times10^{-9}\left[\text{mg.cm}^{-3}\right]^2$. This term will be very insignificant compared to the other correlations described in Eq. (2.3.6) due to volume elements common to two mass comparisons.

An objection might be raised that the functional form of the "Weighing Equation", (2.2.14), should be expanded to include the functional form of the air density equation, (3.1.15). The complete uncertainty evaluation could then be processed with the expanded equation. However, this will give the same results as the present method which is to be preferred since the separation allows a clearer discussion of the form of each equation, and also easier analysis of the various

contributions to the standard uncertainty of each. This of course assumes that there are no covariances between the influence quantities of Eq (2.2.14) and those of Eq. (3.1.15). The possibility of such has been disposed of earlier.

Thus Eqs (2.2.14) and (3.1.15) allow a complete evaluation of the measurand—the mass difference of two or more standards, while eqs. (2.3.3) & (3.2.5) allow an evaluation of the standard uncertainty of this estimate.

We must now proceed to look at the case of multiple combinations of such mass differences, the evaluation of which is necessary to allow statistical parameter estimation of the mass *values* [as opposed to *differences*] of the standards, which is the final goal of our investigations. To establish the input information in such cases we require less cumbersome tools than the ones so far developed, and we now proceed to develop the Unified Model in a more elegant formalism using matrix algebra.

****

4. Mass Dissemination / "Within -Group Comparisons"

4.0 Summary

In this chapter we introduce the Group Comparison Calibrations used in mass dissemination as well as other calibration exercises. These involve many comparisons being carried out among a group of standards resulting in a set of weighing equations each of which will yield a mass difference term. We introduce matrix algebra as a convenient way to represent the data and also to evaluate the measurands and their standard uncertainties.

We first show how the situation can be expressed in matrix theory, and give the Weighing Equation in this form. We then introduce the important statistical tools needed in uncertainty analysis, in particular the *covariance matrix* which encapsulates all the variance and covariance information about a corresponding vector of parameters and also the *Jacobian* which is a matrix of partial derivatives of a functional relationship among a set of influence quantities.

Before developing the theory for the Weighing Equation, an example is given using electrical measurements, taken from the ISO Guide (ISO, 1993) but re-worked here in matrix notation.

The evaluation of the Covariance Matrix of the Weighing Equation is then developed resulting in a single equation from which the complete covariance matrix of the set of measurands (mass differences from the comparison calibrations) can be easily evaluated.

The great advantage of the technique is that all the influence information is then included in one measurand vector and one covariance matrix. We will see in Chapters 5 to 8, where we deal with parameter estimation techniques, just how convenient this is. Other approaches have more difficulty in including all available information, in particular the secondary influence quantities such as those due to the buoyancy correction, see for example Schwartz (1991). The present technique has not been widely used in mass calibration to date, but is to be highly recommended on account of both its mathematical conciseness and indeed its unified approach to uncertainty analysis.

Fig. (4.0.1) below is a schematic outline of the analysis process used. The various terms are explained in detail in the body of the chapter.

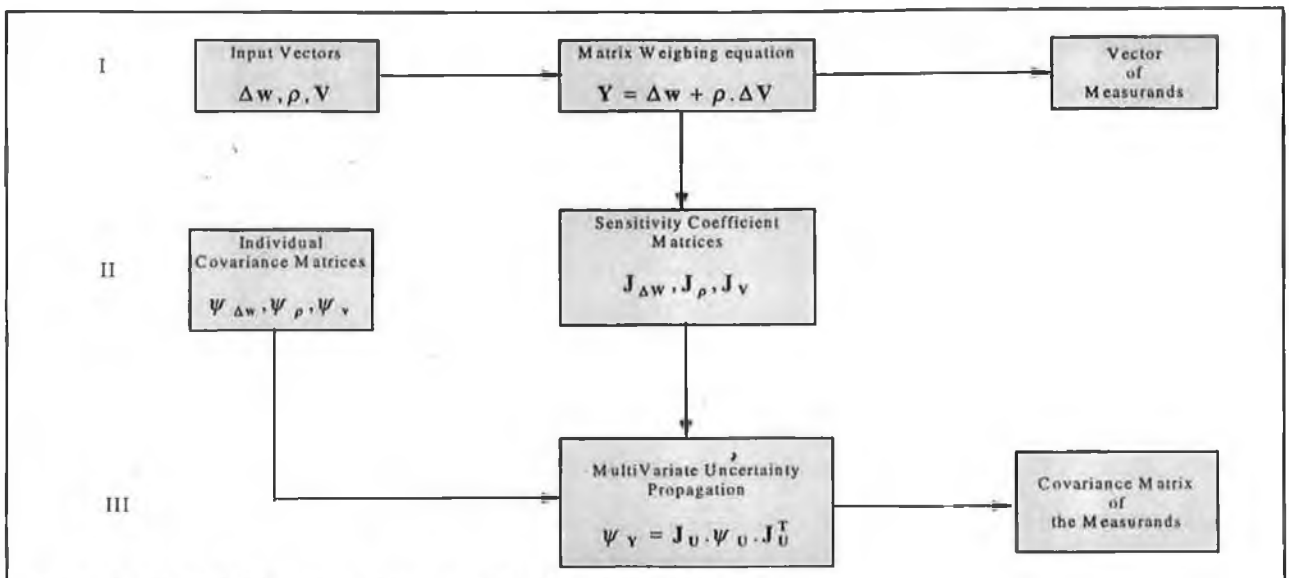


Fig. 4.0.1: Unified Uncertainty Analysis in the MultiVariate Case

4.1 Introduction

One of the crucial features about comparison calibrations, from an experimental point of view, is that the comparisons are carried out between nominally equal standards. Thus only residual differences are measured with resulting greater accuracy. In establishing a mass scale, we are starting from a prototype of nominal value 1kg. Thus comparisons of multiples and sub-multiples of this unit require combinations of standards to be built up in order to ensure comparisons among units of nominally equal value. These are usually called "Within-Group" Calibrations, since there is a group of unknowns, and usually only one known standard. At this point we define our parameters to be the mass *values* of these various artifacts. Note carefully that these parameters are to be distinguished from the measurands discussed previously. As was stated then, the measurands are the mass *differences* which are determined from experiment. How the parameters are calculated from the corrected experimental realizations of the measurands is the subject of later sections of this work. At this stage we are just interested in evaluating the estimates of the measurands and their standard uncertainties. (For further discussion of mass dissemination see Kochsiek (1984), Davis (1985) and also Prowse (1982)).

4.2 MultiVariate Functional Relationship

Consider the following example of a typical set of comparisons:

$$\begin{aligned} \Delta m_1 &= m_1 - m_2 \quad\quad = y_1 \\ \Delta m_2 &= m_1 \quad\quad - m_3 = y_2 \\ \Delta m_3 &= \quad\quad m_2 - m_3 = y_3 \end{aligned} \tag{4.2.1}$$

The above indicates the possible comparisons which could be carried out with three mass standards of value m_1, m_2 & m_3. If we introduce the weighing equation, (2.2.14), we realise that (4.2.1) becomes:

$$\begin{aligned} y_1 &= \Delta w_1 + \rho_1(v_1 - v_2)(1+\alpha(t_1-20)) - \frac{2d_1 m_1^n}{r_e} \\ y_2 &= \Delta w_2 + \rho_2(v_1 - v_3)(1+\alpha(t_2-20)) - \frac{2d_2 m_2^n}{r_e} \\ y_3 &= \Delta w_3 + \rho_3(v_2 - v_3)(1+\alpha(t_3-20)) - \frac{2d_3 m_3^n}{r_e} \end{aligned} \qquad (4.2.2)$$

In (4.2.2), m_i^n is the i^{th} nominal mass in each case. Since deviations from nominal value are always of the order of *mg* or μg, the difference to the gravitational correction resulting from using nominal rather than actual mass value will be negligible. Without the weighing equation, (4.2.1) can also be expressed in matrix terms as:

$$\mathbf{Y} = \mathbf{X}\beta \qquad (4.2.3)$$

For **Y** an *n x 1* vector of corrected realised data, β a *p x 1* vector of parameters and **X** an *n x p* design matrix describing the form of the comparisons. Eq. (4.2.3) can be referred to as the "system model" for the measurement process. In the example above, we would have (where $b_i \equiv m_i$):

$$\mathbf{X} = \begin{bmatrix} 1 & -1 & 0 \\ 1 & 0 & -1 \\ 0 & 1 & -1 \end{bmatrix}; \quad \beta = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} \& \quad \mathbf{Y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}$$

Continuing, it will be noted that (4.2.2) can be expressed in matrix notation (Bich *et al* 1993/94) as:

$$\mathbf{Y} = \Delta\mathbf{w} + \{(1-20\alpha)\mathbf{I_n} + \alpha\,\mathbf{T}\}\rho\mathbf{X}\mathbf{V} - \left(\frac{2}{r_e}\right)\mathbf{D}\mathbf{M_n} \qquad (4.2.4)$$

which is the matrix form of the weighing equation, where:

Y is an *n* x 1 vector of measurand estimates.

Δw is an *n* x 1 vector of experimentally realised (i.e. uncorrected) quantities.

$\rho = \text{diag}\{\rho_\mathbf{a}\}$ is an *n* x *n* matrix where $\rho_\mathbf{a}$ is an *n* x 1 vector of air densities.

X is an *n* x *p* matrix giving the design scheme for the comparisons.

V is an *p* x 1 vector of volumes of the standards

$\mathbf{I_n}$ is an identity matrix of order *n*

$\mathbf{T} = \text{diag}\{\mathbf{t}\}$ is an *n* x *n* matrix where **t** is an *n* x 1 vector of air temperatures.

α is a scalar constant, the volume expansion coefficient of stainless steel.

$\mathbf{D} = \text{diag}\{\mathbf{d}\}$ is an *n x n* matrix where **d** is an *n* x 1 vector of center of gravity differences between the standard and test ensembles.

$\mathbf{M_n}$ is an *n* x 1 vector of nominal mass values for each of the *n* comparisons involved in the calibration exercise.

Observe that, leaving aside the volume correction factor, the term **XV** generates the required volume differences for each comparison. It is necessary to re-express ρ_a as a diagonal matrix in order to facilitate the required matrix multiplication—the other alternative would be to use the Hadamard matrix formalism, but we will retain the conventional expressions for clarity and ease of understanding.

4.3 Important Statistical Terms in Matrix Form

We now wish to evaluate the variances and covariances of the data described by (4.2.4). First we note that the scalar forms of expectation value and variance/covariance must now be supplanted by their vector/matrix equivalents. (See Allisy). For example, for a column vector $\mathbf{A}_{n\times 1}$,

$$E[\mathbf{A}]=\begin{bmatrix} E(a_1) \\ E(a_2) \\ . \\ . \\ E(a_n) \end{bmatrix} \quad \text{for} \quad \mathbf{A}=\begin{bmatrix} a_1 \\ a_2 \\ . \\ . \\ a_n \end{bmatrix} \tag{4.3.1}$$

Similarly to $s^2(y)=E\left[y-E[y]\right]^2$, we can define the variance-covariance matrix by:

$$\mathbf{cov}(\mathbf{A}_{n\times 1})=E\left\{(\mathbf{A}-E[\mathbf{A}])(\mathbf{A}-E[\mathbf{A}])^{\mathbf{T}}\right\} \tag{4.3.2}$$

the 'T' denoting matrix transposition. Then (4.3.2) can be expanded to give:

$$\begin{aligned}\mathbf{cov}(\mathbf{A}) &= E\left\{\begin{bmatrix} a_1-E[a_1] \\ a_2-E[a_2] \\ . \\ . \\ a_n-E[a_n] \end{bmatrix}\cdot\begin{bmatrix} a_1-E[a_1] & a_2-E[a_2] & . & . & a_n-E[a_n] \end{bmatrix}\right\} \\ &= E\left\{\begin{bmatrix} \{a_1-E[a_1]\}^2 & \{a_1-E[a_1]\}\{a_2-E[a_2]\} & . & . & . \\ \{a_2-E[a_2]\}\{a_1-E[a_1]\} & \{a_2-E[a_2]\}^2 & & & \\ . & & . & & \\ . & & & . & \\ . & & & & \{a_n-E[a_n]\}^2 \end{bmatrix}\right\} \\ &= \begin{bmatrix} s^2(a_1) & s(a_1,a_2) & . & . & s(a_1,a_n) \\ s(a_2,a_1) & s^2(a_2) & & & \\ . & & . & & \\ . & & & . & \\ s(a_n,a_1) & & & & s^2(a_n) \end{bmatrix}\end{aligned} \tag{4.3.3}$$

Thus the covariance matrix of **A**, **cov(A)**, consists of diagonal elements giving the variance components of **A**, while the off-diagonal elements give the covariances between the components, so that for example, if $b=a_1+a_2$ then

$s^2(b) = s^2(a_1) + s^2(a_2) + 2\left(s(a_1, a_2)\right)$. The terms for this can then be easily extracted from the matrix in (4.3.3). We will see in succeeding sections that this matrix plays a very important role in the Parameter Estimation Techniques we will investigate later.

One important property of the covariance matrix which we shall need is the covariance of linear combinations. Consider:

$$\mathbf{Z} = \mathbf{XY} \tag{4.3.4}$$

where $\mathbf{X}$ is a constant matrix and $\mathbf{Y}$ is a vector of subjective information (i.e. finite, or limited-accuracy information, hence needing degrees of belief or distributional information assigned to each of its components.) We wish to evaluate **cov(Z)**:

$$\begin{aligned} \mathbf{cov}(\mathbf{Z}) &= E\left[\{\mathbf{Z} - E[\mathbf{Z}]\}\{\mathbf{Z} - E[\mathbf{Z}]\}^{\mathrm{T}}\right] \\ &= E\left[\{\mathbf{XY} - E[\mathbf{XY}]\}\{\mathbf{XY} - E[\mathbf{XY}]\}^{\mathrm{T}}\right] \\ &= E\left[\mathbf{X}\{\mathbf{Y} - E[\mathbf{Y}]\}\{\mathbf{Y} - E[\mathbf{Y}]\}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\right] \end{aligned} \tag{4.3.5}$$

where, for a matrix product **AB**, one has $(\mathbf{AB})^{\mathrm{T}} = \mathbf{B}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}$. Since $\mathbf{X}$ is a constant, Eq. (4.3.5) can be further expressed as:

$$\begin{aligned} &\mathbf{X}\, E\left[\{\mathbf{Y} - E[\mathbf{Y}]\}\{\mathbf{Y} - E[\mathbf{Y}]\}^{\mathrm{T}}\right]\mathbf{X}^{\mathrm{T}} \\ &= \mathbf{X}\,\psi_{\mathbf{Y}}\,\mathbf{X}^{\mathrm{T}} \end{aligned} \tag{4.3.6}$$

where $\psi_{\mathbf{Y}} = \mathbf{cov}(\mathbf{Y})$ from (4.3.2) above. This important relation will appear frequently in what follows.

In most statistical treatments, one considers the vector $\mathbf{Y}$ to contain additive, zero-mean errors such that $\mathbf{Y} = \eta + \varepsilon$ where $E[\varepsilon] = \mathbf{0}$; $\mathbf{cov}[\eta] = \mathbf{0}$ & $\mathbf{cov}[\varepsilon] = \psi$. In other words η is the "true value" and ε represents the random errors on the measurements. In that case the distributions are attached to the errors rather than to the parameters. However, as has been repeatedly pointed out in the unified approach, we do not know η and can never know it, so it is not really meaningful to use it. Therefore we consider $\mathbf{Y}$ to be a subjective estimate based on whatever information has been to hand and we consider our distributional information to be based around our estimate, $\mathbf{Y}$. Hence in the following we do not emphasise the role of random errors and can leave the analysis much more general. (Refer again to Chapter 1 for further comments on the meaning of subjective information in parameter estimation).

Let us now consider:

$$\mathbf{Z} = f(\mathbf{U}) \tag{4.3.7}$$

where $\mathbf{U}^{\mathrm{T}} = [u_1 \quad u_2 \quad . \quad . \quad u_n]$ are the n influence quantities involved in the corrected functional relationship. Recall that in the scalar case we started from a Taylor series expansion to Order 1, i.e.:

$$z_i + \delta z_i = z_i + \sum_{j=1}^{n} \frac{\partial f}{\partial u_j} \delta u_j \tag{4.3.8}$$

This can be written as:

$$\delta z_i \equiv \left(z_i - E[z_i]\right) \equiv \sum_{j=1}^{n} \frac{\partial f}{\partial u_j} \left(u_j - E[u_j]\right) \tag{4.3.9}$$

In matrix notation we can write:

$$(\mathbf{Z} - E[\mathbf{Z}]) = \mathbf{J}_U (\mathbf{U} - E[\mathbf{U}]) \tag{4.3.10}$$

Where the Jacobian $\mathbf{J}_u$ is defined by the following m x n matrix of partial derivatives:

$$\mathbf{J}_U \equiv \nabla_U \mathbf{U}^T = \begin{bmatrix} \frac{\partial f_1}{\partial u_1} & \frac{\partial f_1}{\partial u_2} & \cdot & \cdot & \frac{\partial f_1}{\partial u_n} \\ \frac{\partial f_2}{\partial u_1} & \frac{\partial f_2}{\partial u_2} & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \frac{\partial f_m}{\partial u_1} & \cdot & \cdot & \cdot & \frac{\partial f_m}{\partial u_n} \end{bmatrix} \tag{4.3.11}$$

where we have assumed that $\begin{Bmatrix} z_1 = f_1(u_1, u_2, \ldots u_n) \\ z_2 = f_2(u_1, u_2, \ldots u_n) \\ \cdot \\ \cdot \\ z_m = f_m(u_1, u_2, \ldots u_n) \end{Bmatrix}$ is true. Of course it may well be

that some, or all of the functions $f_1 \ldots f_m$ are identical. If we now square both sides of (4.3.10) and take expectations we find:

$$E\left[(\mathbf{Z} - E[\mathbf{Z}])(\mathbf{Z} - E[\mathbf{Z}])^T\right] = E\left[\mathbf{J}_U (\mathbf{U} - E[\mathbf{U}])\left(\mathbf{J}_U (\mathbf{U} - E[\mathbf{U}])\right)^T\right] \tag{4.3.12}$$

From (4.3.2) & (4.3.6) we see that this is in fact the covariance matrix of $\mathbf{Z}$:

$$\mathbf{cov}(\mathbf{Z}) = \mathbf{J}_U \mathbf{cov}(\mathbf{U}) \mathbf{J}_U^T$$

which we can write as:

$$\boxed{\Psi_Z = \mathbf{J}_U \Psi_U \mathbf{J}_U^T} \tag{4.3.13}$$

Equation (4.3.13) is the Gaussian Law of Error Propagation in its most general form (compare with the scalar form in (1.4.8)). Its simplicity and clarity highlight the superior convenience of the multivariate approach to the problem. Ψ_u will list all the variances and covariances of the influence quantities; $\mathbf{J}_u$ gives all the sensitivity coefficients (partial derivatives) and then Ψ_z gives the complete variance/covariance matrix for the output quantities.

4.4 An Example

Before proceeding with the application of (4.3.13) to the weighing equation–Eq. (4.2.4)–we pause to quote a useful example, illustrating the method. This example is taken from Annex H.2 of the Guide (ISO, 1993), however here we have changed it slightly to present the information in matrix notation. (A similar example is given in Weise (1985)). We reproduce this example here because of its excellent illumination of the concepts of the preceding section.

Consider an experiment where measurements are made of voltage, V, current I and phase angle ϕ in some circuit. Since the measurements are made contemporaneously on the one circuit, covariances and correlations can be expected between these three input quantities. The measurands of interest are:

$$\begin{aligned} &\text{Resistance, } R = \frac{V}{I}\cos\phi\,, \\ &\text{Reactance, } X = \frac{V}{I}\sin\phi\,, \\ &\text{Impedance, } |Z| = \left|\frac{V}{I}\right| = \left(R^2 + X^2\right)^{1/2} \end{aligned} \tag{4.4.1}$$

Thus for $\mathbf{Y} = \mathrm{f}(\mathbf{U})$ we have:

$$\mathbf{Y} = \begin{bmatrix} R \\ X \\ Z \end{bmatrix}; \quad \mathbf{U} = \begin{bmatrix} V \\ I \\ \phi \end{bmatrix} \;\&\; \mathbf{f}(\mathbf{U}) = \begin{bmatrix} f_1(V,I,\phi) \\ f_2(V,I,\phi) \\ f_3(V,I,\phi) \end{bmatrix} \tag{4.4.2}$$

We wish to evaluate $\psi_{\mathbf{Y}}$. First we calculate: (4.4.3)

$$\mathbf{J}_{\mathrm{U}} = \begin{bmatrix} \frac{\partial f_1}{\partial V} & \frac{\partial f_1}{\partial I} & \frac{\partial f_1}{\partial \phi} \\ \frac{\partial f_2}{\partial V} & \frac{\partial f_2}{\partial I} & \frac{\partial f_2}{\partial \phi} \\ \frac{\partial f_3}{\partial V} & \frac{\partial f_3}{\partial I} & \frac{\partial f_3}{\partial \phi} \end{bmatrix} = \begin{bmatrix} \frac{\cos\phi}{I} & \frac{-V\cos\phi}{I^2} & \frac{-V}{I}\sin\phi \\ \frac{\sin\phi}{I} & \frac{-V\sin\phi}{I^2} & \frac{V}{I}\cos\phi \\ \frac{1}{I} & \frac{-V}{I^2} & 0 \end{bmatrix}$$

We now need to evaluate ψ_{u}. To do this we realise that:

$$\psi_{\mathrm{U}} = \begin{bmatrix} s^2(V) & s(V,I) & s(V,\phi) \\ s(I,V) & s^2(I) & s(I,\phi) \\ s(\phi,V) & s(\phi,I) & s^2(\phi) \end{bmatrix} \tag{4.4.4}$$

The components of this matrix are evaluated from the actual data obtained in the experiment, using the expressions for variances and covariance given in Chapter 1. Eq. (4.3.13) can then be evaluated to give:

$$\psi_{\mathrm{Y}} = \begin{bmatrix} s^2(R) & s(R,X) & s(R,Z) \\ s(X,R) & s^2(X) & s(X,Z) \\ s(Z,R) & s(Z,X) & s^2(Z) \end{bmatrix} \tag{4.4.5}$$

Each term of this matrix evaluates to give quite complex algebraic expressions, if calculated theoretically. For example,

$$s^2(R)=\left(\frac{\cos\phi}{I}\right)^2 s^2(V)+\left(\frac{V}{I^2}\cos\phi\right)^2 s^2(I)+\left(\frac{V\sin\phi}{I}\right)^2 s^2(\phi)$$
$$-2\frac{V\cos^2\phi}{I^3}s(V,I)-2\frac{V\sin\phi\cos\phi}{I^2}s(V,I)+2\frac{V^2\sin\phi\cos\phi}{I^3}s(I,\phi) \qquad (4.4.6)$$

is the complete expression for the variance of the estimate of R, incorporating the variance components due to V, I, ϕ and their respective covariance terms.

4.5 Uncertainty Propagation in the Weighing Equation

At this point we wish to proceed and apply the Error Propagation Theory to Eq. (4.2.4), the matrix form of the weighing equation.(Bich *et al* (1993/94)) From Eq. (4.2.4) we can see that

$$\mathbf{Y}=f(\Delta\mathbf{w},\rho,\mathbf{X},\mathbf{T},\mathbf{V},\mathbf{D},\mathbf{M_n},\alpha) \qquad (4.5.1)$$

However, $\mathbf{X}$, $\mathbf{M_n}$ & α are constants and so contribute nothing to the covariance of $\mathbf{Y}$ [of course α is not really a constant but the influence of its variance would be so small that we can neglect it] Thus:

$$\mathbf{U}=\left[\Delta\mathbf{w}\quad \rho_a\quad \mathbf{V}\quad \mathbf{T}\quad \mathbf{D}\right]^{\mathbf{T}} \qquad (4.5.2\,a)$$

is the vector of influence quantities of interest and:

$$\mathbf{J_U}=\left[\mathbf{J}_{\Delta w}\ \vdots\ \mathbf{J}_{\rho_a}\ \vdots\ \mathbf{J_V}\ \vdots\ \mathbf{J_T}\ \vdots\ \mathbf{J_D}\right] \qquad (4.5.2b)$$

is the matrix of sensitivity coefficients to be evaluated. With (4.2.4) we can now establish the sub-matrices of (4.5.2):

$$\mathbf{J}_{\Delta w}=\begin{bmatrix} \frac{\partial y_1}{\partial \Delta w_1} & \frac{\partial y_1}{\partial \Delta w_2} & \cdots & \frac{\partial y_1}{\partial \Delta w_n} \\ \frac{\partial y_2}{\partial \Delta w_1} & \frac{\partial y_2}{\partial \Delta w_2} & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \frac{\partial y_n}{\partial \Delta w_n} \end{bmatrix} = \mathbf{I_n} \qquad (4.5.3)$$

Since $\frac{\partial y_i}{\partial \Delta w_j}=1$ when $i=j$ and $=0$ when $i\neq j$, there are thus no off-diagonal terms in (4.5.3) and hence $\mathbf{J}_{\Delta w}=\mathbf{I_n}$. Similarly, for the next term in (4.5.2b) we get:

$$\mathbf{J}_{\rho_a} = \begin{bmatrix} \frac{\partial y_1}{\partial \rho_{a_1}} & & & \\ & \frac{\partial y_2}{\partial \rho_{a_2}} & \mathbf{0} & \\ & & \ddots & \\ & \mathbf{0} & & \frac{\partial y_n}{\partial \rho_{a_n}} \end{bmatrix} \tag{4.5.4}$$

Which can be evaluated in the light of (4.2.4), to yield:

$$\mathbf{J}_{\rho} = diag\left[\{(1-20\alpha)\mathbf{I}_n + \alpha\mathbf{T}\}\mathbf{XV}\right] \tag{4.5.6}$$

The next term in (4.5.2b) is:

$$\mathbf{J}_{\mathbf{V}} = \begin{bmatrix} \frac{\partial y_1}{\partial V_1} & \frac{\partial y_1}{\partial V_2} & \cdots & \frac{\partial y_1}{\partial V_p} \\ \frac{\partial y_2}{\partial V_1} & \frac{\partial y_2}{\partial V_2} & \cdots & \vdots \\ \vdots & \vdots & \cdots & \vdots \\ \vdots & \vdots & \cdots & \vdots \\ \vdots & \vdots & \cdots & \frac{\partial y_n}{\partial V_p} \end{bmatrix} \tag{4.5.7}$$

This term, unlike the preceding two, will not be diagonal: the exact form will depend upon the standards used in each comparison, but we can expect that some off-diagonal terms will exist on each row. Which standards are used in each comparison depends upon the form of $\mathbf{X}$. So we would expect $\mathbf{J}_{\mathbf{V}} \propto \mathbf{X}$. Indeed, the general form of (4.5.7), from (4.2.4) is:

$$\mathbf{J}_{\mathbf{V}} = \{(1-20\alpha)\mathbf{I}_n + \alpha\mathbf{T}\}\rho\mathbf{X} \tag{4.5.8}$$

Since the temperature measurements are independent we can expect no correlation between t_i & t_j so that:

$$\mathbf{J}_{\mathbf{T}} = \begin{bmatrix} \frac{\partial y_1}{\partial t_1} & & & \\ & \frac{\partial y_2}{\partial t_2} & \mathbf{0} & \\ & & \ddots & \\ & \mathbf{0} & & \frac{\partial y_n}{\partial t_n} \end{bmatrix} \tag{4.5.9}$$

and from (4.2.4) we find the specific form of $\mathbf{J}_{\mathbf{T}}$ to be:

$$\mathbf{J}_{\mathbf{T}} = \alpha\, diag\{\rho\mathbf{XV}\} \tag{4.5.10}$$

and similarly for $\mathbf{J}_{\mathbf{D}}$ we simply have:

$$\mathbf{J}_{\mathbf{D}} = \left(\frac{2}{r_e}\right) diag\{\mathbf{M}_n\} \tag{4.5.11}$$

Now we can expand (4.5.2b) to give:

$$\mathbf{J_U} = \left[\mathbf{I}_n \;\vdots\; diag\left[\{(1-20\alpha)\,\mathbf{I}_n + \alpha\mathbf{T}\}\mathbf{XV}\right] \;\vdots\; \{(1-20\alpha)\mathbf{I}_n + \alpha\,\mathbf{T}\}\rho\mathbf{X} \;\vdots\; \alpha\, diag\{\rho\mathbf{XV}\} \;\vdots\; \left(\frac{2}{r_e}\right) diag(\mathbf{M_n}) \right] \tag{4.5.12}$$

The dimensions of this augmented matrix are *(n × (4n+p))*.

Now from (4.5.2a) we can see that **cov**(**U**) is:

$$\psi_{\mathbf{U}} = diag\left\{\psi_{\Delta w}, \psi_\rho, \psi_{\mathbf{V}}, \psi_{\mathbf{T}}, \psi_{\mathbf{D}}\right\} \tag{4.5.13}$$

where, similarly to the univariate case, we have assumed there are no covariances between the input influence quantities in **U**. We further assume that each of the sub-matrices in (4.5.13) is itself diagonal, as each represents the covariance matrix of a vector of independent quantities. Note: it may be recalled from Section 3.2 that we stated that a covariance did exist between ρ_{a_i} & ρ_{a_j} due to the common terms of *R*, the gas constant, M_v the water vapour molar mass and M_a, the dry air molar mass, which terms appear in the functional form of the air density equation (3.1.5). Thus ψ_ρ in (4.5.13) above should be completely non-diagonal; however, as was pointed out in Eq. (3.2.7), the contribution to $\psi_{\mathbf{Y}}$ due to this covariance is so small compared to that which arises from the volume terms in the weighing equation, that its neglect is entirely justified. We now evaluate (4.3.13) with (4.5.12) & (4.5.13) as follows:

$$\psi_{\mathbf{Y}} = \mathbf{J_U}\,\psi_{\mathbf{U}}\,\mathbf{J_U^T}$$

$$= \left[\mathbf{I}_n \;\vdots\; diag\left\{\left[(1-20\alpha)\mathbf{I_n} + \alpha\mathbf{T}\right]\mathbf{XV}\right\} \;\vdots\; \left[(1-20\alpha)\mathbf{I_n} + \alpha\mathbf{T}\right]\rho\mathbf{X} \;\vdots\; \alpha.diag\{\rho\mathbf{XV}\} \;\vdots\; \left(\frac{2}{r_e}\right) diag\{\mathbf{M_N}\} \right]$$

$$\bullet \begin{bmatrix} \psi_{\Delta w} & & & & \\ & \psi_\rho & & & \\ & & \psi_{\mathbf{V}} & & \\ & & & \psi_{\mathbf{T}} & \\ & & & & \psi_{\mathbf{D}} \end{bmatrix} \bullet \begin{bmatrix} \mathbf{I_n} \\ diag\left\{\left[(1-20\alpha)\mathbf{I_n} + \alpha\mathbf{T}\right]\mathbf{XV}\right\} \\ \mathbf{X^T}\rho^{\mathbf{T}}\left[(1-20\alpha)\mathbf{I_n} + \alpha\mathbf{T}\right] \\ \alpha\, diag\{\rho\mathbf{XV}\} \\ \frac{2}{r_e} diag(\mathbf{M_n}) \end{bmatrix} \tag{4.5.14}$$

where we have noted that symmetric matrices remain unchanged on transposition. Evaluating and simplifying (4.5.14) yields:

$$\begin{aligned} \psi_{\mathbf{Y}} = \;& \psi_{\Delta w} \\ & + diag\{[...].\mathbf{XV}\}\,\psi_\rho\, diag\{[...]\mathbf{XV}\} \\ & + [...]\rho\mathbf{X}\psi_{\mathbf{V}}\mathbf{X^T}\rho^{\mathbf{T}}[...] \\ & + (\alpha)^2\, diag\{\rho\mathbf{XV}\}\psi_{\mathbf{T}} diag\{\rho\mathbf{XV}\} \\ & + \left(\frac{2}{r_e}\right)^2 diag\{\mathbf{M_n}\}\psi_{\mathbf{D}} diag\{\mathbf{M_n}\} \end{aligned} \tag{4.5.15}$$

where [...] represents $\left[(1-20\alpha)\mathbf{I}_n + \alpha\mathbf{T}\right]$ which is a symmetric matrix.

Equation (4.5.15) gives the complete covariance matrix for $\mathbf{Y}$ in terms of the covariance matrices of the influence quantities. While we have assumed all of the latter (ψ_ρ, $\psi_{\Delta w}$ etc.) to be diagonal, $\psi_{\mathbf{Y}}$ is nonetheless *not* diagonal as a result of the third term on the r.h.s. of (4.5.15). This is the term due to the volume influence.

We can simplify things quite a bit if we assume the two small systematic corrections, due to volume expansion coefficients and center of gravity differences, can be neglected. Then we would obtain, instead of (4.2.4) and (4.5.15):

$$\boxed{\mathbf{Y} = \Delta\mathbf{w} + \rho\mathbf{X}\mathbf{V}} \tag{4.5.16}$$

$$\boxed{\psi_{\mathbf{Y}} = \psi_{\Delta w} + diag\{\mathbf{X}\mathbf{V}\}\psi_\rho diag\{\mathbf{X}\mathbf{V}\} + \rho\mathbf{X}\psi_{\mathbf{V}}\mathbf{X}^{\mathrm{T}}\rho^{\mathrm{T}}} \tag{4.5.17}$$

So equations (4.2.4) & (4.5.15) or (4.5.16) & (4.5.17) provide complete analytical tools for evaluating all the necessary information about our measurand estimates, the Δm_i, or mass differences, for a set of comparisons among various combinations of mass standards. Fig. (4.0.1) at the beginning of this chapter now provides us with a useful conclusion: the three stages of the process are highlighted—i.e. first identifying the input quantities and forming the weighing equation; secondly evaluating the sensitivity coefficient matrices and the input covariance matrices; and finally evaluating the contribution of each of these to the overall covariance matrix.

We must now proceed to the real task at hand, which is to use the information presented by the methods described here to estimate the *values* of the parameters themselves, i.e. the $m_{j,k}$ in $\Delta m_i = \sum m_j - \sum m_k$.

In order to do this we must now turn our attention to the study of Parameter Estimation techniques.

****

5. Parameter Estimation Techniques in Mass Calibration

5.0 Summary

At this point we have established relationships for calculating the measurands (mass differences) and their combined standard uncertainties both in univariate and multivariate form. This has been done in a generalised and uniform manner, treating all influence quantities equally and including all known information in the analysis.

Before proceeding we need to identify some terms: as shown in Fig. (5.0.1) below, we have first the influence quantities leading to the measurand via the functional relationship. In the case of mass calibration the measurands are mass differences resulting from comparison calibrations. However, ultimately we require mass values (the parameters) so further analysis is needed. The series of comparisons carried out is described by a system model which relates the experimental measurands to the parameters via a design matrix.

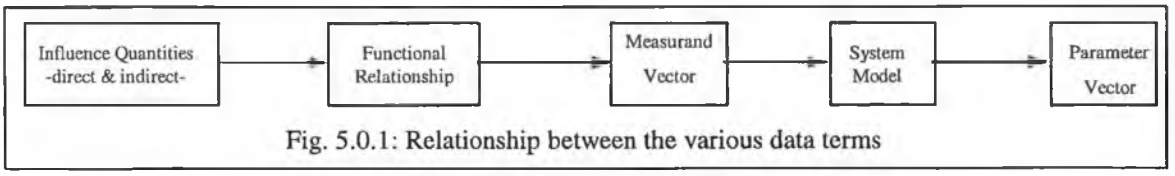


Fig. 5.0.1: Relationship between the various data terms

With this much established, it is our purpose in the next four chapters to investigate the Estimation Methods appropriate for determining the parameter vector and its covariance matrix. Since the design matrix will be chosen so as to have an over-determined system we can implement statistical estimation techniques.

We first investigate the Least Squares (LS) method and find it inappropriate since the observation vector of measurands only contains differences among the parameters and thus absolute values cannot be determined without some extra information. This leads us to LS subject to constraints—Restrained Least Squares (RLS), in which the constraints used are the previously determined values of one or more of the standards involved in the calibration exercise. Fig. (5.0.2) below illustrates the method.

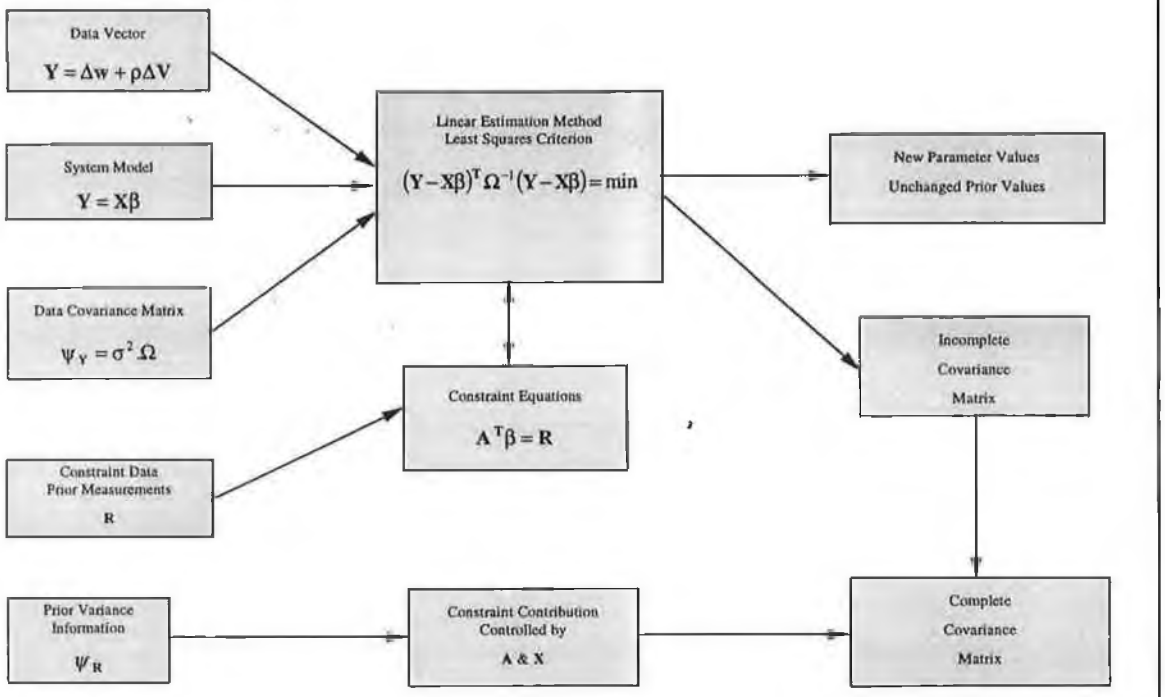


Fig. 5.0.2: The RLS method of Parameter Estimation in Mass Determination

An extended treatment is given of RLS since a discussion of constraints is critical for Estimation Techniques in Mass Determination. It is shown how RLS treats the constraints (prior information) deterministically leading to a solution vector with an incomplete covariance matrix. A final solution is then found by combining the correct contribution of the prior variance/covariance information as determined by the estimation method. The method is discussed at some length, two significant flaws being highlighted: Firstly the constraint information is treated inconsistently, being viewed as fixed, or deterministic to obtain a parameter estimate, but then viewed stochastically to get a complete covariance matrix. This is shown to be mathematically unsatisfactory. Secondly the approach is anything but uniform in its treatment of the various data sets. Rather, it is shown to be in agreement with the policy of separating random & systematic uncertainty components as in conventional analysis and indeed can perhaps be justified in that light. However, the goal in this work is to produce a unified analysis at all stages so it is not acceptable to treat the constraint information in this manner. Compare Fig. (5.0.2) with Fig. (1.0.1) to see the underlying differences between the approaches.

We then present an alternative method, the Augmented Design approach (AD), Bich (1992), in which the fundamental difference is that the constraints are considered as just extra data needed to get a solution. Thus its covariance information can be

included in the estimation process, leading to complete solutions. Fig. (5.0.3) illustrates the method, showing how the augmented data is produced.

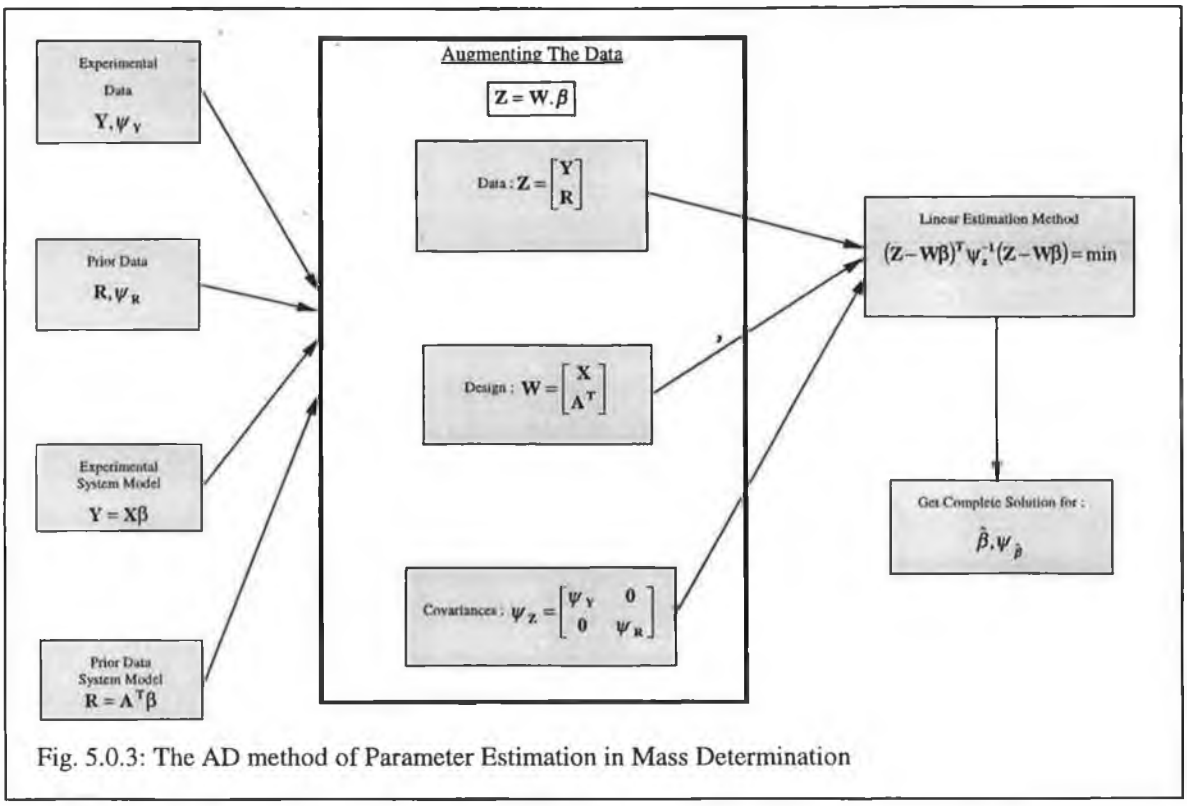


Fig. 5.0.3: The AD method of Parameter Estimation in Mass Determination

The significant feature of this method is that it leads to a *smaller* covariance matrix than does RLS, and indeed *reduces* the uncertainties of the Reference Standards (the prior information). This seems surprising at first but can be interpreted by remembering that the constraints are just being viewed as extra stochastic data. Three examples are included in this chapter, treated by both RLS & AD to highlight the attributes of each, and also their key differences.

5.1 Introduction

We have already established the form of the Functional Relationship used in comparison calibrations involving mass standards and have calculated the corresponding covariance matrix. We have noted that a final aim of the procedure is the evaluation of *mass values*, while the experimental method allows us to establish *mass differences*. Some further data reduction is now needed. From Eq. (4.2.4) we know that:

$$\mathbf{Y} = f(\Delta\mathbf{W}, \mathbf{V}, \rho, \mathbf{T}, \mathbf{D}, \mathbf{M_n}, \mathbf{X}) \tag{5.1.1}$$

which gives us the vector of mass differences:

$$[\mathbf{Y}] = \begin{bmatrix} y_1 \\ y_2 \\ . \\ y_n \end{bmatrix} = \begin{bmatrix} \Delta m_1 \\ \Delta m_2 \\ . \\ \Delta m_n \end{bmatrix}$$

We also have:

$$\mathbf{Y} = \mathbf{X}.\beta \tag{5.1.2}$$

as our system model describing the measurements. The n x p matrix $\mathbf{X}$ is the design matrix, indicating the form of the comparisons, while β is a p x 1 vector of the required parameters (i.e. the mass values of the standards). To estimate the parameters, the minimum requirement is that $\mathbf{X}$ should contain at least p independent rows—indeed there can only be p independent rows in a system with p parameters, but there could be less, in which case the system would not be solvable.

Because our observations, $\mathbf{Y}$, are stochastic in nature, it is beneficial to maintain the redundant information present in an over-determined design scheme where linearly dependent rows are present in $\mathbf{X}$. This allows the extra, statistical, information so presented to be used via some parameter estimation technique in order to establish so-called "best fit" values for the parameters. Our purpose now is to examine several such estimation techniques in order to find one which is most appropriate to the situation at hand.

5.2 Least Squares Methods

Probably the most widely known and used procedure is the well-known Ordinary Least Squares (OLS) Solution which involves a minimisation of the sum:

$$S_{LS} = (\mathbf{Y} - \mathbf{X}\beta)^{\mathrm{T}}(\mathbf{Y} - \mathbf{X}\beta) \tag{5.2.1}$$

for a model given by (5.1.2). The estimator satisfying this criteria is the OLS estimator (see for e.g. Beck & Arnold (1977), Eadie (1971), Luenberger (1968), Zelen (1962) & Mandel (1964)):

$$\boxed{\hat{\beta}_{\mathrm{ols}} = \left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{Y}} \tag{5.2.2}$$

The only immediate requirement here is that $n \geq p$ so that the product $\mathbf{X}^{\mathrm{T}}\mathbf{X}$ is non-singular. For the estimation to produce any information that could not be obtained by simple algebra, we further need to have $n \geq (p+1)$. If $\mathbf{X}$ & β are non-stochastic and the dispersion characteristics of $\mathbf{Y}$ imply additive, zero-mean errors, it then follows that

$$E\left[\hat{\beta}_{\mathrm{ols}}\right] = \beta \tag{5.2.3}$$

or, the Ordinary Least Squares Estimator is unbiased with respect to the parameters. Noting that (5.2.2) can be expressed as:

$$\hat{\beta}_{\mathrm{ols}} = \mathbf{A}\,\mathbf{Y} \quad \text{with} \quad \mathbf{A} = \left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathrm{T}}$$

we have, from (Eq 4.3.6) that:

$$\mathbf{cov}\left[\hat{\beta}_{ols}\right] = \psi_{\hat{\beta}_{ols}} = \mathbf{A}\,\psi_{\mathbf{Y}}\,\mathbf{A}^{\mathbf{T}} = \left(\mathbf{X}^{\mathbf{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathbf{T}}\,\psi_{\mathbf{Y}}\,\mathbf{X}\left(\mathbf{X}^{\mathbf{T}}\mathbf{X}\right)^{-1} \tag{5.2.4}$$

in which $\mathbf{X}^{\mathbf{T}}\mathbf{X}$ is a symmetric matrix. However, (5.2.4) will not be a minimum variance estimator without the added restriction that

$$\psi_{\mathbf{Y}} = \sigma^2\mathbf{I} \tag{5.2.5}$$

resulting in

$$\psi_{\hat{\beta}_{ols}} = \sigma^2\left(\mathbf{X}^{\mathbf{T}}\mathbf{X}\right)^{-1} \tag{5.2.6}$$

Thus providing (5.2.3) & (5.2.5) are satisfied, (5.2.2) & (5.2.6) give the Best Linear Unbiased Estimator (BLUE !) for (5.1.2).

We need to check of course, that this is really valid before proceeding. It is most unlikely that (5.2.5) will hold since the diagonal terms of $\psi_{\mathbf{Y}}$ will not usually be identical in practice, although they could be approximated as identical by choosing a "worst case" largest variance, but this is not very satisfactory since we do place some importance on achieving *optimal* estimation which will be a realistic reflection of what we have observed. In any event, this problem is eclipsed by the much greater difficulty of $\psi_{\mathbf{Y}}$ being non-diagonal; as we noted in Sec. 4.5, the form of $\psi_{\mathbf{Y}}$ is such that there are usually significant off-diagonal terms present.

There are two methods of dealing with this problem: one is to invoke Weighted Least Squares (WLS), and the other attacks the problem via the Gauss Markov Theorem (see Luenberger (1968), Zelen (1962), Rao (1973)). In both cases it is assumed that:

$$\psi_{\mathbf{Y}} = \sigma^2\Omega \tag{5.2.7}$$

where the form of Ω is known but σ^2 may not be. In other words, $\psi_{\mathbf{Y}}$ is known to within a multiplicative constant (σ). In this case the estimators become:

$$\boxed{\hat{\beta}_{gm} = \hat{\beta}_{wls} = \left(\mathbf{X}^{\mathbf{T}}\,\Omega^{-1}\,\mathbf{X}\right)^{-1}\mathbf{X}^{\mathbf{T}}\,\Omega^{-1}\,\mathbf{Y}} \tag{5.2.8}$$

$$\boxed{\psi_{\hat{\beta}_{gm}} = \psi_{\hat{\beta}_{wls}} = \sigma^2\left(\mathbf{X}^{\mathbf{T}}\,\Omega^{-1}\,\mathbf{X}\right)^{-1}} \tag{5.2.9}$$

However, while this method circumvents the problem with condition (5.2.5), and allows BLUE's to be obtained, another obstacle appears in that $\left|\mathbf{X}^{\mathbf{T}}\mathbf{X}\right| = 0$. This is unavoidable when only differences in parameters are measured. The inevitable lack of information resulting in such cases leads to the singularity in $\mathbf{X}^{\mathbf{T}}\mathbf{X}$ above, and the parameters being non-estimable—that is no unique solution to (5.2.8) can be found.

5.3 Restrained Least Squares

In order then to obtain a solution, it is necessary to include some external information (see Cameron *et al*, (1977) Bich (1992), Nielson (1997) and for another approach Hughes & Musk, (1972)). That is, the estimators $\hat{\beta}$ should satisfy some set of linearly independent restraints such that:

$$\mathbf{A}^{\mathrm{T}}\,\beta - \mathbf{R} = \mathbf{0} \tag{5.3.1}$$

where **A** is the "design matrix" of constraints and **R** is the vector of constraints. There may be m such constraints such that **R** is of order m x 1. Then for the p parameters one would have:

$$\begin{aligned} &a_{11}\beta_1 + a_{12}\beta_2 + \ldots\ldots\ldots + a_{1p}\beta_p = r_1 \\ &\vdots \\ &a_{m1}\beta_1 + a_{m2}\beta_2 + \ldots\ldots\ldots + a_{mp}\beta_p = r_m \end{aligned} \tag{5.3.2}$$

so that we have:

$$\begin{bmatrix} a_{11} & a_{12} & \cdot & \cdot & a_{1p} \\ a_{21} & & & & \\ \cdot & & \cdot & & \\ \cdot & & & \cdot & \\ a_{m1} & \cdot & \cdot & \cdot & a_{mp} \end{bmatrix} \cdot \begin{bmatrix} \beta_1 \\ \beta_2 \\ \cdot \\ \cdot \\ \beta_p \end{bmatrix} = \begin{bmatrix} r_1 \\ \cdot \\ \cdot \\ \cdot \\ r_m \end{bmatrix} \tag{5.3.3}$$

In our practical case this implies that there are some linear combinations of the p parameters [mass standards] whose values are known *a priori* and are not linearly related to the combinations represented by **X**.β. In practice, this prior knowledge would often be the previously determined values of some or all of the parameters. Then the elements of $\mathbf{A}^{\mathrm{T}}$ would be either 1 or 0.

It is now necessary to solve the minimisation of the sum given in (5.2.1) subject to the constraint in (5.3.1). Note that because of (5.2.7), we write (5.2.1) now as:

$$S = f(\beta) = (\mathbf{Y} - \mathbf{X}\beta)^{\mathrm{T}}\,\Omega^{-1}\,(\mathbf{Y} - \mathbf{X}\beta) \tag{5.3.4}$$

and from (5.3.2) we have:

$$g(\beta) = \mathbf{A}^{\mathrm{T}}\,\beta - \mathbf{R} = \mathbf{0} \tag{5.3.5}$$

In the Lagrange Undetermined Multipliers method, we have the following simultaneous equations:

$$\frac{\partial \phi(\beta)}{\partial \beta} = \mathbf{0} \tag{5.3.6a}$$

$$g(\beta) = \mathbf{0} \tag{5.3.6b}$$

where:

$$\phi(\beta) = f(\beta) + 2\lambda g(\beta) \tag{5.3.6c}$$

λ being a (p x 1) vector of Lagrange multipliers. Then the solution will be found from:

$$\nabla_\beta \left\{ (\mathbf{Y}-\mathbf{X}\beta)^T \, \Omega^{-1} \, (\mathbf{Y}-\mathbf{X}\beta) + 2\lambda^T \left(\mathbf{A}^T\beta - \mathbf{R} \right) \right\} \Big|_{\substack{\beta=\hat{\beta} \\ \lambda=\hat{\lambda}}} = \mathbf{0} \qquad (5.3.7)$$

We first note some properties of ∇_β where β is a p x 1 vector: If $\mathbf{C}$ is a p x 1 vector which is not a function of β then:

$$\nabla_\beta \, \mathbf{C}^T \beta = \mathbf{C} \qquad (5.3.8a)$$

Also, if $\mathbf{B}$ is a p x m matrix which is not a function of β then:

$$\nabla_\beta \, \beta^T \mathbf{B} = \mathbf{B} \qquad (5.3.8b)$$

Furthermore, if $\mathbf{Q} = \mathbf{A}^T \Phi \, \mathbf{A}$ where $\mathbf{A}$ is an n x 1 vector and Φ is an n x n symmetric matrix, then, if $\mathbf{A}$ is a function of β and Φ is not, we have:

$$\nabla_\beta \, \mathbf{Q} = 2 \left(\nabla_\beta \, \mathbf{A}^T \right) \Phi \, \mathbf{A}$$

If $\mathbf{A} = \mathbf{X}\beta$ for $\mathbf{X}$ an n x p matrix, as often occurs in linear estimation, the above expression becomes:

$$\nabla_\beta \, \mathbf{Q} = 2 \left(\nabla_\beta \, \beta^T \mathbf{X}^T \right) \Phi \, \mathbf{X} \, \beta$$

$$\Rightarrow \nabla_\beta \, \mathbf{Q} = 2 \, \mathbf{X}^T \Phi \, \mathbf{X} \, \beta \qquad (5.3.8c)$$

where (5.3.8b) has been used.

With these 3 equations, we can now evaluate (5.3.7):

$$\nabla_\beta \left(\mathbf{Y}^T\Omega^{-1}\mathbf{Y} - \mathbf{Y}^T\Omega^{-1}\mathbf{X}\hat{\beta} - \hat{\beta}^T\mathbf{X}^T\Omega^{-1}\mathbf{Y} + \hat{\beta}^T\mathbf{X}^T\Omega^{-1}\mathbf{X}\beta + 2\hat{\lambda}^T\mathbf{A}^T\hat{\beta} - 2\hat{\lambda}^T\mathbf{R} \right) = \mathbf{0}$$

$$\Rightarrow \mathbf{0} - \mathbf{X}^T\Omega^{-1}\mathbf{Y} - \mathbf{X}^T\Omega^{-1}\mathbf{Y} + 2\mathbf{X}^T\Omega^{-1}\mathbf{X}\hat{\beta} + 2\mathbf{A}\hat{\lambda} - \mathbf{0} = \mathbf{0}$$

$$\text{or} \quad -2\mathbf{X}^T\Omega^{-1}\mathbf{Y} + 2\mathbf{X}^T\mathbf{X}\hat{\beta} + 2\mathbf{A}\hat{\lambda} = \mathbf{0} \qquad (5.3.9)$$

So we get the two equations:

$$\mathbf{X}^T\Omega^{-1}\mathbf{X}\hat{\beta} + \mathbf{A}\hat{\lambda} = \mathbf{X}^T\Omega^{-1}\mathbf{Y} \qquad (5.3.10a)$$

$$\& \quad \mathbf{A}^T\hat{\beta} = \mathbf{R} \qquad (5.3.10b)$$

This can also be expressed as:

$$\begin{bmatrix} \mathbf{X}^T\Omega^{-1}\mathbf{X} & \mathbf{A} \\ \mathbf{A}^T & \mathbf{0} \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \hat{\lambda} \end{bmatrix} = \begin{bmatrix} \mathbf{X}^T\Omega^{-1}\mathbf{Y} \\ \mathbf{R} \end{bmatrix} \qquad (5.3.11)$$

Thus if we define:

$$\begin{bmatrix} \mathbf{c}_1 & \mathbf{c}_2 \\ \mathbf{c}_2^T & \mathbf{c}_3 \end{bmatrix} = \begin{bmatrix} \mathbf{X}^T\Omega^{-1}\mathbf{X} & \mathbf{A} \\ \mathbf{A}^T & \mathbf{0} \end{bmatrix}^{-1} \qquad (5.3.12)$$

we can then say:

$$\begin{bmatrix} \hat{\beta} \\ \hat{\lambda} \end{bmatrix} = \begin{bmatrix} \mathbf{c}_1 & \mathbf{c}_2 \\ \mathbf{c}_2^T & \mathbf{c}_3 \end{bmatrix} \begin{bmatrix} \mathbf{X}^T\Omega^{-1}\mathbf{Y} \\ \mathbf{R} \end{bmatrix} \qquad (5.3.13)$$

(Aside: we will see this type of analysis again later when we examine pseudo-inverses and Generalised Least Squares (see Chapter 6).)

Now if we pre-multiply the l.h.s. of (5.3.12) by the inverse its r.h.s. we get:

$$\begin{bmatrix} \mathbf{X}^T\Omega^{-1}\mathbf{X} & \mathbf{A} \\ \mathbf{A}^T & \mathbf{0} \end{bmatrix}\begin{bmatrix} \mathbf{c}_1 & \mathbf{c}_2 \\ \mathbf{c}_2^T & \mathbf{c}_3 \end{bmatrix} = \mathbf{I} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix}$$

$$\Rightarrow \mathbf{X}^T\Omega^{-1}\mathbf{X}\mathbf{c}_1 + \mathbf{A}\mathbf{c}_2^T = \mathbf{I} \quad (5.3.14a)$$

$$\mathbf{X}^T\Omega^{-1}\mathbf{X}\mathbf{c}_2 + \mathbf{A}\mathbf{c}_3 = \mathbf{0} \quad (5.3.14b)$$

$$\mathbf{A}^T\mathbf{c}_1 = \mathbf{0} \quad (5.3.14c)$$

$$\mathbf{A}^T\mathbf{c}_2 = \mathbf{I} \quad (5.3.14d)$$

Now because $\left(\mathbf{X}^T\Omega^{-1}\mathbf{X}\right)$ is singular–the reason for this approach initially–we need to make the following definition in order to proceed:

$$\mathbf{X}^T\Omega^{-1}\mathbf{X} = \mathbf{a}_0 + \mathbf{D}\mathbf{A}\mathbf{A}^T \quad (5.3.15)$$

where $\mathbf{a}_0$ is a non-singular diagonal matrix. We can choose $\mathbf{D} = -\mathbf{I}$ so that:

$$\boxed{\mathbf{X}^T\Omega^{-1}\mathbf{X} = \mathbf{a}_0 - \mathbf{A}\mathbf{A}^T} \quad (5.3.16)$$

Substituting this into (5.3.14a) and noting (5.3.14c) results in:

$$\mathbf{c}_1 = \mathbf{a}_0^{-1}\left(\mathbf{I} - \mathbf{A}.\mathbf{c}_2^T\right) \quad (5.3.17)$$

Similarly, substituting (5.3.16) into (5.3.14b) and noting (5.3.14d) yields:

$$\mathbf{c}_2 = \mathbf{a}_0^{-1}\mathbf{A}(\mathbf{I} - \mathbf{c}_3) \quad (5.3.18)$$

Pre-multiplying (5.3.18) by $\mathbf{A}^T$ gives:

$$\mathbf{A}^T\mathbf{c}_2 = \left(\mathbf{A}^T\mathbf{a}_0^{-1}\mathbf{A}\right)(\mathbf{I} - \mathbf{c}_3) = \mathbf{I} \quad \text{by (5.3.14d)}$$

$$\mathbf{c}_3 = \mathbf{I} - \left(\mathbf{A}^T\mathbf{a}_0^{-1}\mathbf{A}\right)^{-1} \quad (5.3.19)$$

We can now use Eqs. (5.3.17) - (5.3.19) in (5.3.13) to obtain:

$$\begin{bmatrix} \hat{\beta} \\ \hat{\lambda} \end{bmatrix} = \begin{bmatrix} \mathbf{a}_0^{-1}\left(\mathbf{I} - \mathbf{A}\mathbf{c}_2^T\right) & \mathbf{a}_0^{-1}\mathbf{A}(\mathbf{I} - \mathbf{c}_3) \\ (\mathbf{I} - \mathbf{c}_3)^T\mathbf{A}^T\mathbf{a}_0^{-1} & \mathbf{I} - \left(\mathbf{A}^T\mathbf{a}_0^{-1}\mathbf{A}\right)^{-1} \end{bmatrix}\begin{bmatrix} \mathbf{X}^T\Omega^{-1}\mathbf{Y} \\ \mathbf{R} \end{bmatrix} \quad (5.3.20)$$

Note that:

$$\mathbf{c}_1 = \mathbf{a}_0^{-1}\left(\mathbf{I} - \mathbf{A}\left(\mathbf{A}^T\mathbf{a}_0^{-1}\mathbf{A}\right)^{-1}\mathbf{A}^T\mathbf{a}_0^{-1}\right) \quad (5.3.21)$$

$$\& \quad \mathbf{c}_2 = \mathbf{a}_0^{-1}\mathbf{A}\left(\mathbf{A}^T\mathbf{a}_0^{-1}\mathbf{A}\right)^{-1} \quad (5.3.22)$$

Evaluating (5.3.20) results in:

$$\boxed{\hat{\beta} = \mathbf{a}_0^{-1}\left[\mathbf{X}^T\Omega^{-1}\mathbf{Y} + \mathbf{A}\left(\mathbf{A}^T\mathbf{a}_0^{-1}\mathbf{A}\right)^{-1}\left\{\mathbf{R} - \mathbf{A}^T\mathbf{a}_0^{-1}\mathbf{X}^T\Omega^{-1}\mathbf{Y}\right\}\right]} \quad (5.3.23)$$

$$\& \quad \boxed{\hat{\lambda} = \mathbf{R} + \left(\mathbf{A}^T\mathbf{a}_0^{-1}\mathbf{A}\right)^{-1}\left\{\mathbf{A}^T\mathbf{a}_0^{-1}\mathbf{X}^T\Omega^{-1}\mathbf{Y} - \mathbf{R}\right\}} \quad (5.3.24)$$

with the aid of (5.3.24), it is also possible to write (5.3.23) as:

$$\boxed{\hat{\beta} = \mathbf{a}_0^{-1}\left[\mathbf{X}^T\Omega^{-1}\mathbf{Y} + \mathbf{A}\left(\mathbf{R} - \hat{\lambda}\right)\right]} \quad (5.3.25)$$

So Eqs (5.3.16), (5.3.24) & (5.3.25) allow a Restrained Least Squares estimate of the parameter vector β in $\mathbf{Y} = \mathbf{X}\beta$, subject to the constraints in $\mathbf{A}^T\beta = \mathbf{R}$. It will be noted from (5.3.13) that, in fact, $\hat{\beta}$ takes the form:

$$\hat{\beta} = \mathbf{c}_1\mathbf{X}^T\Omega^{-1}\mathbf{Y} + \mathbf{c}_2\mathbf{R} \tag{5.3.26}$$

This is significant, because, if compared with (5.2.8), where:

$$\hat{\beta}_{gm} = \hat{\beta}_{wls} = \left(\mathbf{X}^T\Omega^{-1}\mathbf{X}\right)^{-1}\mathbf{X}^T\Omega^{-1}\mathbf{Y}$$

it can be seen that in (5.2.8), the parameter estimates are a linear combination of the observations, i.e.:

$$\hat{\beta} = \mathbf{L}\mathbf{Y} \tag{5.3.27}$$

$$\text{where } \mathbf{L} = \left(\mathbf{X}^T\Omega^{-1}\mathbf{X}\right)^{-1}\mathbf{X}^T\Omega^{-1}$$

But in (5.3.26) we have:

$$\hat{\beta} = \mathbf{M}\mathbf{Y} + \mathbf{c}_2\mathbf{R} \tag{5.3.28}$$

with $\mathbf{M} = \mathbf{c}_1\mathbf{X}^T\Omega^{-1}$.

Comparing (5.3.27) & (5.3.28), the chief difference is that a linear combination of the constraints is also added to the linear combination of the observations for purposes of establishing the parameter estimates. Hence the constraints need to be chosen to be something physically meaningful with respect to the parameters under scrutiny.

Now we need to look at the covariance matrix for $\hat{\beta}$ as given by (5.3.26) or (5.3.28). We can say that:

$$\mathbf{cov}\left[\hat{\beta}\right] \equiv \psi_{\hat{\beta}} = \mathbf{M}\psi_Y\mathbf{M}^T \tag{5.3.29}$$

since the constraints are considered to be deterministic, or constant; otherwise the Lagrange method would not be appropriate. So from (5.3.26), (5.3.29) becomes:

$$\psi_{\hat{\beta}} = \mathbf{c}_1\mathbf{X}^T\Omega^{-1}\psi_Y\Omega^{-1}\mathbf{X}\mathbf{c}_1^T \tag{5.3.30}$$

Note that $\psi_Y = \sigma^2\Omega$ and that in many practical cases we can assume $\sigma^2 = 1$—i.e. that Ω is fully known. In these circumstances we find:

$$\psi_{\hat{\beta}} = \mathbf{c}_1\mathbf{X}^T\Omega^{-1}\mathbf{X}\mathbf{c}_1^T \tag{5.3.31}$$

Aside : We will see later that in fact $\mathbf{c}_1$ is a *g-inverse* of $\mathbf{X}^T\Omega^{-1}\mathbf{X}$ and so, since $\mathbf{c}_1$ is a symmetric matrix it follows that $\psi_{\hat{\beta}} = \mathbf{c}_1$, but more on this later. (See Secs. 6.2 / 6.3)

5.4 Discussion (Refer again to Fig. 5.0.2)

We now analyse the RLS method in the light of the general criteria for the unified approach that we wish to establish. By introducing the constraint we have permitted a solution to be found and also a covariance matrix. That $\hat{\beta}$ is unbiased in spite of the extra $\mathbf{c}_2\mathbf{R}$ term in (5.3.28) can be verified by considering (from Eq. (5.3.25)):

$$E\left[\hat{\beta}\right] = \mathbf{a}_0^{-1}\mathbf{X}^T\Omega^{-1}E[\mathbf{Y}] + \mathbf{a}_0^{-1}\mathbf{A}E[\mathbf{R}] - \mathbf{a}_0^{-1}\mathbf{A}E\left[\hat{\lambda}\right] \tag{5.4.1}$$

Now, we know $E[\mathbf{Y}]=\mathbf{X}\beta$ from (5.1.2) and $E[\mathbf{R}]=\mathbf{A}^{\mathrm{T}}\beta$ from (5.3.1). So (5.4.1) becomes:

$$E\left[\hat{\beta}\right]=\mathbf{a}_0^{-1}\mathbf{X}^{\mathrm{T}}\Omega^{-1}\mathbf{X}\beta+\mathbf{a}_0^{-1}\mathbf{A}\mathbf{A}^{\mathrm{T}}\beta-\mathbf{a}_0^{-1}\mathbf{A}\,E\left[\hat{\lambda}\right]$$
$$=\mathbf{a}_0^{-1}\left(\mathbf{X}^{\mathrm{T}}\Omega^{-1}\mathbf{X}+\mathbf{A}\mathbf{A}^{\mathrm{T}}\right)\beta-\mathbf{a}_0^{-1}\mathbf{A}\,E\left[\hat{\lambda}\right] \qquad (5.4.2)$$

From (5.3.16) we have $\mathbf{a}_0=\mathbf{X}^{\mathrm{T}}\Omega^{-1}\mathbf{X}+\mathbf{A}\mathbf{A}^{\mathrm{T}}$ So then:

$$E\left[\hat{\beta}\right]=\beta-\mathbf{a}_0^{-1}\mathbf{A}\,E\left[\hat{\lambda}\right] \qquad (5.4.3)$$

Similarly, from (5.3.24),

$$E\left[\hat{\lambda}\right]=E[\mathbf{R}]+\left(\mathbf{A}^{\mathrm{T}}\mathbf{a}_0^{-1}\mathbf{A}\right)^{-1}\mathbf{A}^{\mathrm{T}}\mathbf{a}_0^{-1}\mathbf{X}^{\mathrm{T}}\Omega^{-1}\,E[\mathbf{Y}]-\left(\mathbf{A}^{\mathrm{T}}\mathbf{a}_0^{-1}\mathbf{A}\right)^{-1}\,E[\mathbf{R}] \qquad (5.4.4)$$

$$=\left(\mathbf{I}-\left(\mathbf{A}^{\mathrm{T}}\mathbf{a}_0^{-1}\mathbf{A}\right)^{-1}\right)E[\mathbf{R}]+\left(\mathbf{A}^{\mathrm{T}}\mathbf{a}_0^{-1}\mathbf{A}\right)^{-1}\mathbf{A}^{\mathrm{T}}\mathbf{a}_0^{-1}\mathbf{X}^{\mathrm{T}}\Omega^{-1}\,E[\mathbf{Y}]$$

$$=\left(\mathbf{I}-\left(\mathbf{A}^{\mathrm{T}}\mathbf{a}_0^{-1}\mathbf{A}\right)^{-1}\right)\mathbf{A}^{\mathrm{T}}\beta+\left(\mathbf{A}^{\mathrm{T}}\mathbf{a}_0^{-1}\mathbf{A}\right)^{-1}\mathbf{A}^{\mathrm{T}}\mathbf{a}_0^{-1}\mathbf{X}^{\mathrm{T}}\Omega^{-1}\mathbf{X}\beta$$

$$\left[\mathbf{A}^{\mathrm{T}}-\left(\mathbf{A}^{\mathrm{T}}\mathbf{a}_0^{-1}\mathbf{A}\right)^{-1}\mathbf{A}^{\mathrm{T}}+\left(\mathbf{A}^{\mathrm{T}}\mathbf{a}_0^{-1}\mathbf{A}\right)^{-1}\mathbf{A}^{\mathrm{T}}\mathbf{a}_0^{-1}\mathbf{X}^{\mathrm{T}}\Omega^{-1}\mathbf{X}\right]\beta \qquad (5.4.5)$$

but $\mathbf{X}^{\mathrm{T}}\Omega^{-1}\mathbf{X}=\mathbf{a}_0-\mathbf{A}\mathbf{A}^{\mathrm{T}}$, so then:

$$E\left[\hat{\lambda}\right]=\left[\mathbf{A}^{\mathrm{T}}-\left(\mathbf{A}^{\mathrm{T}}\mathbf{a}_0^{-1}\mathbf{A}\right)^{-1}\mathbf{A}^{\mathrm{T}}+\left(\mathbf{A}^{\mathrm{T}}\mathbf{a}_0^{-1}\mathbf{A}\right)^{-1}\mathbf{A}^{\mathrm{T}}\mathbf{a}_0^{-1}\mathbf{a}_0-\left(\mathbf{A}^{\mathrm{T}}\mathbf{a}_0^{-1}\mathbf{A}\right)^{-1}\left(\mathbf{A}^{\mathrm{T}}\mathbf{a}_0^{-1}\mathbf{A}\right)\mathbf{A}^{\mathrm{T}}\right]$$

$$\Rightarrow E\left[\hat{\lambda}\right]=\mathbf{A}^{\mathrm{T}}-\mathbf{A}^{\mathrm{T}}=\mathbf{0} \qquad (5.4.6)$$

So from (5.4.3), $E\left[\hat{\beta}\right]=\beta$, since the second term is zero by (5.4.6) and we can conclude that the estimator is unbiased. Application of Eq.(5.3.28) (see examples below) shows that the constraint information remains unchanged by the estimation process. Thus anything which was fixed before the experiment remains unchanged afterwards in spite of whatever information might have surfaced to suggest otherwise; and so it appears that we are not maximising the information potentially available from the experiment. When considered this way, we can see a philosophical weakness with the estimator. However, it is certainly *internally consistent* inasmuch as the deterministically-viewed constraint information would not be expected to change in the Lagrange Multipliers analysis. We would therefore not expect to learn anything new about the constraints in the experiment. This is further emphasised in the covariance matrix, where there is no term due to the **R** vector present [see (5.3.29) & (5.3.31) for example]. The result of this–see examples here and in Chapter 9–is that $\psi_{\hat{\beta}}$ is incomplete: variance and covariance terms due to the elements of **R** are missing. Some have proposed a design matrix **X** such that $\mathbf{X}^{\mathrm{T}}.\mathbf{X}$ is orthogonal (see Prowse & Anderson (1974), Grabe (1978), Zuker et al (1980) & Mihailov & Romanowski (1990)), in which case $\psi_{\hat{\beta}}$ would be expected to be diagonal, but in the present case this is not so, and thus the covariance matrix is incomplete. Indeed orthogonal systems may not always be a good idea. See for example. Morris (1992).

Another aspect on Design Schemes to deal with physical problems in calibration such as adsorption on metal surfaces is given in Ikeda (1986a) & (1986b). The internal consistency mentioned above is obtained with the price of a flawed and improper view of the true situation, in which the standards are not deterministic, and thus violates one of our key criteria of consistency by neglecting valid information. We will say more about this shortly but now let us consider two examples.

Example I

As an example, consider this design for three parameters:

$$\mathbf{X}=\begin{bmatrix} 1 & -1 & 0 \\ 1 & 0 & -1 \\ 0 & 1 & -1 \end{bmatrix} \quad \beta=\begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} \quad , \quad \mathbf{Y}=\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} \tag{5.4.7}$$

To keep things simple, we can let $\psi_Y = \sigma^2.\mathbf{I}$. Now suppose the constraint is the known value of one parameter, b_1, such that $b_1 = m_1$ and then we have

$$\mathbf{A}=\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \;\&\; \mathbf{R}=\begin{bmatrix} m_1 \end{bmatrix}$$

Evaluating Eq. (5.3.23) now yields:

$$\hat{\beta}=\begin{bmatrix} \hat{b}_1 \\ \hat{b}_2 \\ \hat{b}_3 \end{bmatrix} = \begin{bmatrix} m_1 \\ m_1-\frac{2}{3}y_1-\frac{1}{3}y_2+\frac{1}{3}y_3 \\ m_1-\frac{1}{3}y_1-\frac{2}{3}y_2-\frac{1}{3}y_3 \end{bmatrix} \tag{5.4.8a}$$

$$\text{while } \psi_{\hat{\beta}}=\sigma^2.\begin{bmatrix} 0 & 0 & 0 \\ 0 & \frac{2}{3} & \frac{1}{3} \\ 0 & \frac{1}{3} & \frac{2}{3} \end{bmatrix} \tag{5.4.8b}$$

In this case the estimated value of b_1 remains unchanged from the constraint value and, according to (5.4.8b), it has no variance or covariance terms.

Example II

As a second example, let the constraint information be:

$$b_1 + b_3 = m_{13} \tag{5.4.9}$$

This is somewhat unlikely in mass calibration, but is nonetheless valid as an example.

$$\mathbf{A}=\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \;\&\; \mathbf{R}=\begin{bmatrix} m_{13} \end{bmatrix}$$

We now obtain:

$$\hat{\beta} = \begin{bmatrix} \hat{b}_1 \\ \hat{b}_2 \\ \hat{b}_3 \end{bmatrix} = \begin{bmatrix} \frac{1}{2}m_{13} + \frac{1}{6}(y_1 + 2y_2 + y_3) \\ \frac{1}{2}m_{13} + \frac{1}{2}(y_3 - y_1) \\ \frac{1}{2}m_{13} - \frac{1}{6}(y_1 + 2y_2 + y_3) \end{bmatrix} \tag{5.4.10a}$$

$$\text{while } \psi_{\hat{\beta}} = \sigma^2 . \begin{bmatrix} \frac{1}{6} & 0 & -\frac{1}{6} \\ 0 & \frac{1}{2} & 0 \\ -\frac{1}{6} & 0 & \frac{1}{6} \end{bmatrix} \tag{5.4.10b}$$

From (5.4.10a) it is apparent that the sum $\hat{b}_1 + \hat{b}_2 = m_{13}$ remains unchanged and the combined variance of the sum is $s^2(\hat{b}_1) + s^2(\hat{b}_3) + 2s(\hat{b}_1, \hat{b}_3) = \left[\frac{1}{6} + \frac{1}{6} + 2\left(-\frac{1}{6}\right)\right].\sigma^2 = 0$; again showing that the constraint information is unchanged.

Now there is a big problem here because it is well known that the constraint information is really just the result of a previous calibration and is thus not a deterministic quantity, but rather a stochastic one with a particular (previously estimated) dispersion characteristic. This of course presents no problem to the traditional view of uncertainties since the uncertainty of the constraints–usually called "standards"–is treated as a systematic uncertainty and therefore considered as something which cannot be altered by the experiment and so does not need to be included in the analysis. It is simply added to the overall uncertainty figure at the end as a "systematic" component. In the formalism presented above, that component can be calculated from (5.3.28) where we have $\hat{\beta} = \mathbf{MY} + \mathbf{c_2 R}$. We have already developed the first component, so what is now needed is the complete uncertainty, including that due to the constraint information:

$$\psi_{\hat{\beta}} = \mathbf{M}\psi_{\mathbf{Y}}\mathbf{M}^{\mathbf{T}} + \mathbf{c}_2\psi_{\mathbf{R}}\mathbf{c}_2^{\mathbf{T}} \tag{5.4.11}$$

From (5.3.22) we know $\mathbf{c}_2 = \mathbf{a}_0^{-1}\mathbf{A}\left(\mathbf{A}^{\mathbf{T}}\mathbf{a}_0^{-1}\mathbf{A}\right)^{-1}$. Fig.(5.4.1) gives $\mathbf{c}_2$ for the two examples and Fig. (5.4.2) gives the resulting "systematic" uncertainty term. Note that $\psi_{\mathbf{R}} = \sigma^2_{r_1}$ in Eg. I while $\psi_{\mathbf{R}} = \sigma^2_{r_{13}}$ in Eg. II since there is just one piece of constraint information in each case.

$$\mathbf{c}_2 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}; \quad \mathbf{c}_2 = \begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} \\ \frac{1}{2} \end{bmatrix}$$

Eg. I Eg. II

Fig. 5.4.1

$$\sigma^2_{r_1}.\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \quad \sigma^2_{r_{13}}.\begin{bmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{bmatrix}$$

Eg. I Eg. II

Fig. 5.4.2 - "Systematic Components"

Combining (5.4.8b) & (5.4.10b) with their corresponding "systematic" terms from Fig. (5.4.2), we find the respective complete covariance matrices can now be given as in Fig. 5.4.3:

$$\Psi_{\hat{\beta}} = \begin{bmatrix} \sigma_{r_1}^2 & \sigma_{r_1}^2 & \sigma_{r_1}^2 \\ \sigma_{r_1}^2 & \frac{2}{3}\sigma^2 + \sigma_{r_1}^2 & \frac{2}{3}\sigma^2 + \sigma_{r_1}^2 \\ \sigma_{r_1}^2 & \frac{1}{3}\sigma^2 + \sigma_{r_1}^2 & \frac{2}{3}\sigma^2 + \sigma_{r_1}^2 \end{bmatrix} \qquad \Psi_{\hat{\beta}} = \begin{bmatrix} \frac{1}{4}\sigma_{r_{13}}^2 + \frac{1}{6}\sigma^2 & \frac{1}{4}\sigma_{r_{13}}^2 & \frac{1}{4}\sigma_{r_{13}}^2 - \frac{1}{6}\sigma^2 \\ \frac{1}{4}\sigma_{r_{13}}^2 & \frac{1}{4}\sigma_{r_{13}}^2 + \frac{1}{2}\sigma^2 & \frac{1}{4}\sigma_{r_{13}}^2 \\ \frac{1}{4}\sigma_{r_{13}}^2 - \frac{1}{6}\sigma^2 & \frac{1}{4}\sigma_{r_{13}}^2 & \frac{1}{4}\sigma_{r_{13}}^2 + \frac{1}{6}\sigma^2 \end{bmatrix}$$

Eg. I , Eg. II

Fig. 5.4.3 - Complete Covariance Matrices

However, this method is in serious disagreement with the Unified Approach to Uncertainty Analysis which we have developed in Chapters 1 & 2 and indeed contravenes the criteria of Consistent & Logical Reasoning which have been established. What we see here is the constraint information being treated deterministically in order to find a solution, and then being treated stochastically in order to find the correct final covariance matrix as shown in Fig. (5.4.3). This is at best an inconsistent approach and at worst a thoroughly inaccurate one! Since there **are** dispersion characteristics associated with the constraint vector, surely this information should be included in the estimation algorithm, as it may well influence the results obtained? This is an example of neglecting valid information which *a priori* is available, and thus contrary to the traditional view of being objective, would rather seem to be distinctly biased.

We will see in succeeding chapters several other estimation techniques which fully take account of all available information, including uncertainties/variances of the constraints, which are just treated as prior information to be included. We should observe that 'prior' in this context can be interpreted in a logical, rather than chronological manner; although in practice with mass standards it is in fact both. The results of doing this are a very distinct improvement over the Restrained Least Squares method outlined so far. In fact it is possible to *improve* the uncertainty estimate of the prior information through the new information obtained in the experiment. Thus the rationale for treating the prior uncertainties as "systematic" or "fixed" is undermined.

5.5 The Augmented Design Approach

The essential point about this new method (see Bich, (1990), Bich (1992) for e.g.) is that the constraints–necessary to obtain a solution–are viewed simply as data, for which expected values and variances are available. The fact that this data was not obtained in the current experiment is no obstacle as it is *logically* prior information. In the unified approach all information is a reflection of what we know about the parameters under investigation: whether this is new information or previous information is of no consequence. (Aside: there is plenty of evidence in mass metrology that mass standards drift over time. This additional information will modify our prior knowledge, sometimes significantly, and hence can affect the resulting parameter estimates. We will look at this in more detail in Chapter 10, while here we will develop the underlying theory.) The crucial point now is that while we still have

$$\mathbf{Y} = \mathbf{X}\beta \quad \text{with} \quad \mathbf{cov}[\mathbf{Y}] = \psi_Y \tag{5.5.1}$$

we now have for our constraint, or prior, information:

$$\mathbf{R} = \mathbf{A}^T\beta \quad \text{with} \quad \mathbf{cov}[\mathbf{R}] = \psi_R \tag{5.5.2}$$

We now augment both of these together to give:

$$\mathbf{Z} = \mathbf{W}\beta \tag{5.5.3}$$

$$\text{or} \begin{bmatrix} \mathbf{Y} \\ \mathbf{R} \end{bmatrix} = \begin{bmatrix} \mathbf{X} \\ \mathbf{A}^T \end{bmatrix} \beta \quad \text{while} \quad \psi_Z = \begin{bmatrix} \psi_Y & \mathbf{0} \\ \mathbf{0} & \psi_R \end{bmatrix} \tag{5.5.4}$$

We assume there are no correlations between the *mass differences* from the current experiment and the *mass values* from the prior data and hence ψ_z above is diagonal. Such correlations could be possible if the same instrumentation and reference standards were in use in each case, but unless they are carefully estimated and shown to be physically meaningful they should not be assumed. $\mathbf{W}^T\mathbf{W}$ is now no longer singular so we can solve the Gauss-Markov Minimum Variance, or Weighted Least Squares Estimator directly:

$$\boxed{\hat{\beta} = \left(\mathbf{W}^T\psi_z^{-1}\mathbf{W}\right)^{-1}\mathbf{W}^T\psi_Z^{-1}\mathbf{Z}} \tag{5.5.5}$$

$$\& \quad \boxed{\psi_{\hat{\beta}} = \left(\mathbf{W}^T\psi_Z^{-1}\mathbf{W}\right)^{-1}} \tag{5.5.6}$$

If we evaluate (5.5.5) we find, for $\mathbf{W}$, ψ_z & $\mathbf{Z}$ as given in (5.5.4), that:

$$\hat{\beta} = \left[\left(\mathbf{X}^T\psi_Y^{-1}\mathbf{X}\right) + \mathbf{A}\psi_R^{-1}\mathbf{A}^T\right]^{-1}\left[\mathbf{X}^T\psi_Y^{-1}\mathbf{Y} + \mathbf{A}\psi_R^{-1}\mathbf{R}\right] \tag{5.5.7}$$

This illustrates how the prior information features prominently in the estimates obtained via the augmented design—both $\mathbf{R}$ & ψ_R are present. Thus we expect this to be a complete solution giving adequate minimum variance estimators and,

importantly, a complete covariance matrix for these estimators. Eq. (5.5.7) indicates that the singularity of $\left(\mathbf{X}^{\mathrm{T}}\psi_{\mathrm{Y}}^{-1}\mathbf{X}\right)$ is no longer an issue, since the term to be inverted requires only the non-singularity of $\left[\left(\mathbf{X}^{\mathrm{T}}\psi_{\mathrm{Y}}^{-1}\mathbf{X}\right)+\mathbf{A}\psi_{\mathrm{R}}^{-1}\mathbf{A}^{\mathrm{T}}\right]$; and since the constraint/prior information is still linearly independent of the observation data, this condition will always be met.

This estimation method has some interesting properties. If we apply it to Example I in Sec 5.4 above, we will get the same estimates and a complete covariance matrix. We have $\psi_{\mathrm{R}}=\sigma_r^2$ and

$$\mathbf{Z}=\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ m_r \end{bmatrix};\ \mathbf{W}=\begin{bmatrix} 1 & -1 & 0 \\ 1 & 0 & -1 \\ 0 & 1 & -1 \\ 1 & 0 & 0 \end{bmatrix};\ \psi_{\mathbf{Z}}=\begin{bmatrix} \sigma^2 & 0 & 0 & 0 \\ 0 & \sigma^2 & 0 & 0 \\ 0 & 0 & \sigma^2 & 0 \\ 0 & 0 & 0 & \sigma_r^2 \end{bmatrix} \tag{5.5.8}$$

Resulting in:

$$\hat{\beta}\ =\ \begin{bmatrix} m_r \\ m_r-\frac{2}{3}y_1-\frac{1}{3}y_2+\frac{1}{3}y_3 \\ m_r-\frac{1}{3}y_1-\frac{2}{3}y_2-\frac{1}{3}y_3 \end{bmatrix}\ \ \&\ \ \psi_{\hat{\beta}}=\begin{bmatrix} \sigma_r^2 & \sigma_r^2 & \sigma_r^2 \\ \sigma_r^2 & \frac{2}{3}\sigma^2+\sigma_r^2 & \frac{2}{3}\sigma^2+\sigma_r^2 \\ \sigma_r^2 & \frac{1}{3}\sigma^2+\sigma_r^2 & \frac{2}{3}\sigma^2+\sigma_r^2 \end{bmatrix} \tag{5.5.9}$$

—obtained directly with Eqs. (5.5.5) & (5.5.6), without the need for any further processing. The prior information remains unchanged as it must since it is not possible to determine any further information about just one constraint.

Likewise for Example II of Sec. 5.4, $\mathbf{Z}$ and $\psi_{\mathbf{Z}}$ remain as given in (5.5.8) and

$$\mathbf{W}=\begin{bmatrix} 1 & -1 & 0 \\ 1 & 0 & -1 \\ 0 & 1 & -1 \\ 1 & 0 & 1 \end{bmatrix}$$

which is still just one piece of prior information. As before we obtain:

$$\hat{\beta}=\begin{bmatrix} \frac{1}{2}m_r+\frac{1}{6}(y_1+2y_2+y_3) \\ \frac{1}{2}m_r+\frac{1}{2}(y_3-y_1) \\ \frac{1}{2}m_r-\frac{1}{6}(y_1+2y_2+y_3) \end{bmatrix}\quad \psi_{\hat{\beta}}=\begin{bmatrix} \frac{1}{4}\sigma_r^2+\frac{1}{6}\sigma^2 & \frac{1}{4}\sigma_r^2 & \frac{1}{4}\sigma_r^2-\frac{1}{6}\sigma^2 \\ \frac{1}{4}\sigma_r^2 & \frac{1}{4}\sigma_r^2+\frac{1}{2}\sigma^2 & \frac{1}{4}\sigma_r^2 \\ \frac{1}{4}\sigma_r^2-\frac{1}{6}\sigma^2 & \frac{1}{4}\sigma_r^2 & \frac{1}{4}\sigma_r^2+\frac{1}{6}\sigma^2 \end{bmatrix} \tag{5.5.10}$$

These results are indeed the same as those obtained with Restrained Least Squares, but they have been obtained by a much more mathematically acceptable–not to mention simpler–means. The inclusion of the prior information **in** the estimation process, rather than just using it as a restraint **on** the estimation process is a much more unified and consistent use of the known information.

Now if we have more than one piece of independent prior information, the power of this method becomes much more apparent, because it is now possible to obtain extra knowledge on the prior information in the experiment, through extra comparisons; the minimum variance characteristics of the estimator then causing a new estimate of the prior information, of lower variance, to be obtained! This really shows what happens when we build the variability of the prior information into the model. If the constraints are not wholly deterministic why should we pretend they are? The example below shows the benefits of treating the constraints stochastically. We will first calculate the estimates by restrained least squares for comparison purposes.

Example III: We have four parameters:

$$\beta = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix} \quad \mathbf{X} = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 1 & 0 & -1 & 0 \\ 1 & 0 & 0 & -1 \\ 0 & 1 & -1 & 0 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \end{bmatrix} \quad \mathbf{Y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \end{bmatrix} \tag{5.5.11}$$

There are thus six observations on the four parameters. We take the covariance matrix of the observations to be $\psi_Y = \sigma^2 . \mathbf{I}_6$ for convenience. The constraint information concerns the values of b_1 & b_4. Thus we have:

$$\mathbf{A} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix}; \quad \mathbf{R} = \begin{bmatrix} m_1 \\ m_4 \end{bmatrix}; \quad \psi_R = \begin{bmatrix} \sigma_{r_1}^2 & 0 \\ 0 & \sigma_{r_2}^2 \end{bmatrix} \tag{5.5.12}$$

from which we can see that the two "reference" standards are not correlated. The restrained least squares solution now yields:

$$\hat{\beta} = \begin{bmatrix} m_1 \\ \frac{1}{8}(4m_1 + 4m_4 - 3y_1 - 2y_2 + 2y_4 + 3y_5 + y_6) \\ \frac{1}{8}(4m_1 + 4m_4 - y_1 - 3y_2 - 2y_4 + y_5 + 3y_6) \\ m_4 \end{bmatrix} \tag{5.5.13a}$$

while for the covariance information we have:

$$\psi_{\hat{\beta}} = \mathbf{c}_1 \mathbf{X}^T \psi_Y^{-1} \mathbf{X} \mathbf{c}_1^T = \mathbf{c}_1 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & \frac{3}{8}\sigma^2 & \frac{1}{8}\sigma^2 & 0 \\ 0 & \frac{1}{8}\sigma^2 & \frac{3}{8}\sigma^2 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad \& \ \mathbf{c}_2 = \begin{bmatrix} 1 & 0 \\ \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \\ 0 & 1 \end{bmatrix} \tag{5.5.13b}$$

and for the constraint information, the uncertainty contribution is:

$$\mathbf{c_2}\psi_{\mathbf{R}}\mathbf{c_2^T} = \begin{bmatrix} \sigma_{r_1}^2 & \frac{1}{2}\sigma_{r_1}^2 & \frac{1}{2}\sigma_{r_1}^2 & 0 \\ \frac{1}{2}\sigma_{r_1}^2 & \frac{1}{4}\left(\sigma_{r_1}^2+\sigma_{r_2}^2\right) & \frac{1}{4}\left(\sigma_{r_1}^2+\sigma_{r_2}^2\right) & \frac{1}{2}\sigma_{r_2}^2 \\ \frac{1}{2}\sigma_{r_1}^2 & \frac{1}{4}\left(\sigma_{r_1}^2+\sigma_{r_2}^2\right) & \frac{1}{4}\left(\sigma_{r_1}^2+\sigma_{r_2}^2\right) & \frac{1}{2}\sigma_{r_2}^2 \\ 0 & \frac{1}{2}\sigma_{r_2}^2 & \frac{1}{2}\sigma_{r_2}^2 & \sigma_{r_2}^2 \end{bmatrix} \qquad (5.5.13c)$$

Thus, as expected, the values of b_1 & b_4 remain unchanged, as do their variances. The much simpler calculation of the augmented design scheme uses the following inputs:

$$\mathbf{W} = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 1 & 0 & -1 & 0 \\ 1 & 0 & 0 & -1 \\ 0 & 1 & -1 & 0 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}; \mathbf{Z} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ m_1 \\ m_4 \end{bmatrix} \quad \& \quad \psi_{\mathbf{Z}} = \begin{bmatrix} \sigma^2.\mathbf{I}_6 & \mathbf{0} \\ \mathbf{0} & \begin{bmatrix} \sigma_{r_1}^2 & 0 \\ 0 & \sigma_{r_2}^2 \end{bmatrix} \end{bmatrix} \qquad (5.5.14)$$

to yield the (more complex!) results given below:

$$\hat{\beta} = \begin{bmatrix} \frac{4\left(m_4\sigma_{r_1}^2+m_1\sigma_{r_2}^2\right)+2m_1\sigma^2+\sigma_{r_1}^2\left(y_1+y_2+2y_3+y_5+y_6\right)}{2\left(2\sigma_{r_1}^2+2\sigma_{r_2}^2+\sigma^2\right)} \\ \frac{16\left(m_4\sigma_{r_1}^2+m_1\sigma_{r_2}^2\right)+4\sigma^2\left(m_1+m_4\right)+4\sigma_{r_1}^2\left(-y_1+y_3+y_4+2y_5+y_6\right)+4\sigma_{r_2}^2\left(-2y_1-y_2-y_3+y_4+y_5\right)+\sigma^2\left(-3y_1-y_2+2y_4+3y_5+y_6\right)}{8\left(2\sigma_{r_1}^2+2\sigma_{r_2}^2+\sigma^2\right)} \\ \frac{16\left(m_4\sigma_{r_1}^2+m_1\sigma_{r_2}^2\right)+4\sigma^2\left(m_1+m_4\right)+4\sigma_{r_1}^2\left(-y_2+y_3-y_4+y_5+2y_6\right)+4\sigma_{r_2}^2\left(-y_1-2y_2-y_3-y_4+y_6\right)+\sigma^2\left(-y_1-3y_2-2y_4+y_5+3y_6\right)}{8\left(2\sigma_{r_1}^2+2\sigma_{r_2}^2+\sigma^2\right)} \\ \frac{4\left(m_4\sigma_{r_1}^2+m_1\sigma_{r_2}^2\right)+2m_4\sigma^2+\sigma_{r_2}^2\left(-y_1-y_2-2y_3-y_5-y_6\right)}{2\left(2\sigma_{r_1}^2+2\sigma_{r_2}^2+\sigma^2\right)} \end{bmatrix}$$

$$\psi_{\hat{\beta}} = \begin{pmatrix} \frac{\sigma_{r_1}^2\left(2\sigma_{r_2}^2+\sigma^2\right)}{2\sigma_{r_1}^2+2\sigma_{r_2}^2+\sigma^2} & \frac{\sigma_{r_1}^2\left(4\sigma_{r_2}^2+\sigma^2\right)}{2\left(2\sigma_{r_1}^2+2\sigma_{r_2}^2+\sigma^2\right)} & \frac{\sigma_{r_1}^2\left(4\sigma_{r_2}^2+\sigma^2\right)}{2\left(2\sigma_{r_1}^2+2\sigma_{r_2}^2+\sigma^2\right)} & \frac{2\sigma_{r_1}^2\sigma_{r_2}^2}{2\sigma_{r_1}^2+2\sigma_{r_2}^2+\sigma^2} \\ & \frac{8\sigma_{r_1}^2\left(2\sigma_{r_2}^2+\sigma^2\right)+\sigma^2\left(8\sigma_{r_2}^2+3\sigma^2\right)}{8\left(2\sigma_{r_1}^2+2\sigma_{r_2}^2+\sigma^2\right)} & \frac{\left(4\sigma_{r_1}^2+\sigma^2\right)\left(4\sigma_{r_2}^2+\sigma^2\right)}{8\left(2\sigma_{r_1}^2+2\sigma_{r_2}^2+\sigma^2\right)} & \frac{\sigma_{r_2}^2\left(4\sigma_{r_1}^2+\sigma^2\right)}{2\left(2\sigma_{r_1}^2+2\sigma_{r_2}^2+\sigma^2\right)} \\ & & \frac{8\sigma_{r_1}^2\left(2\sigma_{r_2}^2+\sigma^2\right)+\sigma^2\left(8\sigma_{r_2}^2+3\sigma^2\right)}{8\left(2\sigma_{r_1}^2+2\sigma_{r_2}^2+\sigma^2\right)} & \frac{\sigma_{r_2}^2\left(4\sigma_{r_1}^2+\sigma^2\right)}{2\left(2\sigma_{r_1}^2+2\sigma_{r_2}^2+\sigma^2\right)} \\ & & & \frac{\sigma_{r_2}^2\left(2\sigma_{r_1}^2+\sigma^2\right)}{2\left(2\sigma_{r_1}^2+2\sigma_{r_2}^2+\sigma^2\right)} \end{pmatrix}$$

We can see from this that b_1 & b_4 have been changed from their previous values. This is merely reflecting their status as "observations" with a mean and variance; so it is not surprising that it is possible to change their values if new information comes to light. What is particularly interesting is the covariance matrix, $\psi_{\hat{\beta}}$ above, from which we can see that:

$$\sigma^2\left(\hat{b}_1\right) = \frac{\sigma_{r_1}^2\left(2\sigma_{r_2}^2 + \sigma^2\right)}{2\sigma_{r_1}^2 + 2\sigma_{r_2}^2 + \sigma^2} < \sigma_{r_1}^2 \qquad (5.5.15)$$

Since the denominator $2\sigma_{r_1}^2 + 2\sigma_{r_2}^2 + \sigma^2 \quad > \quad \left(2\sigma_{r_2}^2 + \sigma^2\right)$ which appears in the numerator. Thus the new estimated variance for $\hat{b}_1$ will be *smaller* than the variance for the original b_1.

So we can see that the Augmented Design approach, which is a more appropriate way to view the problem considering the true, stochastic nature of all the information available, is also a superior method inasmuch as it can effect a reduced covariance matrix, which Restrained Least Squares cannot do. This is as a result of making full use of all the available information. It is also, as we have already remarked, a more agreeable method, in view of our Uniform Approach to Uncertainty Analysis already used to calculate $\psi_{\mathbf{Y}}$, the covariance matrix of the input data. Further numerical examples will be given in later sections to highlight these methods.

But now we proceed to look at a more generalised estimation technique, of which the two methods outlined in this section are special cases. We will see the significance of a proper view of the extra constraint information in what follows.

****

6. A Generalised Estimation Method

6.0 Summary

A different, generalised estimation technique, with a different theoretical basis to either of the two methods discussed in Chapter 5 is introduced in this chapter. The method is developed by Rao (chapter 4 in Rao (1973)) and known as the Generalised Gauss Markov method. We will see that this model is entirely general and does not make any assumptions about the data, the input covariance matrix or the system model. (In contrast to RLS which requires constraints, and AD which needs an invertible covariance matrix). The GGM technique utilises the matrix theory of *generalised inverses* (sometimes called *pseudo-inverses,* or *g-inverses*)

We show how the method is implemented by forming an augmented matrix, as shown in Fig. (6.0.1) below. The details of the estimation technique are discussed within the chapter, but the principal point to note here is the form of the solutions obtained, illustrated in the figure below. Note that, in the presence of constraints, an extended model can be written down in terms of **Z**, **W** & β as in Fig. (5.0.3), rather than **Y**, **X** & β as below. The GGM model can deal with either. The exact form of the solutions will depend upon the details of the input data and we leave these specifics to Chapter 7.

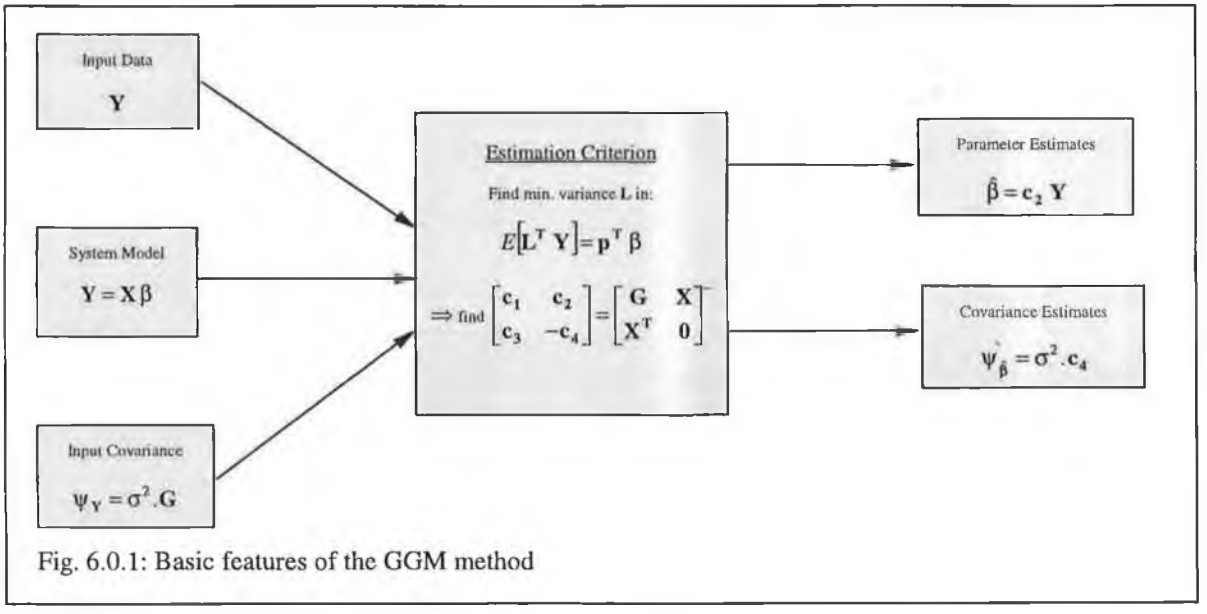


Fig. 6.0.1: Basic features of the GGM method

6.1 Introduction

So far we have seen two principal parameter estimation techniques: RLS & the AD method. In the former, the singularity in $\left(\mathbf{X}^T\psi_Y^{-1}\mathbf{X}\right)$ is dealt with by solving the normal equations subject to a set of constraints which are linearly independent of the observations. The constraints used in mass calibration are usually the previously known values of [some of] the parameters. This allows a solution to be found, but we pointed out that it is both inconsistent and inaccurate in its use of the constraint information which is in fact not deterministic but has previously determined dispersion characteristics.

We then saw how the AD method allows this additional knowledge to be utilised in obtaining a full solution requiring no further post-estimation calculations and indeed allowing the possibility of arriving at a smaller dispersion matrix for the parameters than would otherwise be possible.

One would wonder if the AD method could be applied even if the constraints are considered deterministic. Perhaps in this case an estimate could indeed be obtained but no further information on the prior knowledge could be found since it is considered fixed. This, unfortunately, is impossible since the solution

$$\hat{\beta}_{AD} = \left(\mathbf{X}^T\psi_Y^{-1}\mathbf{X}\right)^{-1}\mathbf{X}^T\psi_Y^{-1}\mathbf{Y}$$

determined to be the BLUE by the G-M theorem, cannot be so determined if ψ_Y is singular, as the G-M theorem does not hold under such conditions. For example, with our model $\mathbf{Y} = \mathbf{X}.\beta$, and prior information $\mathbf{A}^T.\beta = \mathbf{R}$, we form the augmented design:

$$\mathbf{W} = \begin{bmatrix} \mathbf{X} \\ \mathbf{A}^T \end{bmatrix};\ \mathbf{Z} = \begin{bmatrix} \mathbf{Y} \\ \mathbf{R} \end{bmatrix};\ \&\ \psi_Z = \begin{bmatrix} \psi_Y & \mathbf{0} \\ \mathbf{0} & \psi_R \end{bmatrix}$$

But if $\mathbf{R}$ is deterministic then $\psi_R = \mathbf{0}$ and then we would have

$$\psi_Z = \begin{bmatrix} \psi_Y & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$$

which is clearly singular and we can proceed no further. So we need a General Solution which can deal with the possibility of a singular covariance matrix, ψ_Y, and it is to the development of such a method that we now turn.

6.2 The Generalised Gauss-Markov Model (See Chapter 4 in Rao, 1973)

We consider the model $\mathbf{Y} = \mathbf{X}\beta$ where $\mathbf{Y}$ is a vector of experimental observations subject to a dispersion matrix $\psi_Y = \sigma^2\mathbf{G}$. We wish to find the best–i.e. Minimum Variance–linear unbiased estimator of β. Let this estimator be $\hat{\beta}$ such that:

$$E\left[\mathbf{p}^T\hat{\beta}\right] = \mathbf{p}^T\beta \tag{6.2.1}$$

for some suitable vector **p**. Note this is most general and in many cases we need not be concerned about the form of **p**, but we leave it here for generality. Our observations however, are the vector **Y**, so our parameter estimate must come from this source as this is all the information on which we can make decisions (although we do not exclude the possibility that there may be a constraint vector, **R**, needed as well.) Thus we need an estimator **L** such that our estimator is a linear combination of our observations:

$$E\left[\mathbf{L}^T\mathbf{Y}\right] = \mathbf{p}^T\beta \tag{6.2.2}$$

$$\text{i.e. } \mathbf{L}^T\mathbf{X}\beta = \mathbf{p}^T\beta$$

$$\text{or } \mathbf{L}^T\mathbf{X} = \mathbf{p}^T$$

$$\Rightarrow \mathbf{X}^T\mathbf{L} = \mathbf{p} \tag{6.2.3}$$

This means that **p** is a linear combination of $\mathbf{X}^T$, or that **p** lies in the vector space spanned by the columns of $\mathbf{X}^T\mathbf{X}$.

Now **L** in (6.2.2) has been chosen such that it is a linear unbiased estimator of $\mathbf{p}^T\beta$, but we want the minimum variance estimator out of the class of all possible unbiased estimators. How do we find this? Observe that the variance of our estimator is:

$$\mathbf{var}\left[\mathbf{L}^T\mathbf{Y}\right] = \sigma^2\mathbf{L}^T\mathbf{G}\mathbf{L} \tag{6.2.4}$$

since $\sigma^2\mathbf{G}$ = **var**[**Y**]. Thus the best estimate for **L** is the one for which (6.2.4) is minimal. Suppose this optimum choice is **M**, chosen such that:

$$\mathbf{X}^T\mathbf{L} = \mathbf{X}^T\mathbf{M} \tag{6.2.5}$$

which of course follows from (6.2.3) since all the valid estimators are among the class of linear unbiased ones and therefore satisfy (6.2.2) & (6.2.3). Then we can say:

$$\mathbf{L}^T\mathbf{G}\mathbf{L} = \left[(\mathbf{L}-\mathbf{M})+\mathbf{M}\right]^T\mathbf{G}\left[(\mathbf{L}-\mathbf{M})+\mathbf{M}\right] \tag{6.2.6}$$

$$= (\mathbf{L}-\mathbf{M})^T\mathbf{G}(\mathbf{L}-\mathbf{M}) + (\mathbf{L}-\mathbf{M})^T\mathbf{G}\mathbf{M} + \mathbf{M}^T\mathbf{G}(\mathbf{L}-\mathbf{M}) + \mathbf{M}^T\mathbf{G}\mathbf{M} \tag{6.2.7}$$

Note, as an aside, that given **AB** = **C** for **A**, **B** & **C** appropriate matrices, that in general $(\mathbf{AB})^T = \mathbf{B}^T\mathbf{A}^T = \mathbf{C}^T$. Thus in (6.2.7) we have:

$$(\mathbf{L}-\mathbf{M})^T\mathbf{G}\mathbf{M} = \left[\mathbf{M}^T\mathbf{G}(\mathbf{L}-\mathbf{M})\right]^T \tag{6.2.8}$$

where **G** is symmetric. Now with the strict condition that:

$$\textit{iff}\quad \mathbf{M}^T\mathbf{G}(\mathbf{L}-\mathbf{M}) = \mathbf{0} \quad \text{when} \quad \mathbf{X}^T(\mathbf{L}-\mathbf{M}) = \mathbf{0} \tag{6.2.9}$$

—which latter we know is true from (6.2.5); then we can say that:

$$\mathbf{L}^T\mathbf{G}\mathbf{L} = (\mathbf{L}-\mathbf{M})^T\mathbf{G}(\mathbf{L}-\mathbf{M}) + \mathbf{M}^T\mathbf{G}\mathbf{M} \tag{6.2.10}$$

since if (6.2.9) is true, both $\mathbf{M}^T\mathbf{G}(\mathbf{L}-\mathbf{M})$ & $(\mathbf{L}-\mathbf{M})^T\mathbf{G}\mathbf{M}$ are zero when **M** is such that (6.2.5) is true. Therefore we can see that:

$$\boxed{\mathbf{L^TGL \geq M^TGM}} \tag{6.2.11}$$

which means that the particular estimator **M** is always of lower covariance than any estimator **L**; in other words it will generate the minimum variance estimate of $\mathbf{p^T\beta}$, which is $\mathbf{p^T\hat{\beta}}$. From (6.2.9), the condition for this to be true is that

$$\mathbf{M^TG(L-M) = X^T(L-M) = 0} \tag{6.2.12}$$

or that $\mathbf{M^T.G}$ is some linear combination of $\mathbf{X^T}$. We can then say:

$$\begin{aligned} &\mathbf{M^TG = -k^TX^T} \\ \text{or}\quad &\mathbf{GM = -Xk} \end{aligned} \tag{6.2.13}$$

for some appropriate matrix **k**. From (6.2.5) & (6.2.3) we know that $\mathbf{X^TM = p}$, so we can now write the two equations:

$$\mathbf{GM + Xk = 0} \tag{6.2.14a}$$

$$\mathbf{X^TM = p} \tag{6.2.14b}$$

$$\text{or} \begin{bmatrix} \mathbf{G} & \mathbf{X} \\ \mathbf{X^T} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{M} \\ \mathbf{k} \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{p} \end{bmatrix} \tag{6.2.15}$$

for **M** a BLUE of β in the model $\mathbf{Y = X\beta}$ with $\psi_Y = \sigma^2\mathbf{G}$. Note that we made no assumptions so far about the form of **G** or indeed about the form of $\mathbf{X^TX}$ as do both AD & RLS. From (6.2.15) we can say:

$$\begin{bmatrix} \mathbf{M} \\ \mathbf{k} \end{bmatrix} = \begin{bmatrix} \mathbf{c_1} & \mathbf{c_2} \\ \mathbf{c_3} & \mathbf{-c_4} \end{bmatrix} \begin{bmatrix} \mathbf{0} \\ \mathbf{p} \end{bmatrix} \tag{6.2.16}$$

where $\begin{bmatrix} \mathbf{c_1} & \mathbf{c_2} \\ \mathbf{c_3} & \mathbf{-c_4} \end{bmatrix}$ is the Generalised Inverse–or g-inverse–of $\begin{bmatrix} \mathbf{G} & \mathbf{X} \\ \mathbf{X^T} & \mathbf{0} \end{bmatrix}$. (See also Goldman & Zelen (1964)) For any matrix **A**, its g-inverse is denoted $\mathbf{A^-}$, and is defined such that $\mathbf{AA^-A = A}$. If **A** is of full rank then it is not singular, and $\mathbf{A^-}$ is the normal inverse. If **A** is not of full rank, or rectangular, then independent rows, **H**, can be added to **A** so that:

$$\mathbf{A_a} = \begin{bmatrix} \mathbf{A} \\ \mathbf{H} \end{bmatrix}$$

Then the inverse of the augmented matrix

$$\begin{bmatrix} \mathbf{A} & \mathbf{H} \\ \mathbf{H^T} & \mathbf{0} \end{bmatrix}$$

exists and in fact

$$\begin{bmatrix} \mathbf{A} & \mathbf{H} \\ \mathbf{H^T} & \mathbf{0} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{c_1} & \mathbf{c_2^T} \\ \mathbf{c_2} & \mathbf{c_4} \end{bmatrix}$$

In this case $\mathbf{c_1}$ is a g-inverse of **A** and $\mathbf{Ac_1A = A}$. Note: this procedure is essentially analogous to that employed in Restrained Least Squares where $\mathbf{A = X^TX}$ and **H** is the design matrix of the constraints in $\mathbf{H\beta = R}$. (See Section 5.3).

In this theory, however, a g-inverse is not unique. Any matrix **H** which allows the inverse of the augmented matrix above to be computed will be sufficient to evaluate $\mathbf{c}_1$; however, in application, **H** would of course have to be chosen so as to be physically meaningful since it directly influences the results obtained.

In the present problem, without going into the details of the evaluation of the g-inverse, we can immediately say that (6.2.16) gives:

$$\begin{bmatrix}\mathbf{M}\\ \mathbf{k}\end{bmatrix} = \begin{bmatrix}\mathbf{c}_2.\mathbf{p}\\ -\mathbf{c}_4.\mathbf{p}\end{bmatrix} \quad \text{or} \quad \begin{matrix}\mathbf{M}=\mathbf{c}_2\mathbf{p}\\ \mathbf{k}=-\mathbf{c}_4\mathbf{p}\end{matrix} \tag{6.2.17}$$

We will investigate the evaluation of $\mathbf{c}_2$ & $\mathbf{c}_4$ for specific examples in Chapter 7. Recall from Eq. (6.2.2) that we chose a linear unbiased estimator **L** such that:

$$\mathbf{L}^T\mathbf{X}\beta = \mathbf{p}^T\beta$$

and we chose **M** as the best (i.e. minimum variance) linear unbiased estimator–that is, the best form of **L**–which led to the conditions expressed in (6.2.9), (6.2.12) & (6.2.13). Since this is the best estimator, we must be able to equate (6.2.2) & (6.2.1) for **M**:

$$E\left[\mathbf{p}^T\hat{\beta}\right] = E\left[\mathbf{M}^T\mathbf{Y}\right] = \mathbf{p}^T\beta \tag{6.2.18}$$

where $\hat{\beta}$ is the BLUE for β, i.e. $\mathbf{p}^T\hat{\beta} = \mathbf{M}^T\mathbf{Y}$ Thus from (6.2.17), we have:

$$\mathbf{p}^T\hat{\beta} = \mathbf{p}^T\mathbf{c}_2^T\mathbf{Y}$$

$$\Rightarrow \boxed{\hat{\beta} = \mathbf{c}_2^T.\mathbf{Y}} \tag{6.2.19}$$

This is a generalised solution for the best parameter estimate for a Generalised Gauss-Markov model; where we have $\mathbf{Y} = \mathbf{X}\beta$ and a covariance matrix $\psi_Y = \sigma^2\mathbf{G}$ whose form is as yet unspecified. We also need to get the covariance matrix of $\hat{\beta}$ and then look at the specific form of $\mathbf{c}_2$, $\mathbf{c}_4$ for our experimental problem; but before we do this we need some results on the g-inverse in (6.2.16)

6.3 Results on the G-inverse

For some variables **a** & **b**, we may write:

$$\mathbf{G}\mathbf{a} + \mathbf{X}\mathbf{b} = \mathbf{0} \tag{6.3.1a}$$

$$\mathbf{X}^T\mathbf{a} \quad = \mathbf{X}^T\mathbf{d} \tag{6.3.1b}$$

—analogous to Eqs (6.2.14a) & (6.2.14b), for some appropriate **d**. In other words:

$$\begin{bmatrix}\mathbf{G} & \mathbf{X}\\ \mathbf{X}^T & \mathbf{0}\end{bmatrix}\begin{bmatrix}\mathbf{a}\\ \mathbf{b}\end{bmatrix} = \begin{bmatrix}\mathbf{0}\\ \mathbf{X}^T\mathbf{d}\end{bmatrix}$$

$$\Rightarrow \begin{bmatrix}\mathbf{a}\\ \mathbf{b}\end{bmatrix} = \begin{bmatrix}\mathbf{c}_1 & \mathbf{c}_2\\ \mathbf{c}_3 & -\mathbf{c}_4\end{bmatrix}\begin{bmatrix}\mathbf{0}\\ \mathbf{X}^T.\mathbf{d}\end{bmatrix} \tag{6.3.2}$$

where $\begin{bmatrix} \mathbf{c}_1 & \mathbf{c}_2 \\ \mathbf{c}_3 & -\mathbf{c}_4 \end{bmatrix} = \begin{bmatrix} \mathbf{G} & \mathbf{X} \\ \mathbf{X}^T & \mathbf{0} \end{bmatrix}^-$ is a g-inverse as before. Then we have from (6.3.2):

$$\begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix} = \begin{bmatrix} \mathbf{c}_2\mathbf{X}^T\mathbf{d} \\ -\mathbf{c}_4\mathbf{X}^T\mathbf{d} \end{bmatrix} \qquad (6.3.3)$$

$\Rightarrow$ (6.3.1a) & (6.3.1b) can be expressed as:

$$\mathbf{G}\mathbf{c}_2\mathbf{X}^T\mathbf{d} - \mathbf{X}\mathbf{c}_4\mathbf{X}^T\mathbf{d} = \mathbf{0}$$
$$\& \ \mathbf{X}^T\mathbf{c}_2\mathbf{X}^T\mathbf{d} = \mathbf{X}^T\mathbf{d}$$

$$\Rightarrow \ \mathbf{G}\mathbf{c}_2\mathbf{X}^T = \mathbf{X}\mathbf{c}_4\mathbf{X}^T \qquad (6.3.4a) \rightarrow \text{Result (a)}$$
$$\& \ \mathbf{X}^T\mathbf{c}_2\mathbf{X}^T = \mathbf{X}^T \qquad (6.3.4b) \rightarrow \text{Result (b)}$$

The latter of course implies that $\mathbf{c}_2$ is a g-inverse of $\mathbf{X}^T$. Now noting that **A.B.A = A** $\Rightarrow \mathbf{A}^T\mathbf{B}^T\mathbf{A}^T = \mathbf{A}^T$, we can write

$$\begin{bmatrix} \mathbf{G} & \mathbf{X}^T \\ \mathbf{X} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{c}_1^T & \mathbf{c}_3^T \\ \mathbf{c}_2^T & -\mathbf{c}_4^T \end{bmatrix} \begin{bmatrix} \mathbf{G} & \mathbf{X}^T \\ \mathbf{X} & \mathbf{0} \end{bmatrix} = \begin{bmatrix} \mathbf{G} & \mathbf{X}^T \\ \mathbf{X} & \mathbf{0} \end{bmatrix}$$

(in which **G** is of course symmetric) which means that $\begin{bmatrix} \mathbf{c}_1 & \mathbf{c}_2 \\ \mathbf{c}_3 & -\mathbf{c}_4 \end{bmatrix}^T$ is also a g-inverse and therefore:

$$\mathbf{c}_2 \equiv \mathbf{c}_3^T \qquad (6.3.5a)$$
$$\mathbf{c}_4 = \mathbf{c}_4^T \qquad (6.3.5b)$$

Thus (6.3.4a) and (6.3.4b) now become:

$$\mathbf{G}\mathbf{c}_3^T\mathbf{X}^T = \mathbf{X}\mathbf{c}_4^T\mathbf{X}^T \qquad (6.3.6a) \rightarrow \text{Result (c)}$$
$$\& \ \mathbf{X}^T\mathbf{c}_3^T\mathbf{X}^T = \mathbf{X}^T \qquad (6.3.6b) \rightarrow \text{Result (d)}$$

If we multiply (6.3.6a) by $\mathbf{X}.\mathbf{c}_3$ we get:

$$\mathbf{X}\mathbf{c}_3\mathbf{G}\mathbf{c}_3^T\mathbf{X}^T = \mathbf{X}\mathbf{c}_3\mathbf{X}\mathbf{c}_4^T\mathbf{X}^T$$
$$= \mathbf{X}\mathbf{c}_4^T\mathbf{X}^T \text{ by (6.3.6b)}$$

But $\mathbf{c}_4 = \mathbf{c}_4^T$ according to (6.3.5b) so (6.3.6b) now indicates that $\mathbf{X}\mathbf{c}_4\mathbf{X}^T$ is symmetric, i.e.:

$$\mathbf{X}\mathbf{c}_4^T\mathbf{X}^T = \mathbf{X}^T\mathbf{c}_4\mathbf{X}^T$$

Thus we now have:

$$\boxed{\mathbf{X}\mathbf{c}_2^T\mathbf{X} = \mathbf{X}\mathbf{c}_3\mathbf{X} = \mathbf{X}} \qquad \text{Result I}$$

—by means of (6.3.4a) & (6.3.4b); and also:

$$\boxed{\begin{aligned} &\mathbf{X}\mathbf{c}_4\mathbf{X}^T = \mathbf{G}\mathbf{c}_2\mathbf{X}^T = \mathbf{G}\mathbf{c}_3^T\mathbf{X}^T \\ &= \mathbf{X}\mathbf{c}_4^T\mathbf{X}^T = \mathbf{X}\mathbf{c}_2^T\mathbf{G} = \mathbf{X}\mathbf{c}_3\mathbf{G} \end{aligned}} \qquad \text{Result II}$$

Of course we could also write (6.3.1a) & (6.3.1b) as:

$$\mathbf{Ga} + \mathbf{Xb} = \mathbf{Xd} \tag{6.3.7a}$$

$$\& \quad \mathbf{X}^T\mathbf{a} = \mathbf{0} \tag{6.3.7b}$$

Treating these similarly to (6.3.1a) & (6.3.1b) gives:

$$\begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix} = \begin{bmatrix} \mathbf{c}_1\mathbf{Xd} \\ \mathbf{c}_3\mathbf{Xd} \end{bmatrix}$$

which gives for (6.3.7a) & (6.3.7b):

$$\mathbf{Gc}_1\mathbf{Xd} + \mathbf{Xc}_3\mathbf{Xd} = \mathbf{Xd}$$

$$\& \ \mathbf{X}^T\mathbf{c}_1\mathbf{Xd} = \mathbf{0}$$

$$\Rightarrow \mathbf{Gc}_1\mathbf{X} + \mathbf{Xc}_3\mathbf{X} = \mathbf{X}$$

by Result I we have already established that $\mathbf{Xc}_3\mathbf{X} = \mathbf{X}$

$$\therefore \boxed{\mathbf{Gc}_1\mathbf{X} = \mathbf{X}^T\mathbf{c}_1\mathbf{X} = \mathbf{0}} \qquad \text{Result III}$$

that is, they are all null matrices.

6.4 Covariance in the GGM Model

In (6.2.18) we had $E\left[\mathbf{p}^T\hat{\beta}\right] = E\left[\mathbf{M}^T\mathbf{Y}\right] = \mathbf{p}^T\beta$ to give the best estimator $\hat{\beta}$ via a linear combination of $\mathbf{Y}$. Further, analogous to (6.2.3) we have:

$$\mathbf{X}^T\mathbf{M} = \mathbf{p} \tag{6.4.1}$$

Now in (6.2.19) we had $\hat{\beta} = \mathbf{c}_2^T\mathbf{Y}$,

$$\therefore \mathbf{var}\left[\mathbf{p}^T\hat{\beta}\right] = \mathbf{var}\left[\mathbf{p}^T\mathbf{c}_2^T\mathbf{Y}\right] = \mathbf{var}\left[\mathbf{M}^T\mathbf{Xc}_2^T\mathbf{Y}\right] \text{ by (6.4.1)}$$

$$= \sigma^2\mathbf{M}^T\mathbf{Xc}_2^T\mathbf{Gc}_2\mathbf{X}^T\mathbf{M} \quad \text{as } \mathbf{var}[\mathbf{Y}] = \sigma^2\mathbf{G}$$

$$= \sigma^2\mathbf{M}^T\left(\mathbf{Xc}_2^T\mathbf{G}\right)\mathbf{c}_2\mathbf{X}^T\mathbf{M}$$

$$= \sigma^2\mathbf{M}^T\left(\mathbf{Xc}_4\mathbf{X}^T\right)\mathbf{c}_2\mathbf{X}^T\mathbf{M} \text{ by Result II}$$

$$= \sigma^2\mathbf{M}^T\mathbf{Xc}_4\left(\mathbf{X}^T\mathbf{c}_2\mathbf{X}^T\right)\mathbf{M}$$

$$= \sigma^2\mathbf{M}^T\mathbf{Xc}_4\mathbf{X}^T\mathbf{M} \quad \text{by Result I}$$

$$\text{i.e. } \mathbf{var}\left[\mathbf{p}^T\hat{\beta}\right] = \sigma^2\mathbf{p}^T\mathbf{c}_4\mathbf{p}$$

$$\Rightarrow \boxed{\mathbf{var}\left[\hat{\beta}\right] = \sigma^2\mathbf{c}_4} \tag{6.4.2}$$

So, in conclusion, in the General Gauss-Markov model–GGM–$\{\mathbf{Y},\mathbf{X}.\beta,\sigma^2\mathbf{G}\}$, we look for a BLUE for β, given by $\mathbf{p}^T.\hat{\beta}$. This is estimated by a linear combination of the observations, $\mathbf{M}^T\mathbf{Y}$ such that:

$$\begin{bmatrix} \mathbf{G} & \mathbf{X} \\ \mathbf{X}^T & \mathbf{0} \end{bmatrix}\begin{bmatrix} \mathbf{M} \\ \mathbf{k} \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{p} \end{bmatrix} \text{ for a suitable } \mathbf{k}.$$

Then for $\begin{bmatrix} \mathbf{c}_1 & \mathbf{c}_2 \\ \mathbf{c}_3 & -\mathbf{c}_4 \end{bmatrix}$ the g-inverse of $\begin{bmatrix} \mathbf{G} & \mathbf{X} \\ \mathbf{X}^T & \mathbf{0} \end{bmatrix}$, one has $\mathbf{M} = \mathbf{c}_2 . \mathbf{p}$

$\Rightarrow \mathbf{M}^T\mathbf{Y} = \mathbf{p}^T\mathbf{c}_2^T\mathbf{Y} = \mathbf{p}^T\hat{\beta}$; giving us the best estimator $\hat{\beta} = \mathbf{c}_2^T\mathbf{Y}$. It's covariance is then obtained from $\mathbf{cov}\left[\hat{\beta}\right] = \sigma^2\mathbf{c}_4$ since $\mathbf{cov}\left[\mathbf{p}^T\hat{\beta}\right] = \sigma^2\mathbf{p}^T\mathbf{c}_4\mathbf{p}$ Note that it is necessary for the inverse of:

$$\begin{bmatrix} \mathbf{G} & \mathbf{X} \\ \mathbf{X}^T & \mathbf{0} \end{bmatrix}$$

to be calculable for this method to proceed—in other words, specifically with regard to the mass comparison problem, the method outlined above cannot be simply used with the design matrix **X** and data **Y**; since, as we know, this is insoluble by itself—there is just not enough information present to permit a unique solution, irrespective of what estimation method is used. Observe that the specific evaluation of $\mathbf{c}_2$ & $\mathbf{c}_4$ depends upon the physical nature of the problem to be solved. In this chapter we have not shown how to actually calculate $\mathbf{c}_2$ & $\mathbf{c}_4$. So let us now proceed to investigate how we can utilize this method in the mass calibration parameter estimation process.

****

7. GGM Theory in the Mass Model

7.0 Summary

The GGM theory introduced in Chapter 6 is now applied to specific examples with application to mass calibration. There are two cases to be considered, depending upon whether or not the constraints are viewed stochastically. In both cases we use an *extended model*—that is prior information must be included in the analysis in order to obtain a solution.

After some mathematical manipulation, we highlight how GGM generates solutions identical to RLS if the prior information is viewed as constant, while the solution is the same as that due to AD if stochastic constraints are used (Bich, 1992). This is a significant result and from it we can conclude that both RLS and AD are special cases of a general theory, albeit GGM is derived from an entirely different theoretical starting point. Thus RLS is appropriate to use in cases where constant constraints apply, but *this is not so in mass calibration* and thus, contrary to common practice in metrology, we conclude that the RLS method should not be used for parameter estimation in mass determination. Fig. (7.0.1) below illustrates the relationship between the models.

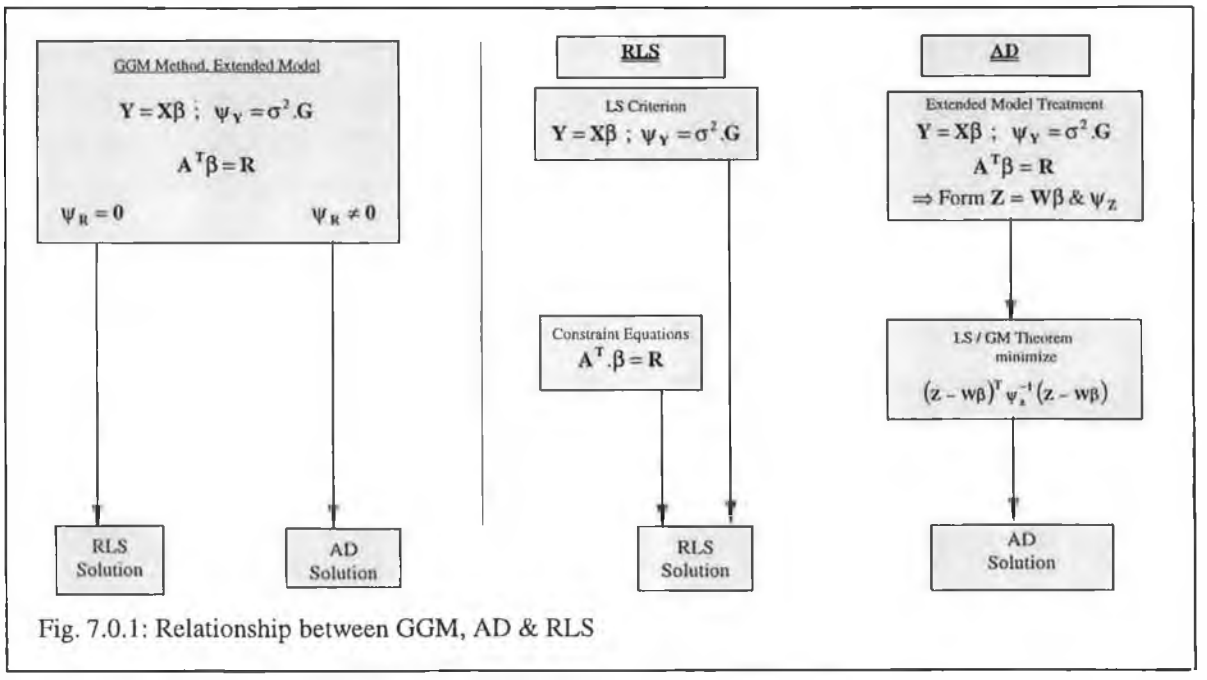


Fig. 7.0.1: Relationship between GGM, AD & RLS

7.1 Introduction

Having developed the GGM theory we now have two cases to consider, which arise from our investigations in Chapter 5 with regard to mass calibration: either the necessary prior information we must supply to obtain a solution is "uncertain" or it is deterministic. In the first case we have:

$$\mathbf{W}=\begin{bmatrix}\mathbf{X}\\ \mathbf{A}^{T}\end{bmatrix};\ \mathbf{Z}=\begin{bmatrix}\mathbf{Y}\\ \mathbf{R}\end{bmatrix};\ \&\ \psi_{Z}=\begin{bmatrix}\psi_{Y} & \mathbf{0}\\ \mathbf{0} & \psi_{R}\end{bmatrix} \tag{7.1.1}$$

and we can use the extended model of $\mathbf{W}$, $\mathbf{Z}$ & ψ_Z in the GGM model. We will see below that this requires ψ_Z to be invertable—which it is. In this case GGM should be the same as the AD solution, since consistency criteria dictate that different methods of analysing the same problem should yield the same solution, which is shown in this chapter to be the case for these two models.

On the other hand, if the constraints are considered deterministic, one now has $\mathbf{W}$ & $\mathbf{Z}$ as above, but:

$$\psi_{Z}=\begin{bmatrix}\psi_{Y} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix} \tag{7.1.2}$$

which is of course singular; the GM theorem does not hold now so the AD method cannot be implemented, but GGM is not incapacitated by this and we will show that it generates a solution identical with RLS. This should not surprise us since it shows that the two methods, AD & RLS are but particular cases of a general theory, depending on how one views the constraints. We have pointed out earlier that the deterministic constraint approach is inconsistent both with the known nature of this prior information and also with our general philosophy of Uncertainty Analysis.

7.2 Deterministic Constraints

Let us first examine the GGM model with fixed, non-stochastic constraints, where we have:

$$\mathbf{Y}=\mathbf{X}\beta\ \&\ \psi_{Y}=\mathbf{G} \tag{7.2.1a}$$

{for convenience we assume $\sigma^2=1$!}. The restraints are:

$$\mathbf{R}=\mathbf{A}^{T}\beta\ \&\ \psi_{R}=\mathbf{0} \tag{7.2.1b}$$

Thus the model we use is that given by (7.1.1) but with ψ_Z given by (7.1.2). According to the GGM theory, what we need to do is evaluate the g-inverse:

$$\begin{bmatrix}\psi_{Z} & \mathbf{W}\\ \mathbf{W}^{T} & \mathbf{0}\end{bmatrix}^{-}=\begin{bmatrix}c_{1} & c_{2}\\ c_{3} & -c_{4}\end{bmatrix}=\mathbf{H}^{-} \tag{7.2.2}$$

then $\hat{\beta} = \mathbf{c}_2^T\mathbf{Z}$ & $\psi_{\hat{\beta}} = \mathbf{c}_4$. In our case the augmented matrix **H** is given by:

$$\left[\begin{array}{cc|c} \psi_Y & \mathbf{0} & \mathbf{X} \\ \mathbf{0} & \mathbf{0} & \mathbf{A}^T \\ \hline \mathbf{X}^T & \mathbf{A} & \mathbf{0} \end{array}\right]$$

however, to obtain the inverse of H using the rule for inverting a partitioned matrix (see, e.g. Beck & Arnold, 1977), we must re-partition it first so that we do not need to invert any singular sub-matrices. Thus:

$$\mathbf{H}^{-} = \left[\begin{array}{c|cc} \psi_Y & \mathbf{0} & \mathbf{X} \\ \hline \mathbf{0} & \mathbf{0} & \mathbf{A}^T \\ \mathbf{X}^T & \mathbf{A} & \mathbf{0} \end{array}\right]^{-1} = \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}^T & \mathbf{C} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{A}^{-1} + \mathbf{F}\mathbf{E}^{-1}\mathbf{F}^T & -\mathbf{F}\mathbf{E}^{-1} \\ -\mathbf{E}^{-1}\mathbf{F}^T & \mathbf{E}^{-1} \end{bmatrix} \quad (7.2.3)$$

where $\mathbf{E} = \mathbf{C} - \mathbf{B}^T\mathbf{F}$ & $\mathbf{F} = \mathbf{A}^{-1}\mathbf{B}$ and $\mathbf{A}^{-1} = \Psi_Y^{-1} = \mathbf{G}^{-1}$. Also:

$$\mathbf{B} = [\mathbf{0} \quad \mathbf{X}];\ \mathbf{C} = \begin{bmatrix} \mathbf{0} & \mathbf{A}^T \\ \mathbf{A} & \mathbf{0} \end{bmatrix};\ \mathbf{F} = \psi_Y^{-1}[\mathbf{0} \quad \mathbf{X}] = \begin{bmatrix} \mathbf{0} & \psi_Y^{-1}\mathbf{X} \end{bmatrix}$$

$$\mathbf{E} = \mathbf{C} - \mathbf{B}^T\mathbf{A}^{-1}\mathbf{B} = \begin{bmatrix} \mathbf{0} & \mathbf{A}^T \\ \mathbf{A} & -\mathbf{X}^T\mathbf{G}^{-1}\mathbf{X} \end{bmatrix} \quad (7.2.4)$$

from the definitions of **A**,**B** & **C**. So we require $\mathbf{E}^{-1}$ to evaluate (7.2.3). If we transform **E** into **D** given by :

$$\mathbf{D} = \begin{bmatrix} -\mathbf{X}^T\mathbf{G}^{-1}\mathbf{X} & \mathbf{A} \\ \mathbf{A}^T & \mathbf{0} \end{bmatrix} \quad (7.2.5)$$

we can proceed in a similar manner to the development of the RLS or GGM theory. The transformation can be effected by:

$$\mathbf{D} = \mathbf{U}^{-1}\mathbf{E}\mathbf{U} \text{ where } \mathbf{U} = \begin{bmatrix} \mathbf{0} & \mathbf{I} \\ \mathbf{I} & \mathbf{0} \end{bmatrix} \quad (7.2.6)$$

We may now proceed with **D** as follows: Let:

$$\mathbf{D}^{-1} = \begin{bmatrix} \mathbf{d}_1 & \mathbf{d}_2 \\ \mathbf{d}_2^T & \mathbf{d}_3 \end{bmatrix} \text{ so that } \mathbf{D}\mathbf{D}^{-1} = \mathbf{I} \quad (7.2.7)$$

$$\text{i.e. } -\mathbf{X}^T\mathbf{G}^{-1}\mathbf{X}\mathbf{d}_1 + \mathbf{A}\mathbf{d}_2^T = \mathbf{I} \quad (7.2.8a)$$

$$-\mathbf{X}^T\mathbf{G}^{-1}\mathbf{X}\mathbf{d}_2 + \mathbf{A}\mathbf{d}_3 = \mathbf{0} \quad (7.2.8b)$$

$$\mathbf{A}^T\mathbf{d}_1 = \mathbf{0} \quad (7.2.8c)$$

$$\mathbf{A}^T\mathbf{d}_2 = \mathbf{I} \quad (7.2.8d)$$

Now because $\mathbf{X}^T\mathbf{G}^{-1}\mathbf{X}$ is singular we will run into problems in evaluating $\mathbf{d}_1$ from (7.2.8a) & $\mathbf{d}_2$ from (7.2.8b), So we define:

$$\mathbf{X}^T\mathbf{G}^{-1}\mathbf{X} = \mathbf{a}_0 + \mathbf{D}\mathbf{A}\mathbf{A}^T \quad (7.2.9)$$

for **D** a non-singular diagonal matrix. We can let **D** = **-I**, so that for (7.2.8a) we get:

$$-\left(\mathbf{a}_0 - \mathbf{A}\mathbf{A}^T\right)\mathbf{d}_1 = \mathbf{I} - \mathbf{A}\mathbf{d}_2^T$$

$$\Rightarrow -\mathbf{a}_0\mathbf{d}_1 + \mathbf{A}\mathbf{A}^T\mathbf{d}_1 = \mathbf{I} - \mathbf{A}\mathbf{d}_2^T \;;\; \text{but } \mathbf{A}^T\mathbf{d}_1 = \mathbf{0} \text{ by (7.2.8c)}$$

$$\therefore \mathbf{d}_1 = \mathbf{a}_0^{-1}\left[\mathbf{A}\mathbf{d}_2^T - \mathbf{I}\right] \qquad (7.2.10a)$$

while for (7.2.8b):

$$-\left(\mathbf{a}_0 - \mathbf{A}\mathbf{A}^T\right).\mathbf{d}_2 = -\mathbf{A}\mathbf{d}_3$$

$$= -\mathbf{a}_0\mathbf{d}_2 + \mathbf{A}\mathbf{A}^T\mathbf{d}_2 = -\mathbf{A}\mathbf{d}_3 \;;\; \text{but } \mathbf{A}^T\mathbf{d}_2 = \mathbf{I} \text{ from (7.2.8d)}$$

$$\text{so } -\mathbf{a}_0\mathbf{d}_2 = -\mathbf{A}\mathbf{d}_3 - \mathbf{A}$$

$$\text{or } \mathbf{d}_2 = \mathbf{a}_0^{-1}\mathbf{A}\left(\mathbf{d}_3 + \mathbf{I}\right) \qquad (7.2.10b)$$

Also, pre-multiplying (7.2.10b) by $\mathbf{A}^T$ gives:

$$\mathbf{A}^T\mathbf{d}_2 = \mathbf{A}^T\mathbf{a}_0^{-1}\mathbf{A}\left(\mathbf{d}_3 + \mathbf{I}\right) = \mathbf{I} \text{ by (7.2.8d)}$$

$$\text{so } \mathbf{d}_3 = \left(\mathbf{A}^T\mathbf{a}_0^{-1}\mathbf{A}\right)^{-1} - \mathbf{I} \qquad (7.2.10c)$$

Thus the components of $\mathbf{D}^{-1}$ are evaluated in terms of $\mathbf{X}$, $\mathbf{G}^{-1}$ & $\mathbf{A}$. We must now transform back to get $\mathbf{E}^{-1}$ by means of $\mathbf{E}^{-1} = \mathbf{U}\mathbf{D}^{-1}\mathbf{U}^{-1}$ where $\mathbf{U}$ is as given in (7.2.6). Then $\mathbf{E}^{-1}$ becomes:

$$\mathbf{E}^{-1} = \begin{bmatrix} \mathbf{d}_3 & \mathbf{d}_2^T \\ \mathbf{d}_2 & \mathbf{d}_1 \end{bmatrix} \qquad (7.2.11)$$

with $\mathbf{d}_1$, $\mathbf{d}_2$ & $\mathbf{d}_3$ as given in (7.2.10a) to (7.2.10c). We can now evaluate the components of the matrix on the r.h.s. of (7.2.3):

$$\mathbf{A}^{-1} + \mathbf{F}\mathbf{E}^{-1}\mathbf{F}^T = \mathbf{G}^{-1} + \begin{bmatrix} \mathbf{0} & \mathbf{G}^{-1}\mathbf{X} \end{bmatrix}\begin{bmatrix} \mathbf{d}_3 & \mathbf{d}_2^T \\ \mathbf{d}_2 & \mathbf{d}_1 \end{bmatrix}\begin{bmatrix} \mathbf{0} \\ \mathbf{X}^T\mathbf{G}^{-1} \end{bmatrix}$$

$$= \mathbf{G}^{-1} + \mathbf{G}^{-1}\mathbf{X}\mathbf{d}_1\mathbf{X}^T\mathbf{G}^{-1}$$

$$= \mathbf{G}^{-1}\left(\mathbf{I} + \mathbf{X}\mathbf{d}_1\mathbf{X}^T\mathbf{G}^{-1}\right) \qquad (7.2.12a)$$

$$-\mathbf{F}\mathbf{E}^{-1} = \begin{bmatrix} \mathbf{0} & -\mathbf{G}^{-1}\mathbf{X} \end{bmatrix}\begin{bmatrix} \mathbf{d}_3 & \mathbf{d}_2^T \\ \mathbf{d}_2 & \mathbf{d}_1 \end{bmatrix} = \begin{bmatrix} -\mathbf{G}^{-1}\mathbf{X}\mathbf{d}_2 & -\mathbf{G}^{-1}\mathbf{X}\mathbf{d}_1 \end{bmatrix} \qquad (7.2.12b)$$

$$-\mathbf{E}^{-1}\mathbf{F}^T = \begin{bmatrix} \mathbf{d}_3 & \mathbf{d}_2^T \\ \mathbf{d}_2 & \mathbf{d}_1 \end{bmatrix}\begin{bmatrix} \mathbf{0} \\ \mathbf{X}^T.\mathbf{G}^{-1} \end{bmatrix} = \begin{bmatrix} -\mathbf{d}_2^T\mathbf{X}^T\mathbf{G}^{-1} \\ -\mathbf{d}_1\mathbf{X}^T\mathbf{G}^{-1} \end{bmatrix} \qquad (7.2.12c)$$

(Note that $\mathbf{d}_1$ is in fact symmetric). Thus from (7.2.3), the inverse is as follows:

$$\left[\begin{array}{c|cc} \mathbf{G}^{-1} + \mathbf{G}^{-1}\mathbf{X}\mathbf{d}_1\mathbf{X}^T\mathbf{G}^{-1} & -\mathbf{G}^{-1}\mathbf{X}\mathbf{d}_2 & -\mathbf{G}^{-1}\mathbf{X}\mathbf{d}_1 \\ \hline -\mathbf{d}_2^T\mathbf{X}^T\mathbf{G}^{-1} & \mathbf{d}_3 & \mathbf{d}_2^T \\ -\mathbf{d}_1\mathbf{X}^T\mathbf{G}^{-1} & \mathbf{d}_2 & \mathbf{d}_1 \end{array}\right] \qquad (7.2.13a)$$

Now, noting that we re-partitioned in Eq. (7.2.3), from Eq. (7.2.2), we can see that the terms $\mathbf{c}_1$, $\mathbf{c}_2$, $\mathbf{c}_3$ & $-\mathbf{c}_4$ will result from partitioning (7.2.13a) as follows:

$$\mathbf{H}^{-} = \left[\begin{array}{cc|c} \mathbf{G}^{-1}+\mathbf{G}^{-1}\mathbf{X}\mathbf{d}_1\mathbf{X}^{\mathrm{T}}\mathbf{G}^{-1} & -\mathbf{G}^{-1}\mathbf{X}\mathbf{d}_2 & -\mathbf{G}^{-1}\mathbf{X}\mathbf{d}_1 \\ -\mathbf{d}_2^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{G}^{-1} & \mathbf{d}_3 & \mathbf{d}_2^{\mathrm{T}} \\ \hline -\mathbf{d}_1\mathbf{X}^{\mathrm{T}}\mathbf{G}^{-1} & \mathbf{d}_2 & \mathbf{d}_1 \end{array}\right] \qquad (7.2.13b)$$

We are interested, in the GGM theory in just $\mathbf{c}_2$ & $-\mathbf{c}_4$ so in fact we must now deal with

$$\mathbf{c}_2 = \begin{bmatrix} -\mathbf{G}^{-1}\mathbf{X}\mathbf{d}_1 \\ \mathbf{d}_2^{\mathrm{T}} \end{bmatrix} \quad \& \quad \mathbf{c}_4 = -\mathbf{d}_1$$

Our estimators are now:

$$\hat{\beta} = \mathbf{c}_2^{\mathrm{T}}\mathbf{Z} = \begin{bmatrix} -\mathbf{d}_1\mathbf{X}^{\mathrm{T}}\mathbf{G}^{-1} & \mathbf{d}_2 \end{bmatrix} \begin{bmatrix} \mathbf{Y} \\ \mathbf{R} \end{bmatrix}$$

$$\Rightarrow \boxed{\hat{\beta} = -\mathbf{d}_1\mathbf{X}^{\mathrm{T}}\mathbf{G}^{-1}\mathbf{Y} + \mathbf{d}_2\mathbf{R}} \qquad (7.2.14)$$

Now with the aid of (7.2.10a) to (7.2.10c) we find (7.2.14) can be expressed as:

$$\boxed{\hat{\beta} = \mathbf{a}_0^{-1}\left(\mathbf{X}^{\mathrm{T}}\mathbf{G}^{-1}\mathbf{Y} + \mathbf{A}\left(\mathbf{A}^{\mathrm{T}}\mathbf{a}_0^{-1}\mathbf{A}\right)^{-1}\left\{\mathbf{R} - \mathbf{A}^{\mathrm{T}}\mathbf{a}_0^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{G}^{-1}\mathbf{Y}\right\}\right)} \qquad (7.2.15)$$

—which is identical with Eq. (5.3.23) obtaining for the restrained least squares analysis! This is a most interesting convergence since the two methods are based on different principles and establishes for us that the GGM is a general theory which is equivalent to RLS under the circumstances of a model $\mathbf{Y} = \mathbf{X}\beta$ subject to deterministic constraints $\mathbf{A}^{\mathrm{T}}\beta = \mathbf{R}$. Note that we have taken care that $\mathbf{a}_0 = \mathbf{X}^{\mathrm{T}}\mathbf{G}^{-1}\mathbf{X} + \mathbf{A}\mathbf{A}^{\mathrm{T}}$ is defined like this for both expressions so that we can directly compare them.

For the covariance matrix in the GGM model we have $\psi_{\hat{\beta}} \equiv -\mathbf{d}_1$ from (7.2.13b), and with (7.2.10a) to (7.2.10c) this yields:

$$\psi_{\hat{\beta}} = \mathbf{a}_0^{-1}\left(\mathbf{I} - \mathbf{A}\left(\mathbf{A}^{\mathrm{T}}\mathbf{a}_0^{-1}\mathbf{A}\right)^{-1}\mathbf{A}^{\mathrm{T}}\mathbf{a}_0^{-1}\right) \qquad (7.2.16)$$

which is also identical with the RLS estimator of covariance.

Note: the reason for the somewhat protracted calculations above is so that (7.2.15) & (7.2.16) are essentially expressed in terms of **X**, **Y**, **R**, **A** & **G**; all of which are known at the start of the work. However, it is not computationally difficult to form the augmented matrices required by GGM. So for the G-M model with restraints one would require to find the g-inverse:

$$\begin{bmatrix} \psi_Z & \mathbf{W} \\ \mathbf{W}^{\mathrm{T}} & \mathbf{0} \end{bmatrix}^{-} \equiv \left[\begin{array}{cc|c} \psi_Y & \mathbf{0} & \mathbf{X} \\ \mathbf{0} & \mathbf{0} & \mathbf{A}^{\mathrm{T}} \\ \hline \mathbf{X}^{\mathrm{T}} & \mathbf{A} & \mathbf{0} \end{array}\right]^{-} = \begin{bmatrix} \mathbf{c}_1 & \mathbf{c}_2 \\ \mathbf{c}_2^{\mathrm{T}} & \mathbf{c}_3 \end{bmatrix} \qquad (7.2.17)$$

Thus as usual

$$\psi_Z\mathbf{c}_1 + \mathbf{W}\mathbf{c}_2^{\mathrm{T}} = \mathbf{I}$$

$$\psi_Z\mathbf{c}_2 + \mathbf{W}\mathbf{c}_3 = \mathbf{0}$$

$$\mathbf{W}^T\mathbf{c}_1 = \mathbf{0}$$
$$\mathbf{W}^T\mathbf{c}_2 = \mathbf{I}$$

From these we get:

$$\mathbf{c}_1 = \mathbf{a}_0^{-1}\left(\mathbf{I} - \mathbf{W}\mathbf{c}_2^T\right) \tag{7.2.18a}$$
$$\mathbf{c}_2 = \mathbf{a}_0^{-1}\mathbf{W}\left(\mathbf{I} - \mathbf{c}_3\right) \tag{7.2.18b}$$
$$\mathbf{c}_3 = \mathbf{I} - \left(\mathbf{W}^T\mathbf{a}_0^{-1}\mathbf{W}\right)^{-1} \tag{7.2.18c}$$
$$\mathbf{a}_0 = \psi_Z + \mathbf{W}\mathbf{W}^T \tag{7.2.18d}$$

In fact $\mathbf{c}_1$ is redundant since we require:

$$\hat{\beta} = \mathbf{c}_2^T\mathbf{Z}$$
$$\& \ \psi_{\hat{\beta}} = -\mathbf{c}_3$$

The augmented matrices ψ_Z & $\mathbf{W}$ can easily be formed and the estimators then found without difficulty.

7.3 Stochastic Constraints:

On the other hand, if the constraints are not deterministic, we have $\mathbf{A}^T\beta = \mathbf{R}$ & $\psi_R \neq \mathbf{0}$ and we can use the expanded model of (7.1.1) in the GGM analysis (i.e. $\mathbf{W}$, ψ_z & $\mathbf{Z}$). Therefore, by Result I from Sec. 6.3, $\mathbf{X}\mathbf{c}_2^T\mathbf{X} = \mathbf{X}$, or in this case, $\mathbf{W}\mathbf{c}_2^T\mathbf{W} = \mathbf{W}$, since we are dealing with the augmented matrix. Now by Result III (Sec. 6.3):

$$\psi_Z\mathbf{c}_1\mathbf{W} = \mathbf{0}$$
$$\therefore \psi_Z\mathbf{c}_1\mathbf{W} + \mathbf{W}\mathbf{c}_2^T\mathbf{W} = \mathbf{W}$$
$$\text{or, } \psi_Z\mathbf{c}_1 + \mathbf{W}\mathbf{c}_2^T = \mathbf{I} \tag{7.3.1a}$$

We also know from Result III that $\mathbf{W}^T\mathbf{c}_1\mathbf{W} = \mathbf{0}$

$$\Rightarrow \mathbf{W}^T\mathbf{c}_1 = \mathbf{0} \tag{7.3.1b}$$

for a non-trivial solution.

Result II tells us that $\mathbf{W}\mathbf{c}_4\mathbf{W}^T = \psi_Z\mathbf{c}_2\mathbf{W}^T$

$$\Rightarrow \psi_Z\mathbf{c}_2 - \mathbf{W}\mathbf{c}_4 = \mathbf{0} \tag{7.3.1c}$$

By Result I again, $\mathbf{W}\mathbf{c}_2^T\mathbf{W} = \mathbf{W}$

$$\Rightarrow \mathbf{W}^T\mathbf{c}_2\mathbf{W}^T = \mathbf{W}^T$$
$$\Rightarrow \mathbf{W}^T\mathbf{c}_2 = \mathbf{I} \tag{7.3.1d}$$

From (7.3.1a) we see that:

$$\mathbf{c}_1 = \psi_Z^{-1}\left(\mathbf{I} - \mathbf{W}\mathbf{c}_2^T\right) \tag{7.3.2}$$

(Observation: Here we see how we can operate when ψ_Z is non-singular. If we had $\psi_R = \mathbf{0}$, we would have $|\psi_Z| = 0$ and (7.3.2) could not be obtained. Thus this development is based on non-deterministic constraints. Note also that Eq. (7.3.2) above and indeed (7.3.4) below can be derived by an identical analysis to that giving

Eqs (7.2.17) - (7.2.18), except that $\mathbf{a_0}$ would not be needed since ψ_z is now directly invertible.)

Continuing, from (7.3.2):

$$\mathbf{W}^T\mathbf{c}_1 = \mathbf{W}^T\psi_Z^{-1}\left(\mathbf{I} - \mathbf{W}\mathbf{c}_2^T\right) \tag{7.3.3}$$

but $\mathbf{W}^T\mathbf{c}_1 = \mathbf{0}$ by (7.3.1b) so:

$$\mathbf{W}^T\psi_Z^{-1} = \mathbf{W}^T\psi_Z^{-1}\mathbf{W}\mathbf{c}_2^T$$

$$\text{or } \mathbf{c}_2^T = \left(\mathbf{W}^T\psi_Z^{-1}\mathbf{W}\right)^{-1}\mathbf{W}^T\psi_Z^{-1} \tag{7.3.4}$$

But in the GGM method, $\hat{\beta} = \mathbf{c}_2^T\mathbf{Z}$ so we can see that the estimator is:

$$\boxed{\hat{\beta} = \left(\mathbf{W}^T\psi_Z^{-1}\mathbf{W}\right)^{-1}\mathbf{W}^T\psi_Z^{-1}\mathbf{Z}} \tag{7.3.5}$$

—which is identical to the G-M based AD approach ! (Eq. 5.5.5), where $\mathbf{Z}$ is the augmented vector of input data, i.e.:

$$\mathbf{Z} = \begin{bmatrix}\mathbf{Y}\\ \mathbf{R}\end{bmatrix}$$

From Eq. (7.3.1c),

$$\psi_Z\mathbf{c}_2 = \mathbf{W}\mathbf{c}_4$$

$$\Rightarrow \mathbf{c}_2 = \psi_Z^{-1}\mathbf{W}\mathbf{c}_4$$

$$\mathbf{c}_4 = \left(\psi_Z^{-1}\mathbf{W}\right)^{-1}\mathbf{c}_2$$

which becomes, using (7.3.4):

$$\mathbf{c}_4 = \left(\psi_Z^{-1}\mathbf{W}\right)^{-1}\left(\psi_Z^{-1}\mathbf{W}\right)\left(\mathbf{W}^T\psi_Z^{-1}\mathbf{W}\right)^{-1}$$

$$\text{or, } \mathbf{c}_4 = \left(\mathbf{W}^T\psi_Z^{-1}\mathbf{W}\right)^{-1} \tag{7.3.6}$$

and since $\psi_{\hat{\beta}} = \mathbf{c}_4$, we now have a covariance estimator also identical with that produced by the Augmented Design approach.

In conclusion, this is an important chapter as it ties together a lot of mathematical development, starting with Chapter 5, and highlights the two primary approaches to parameter estimation. We have seen that AD & RLS are both particular cases of a generalised estimation technique operating on a linear model subject to restraints/prior information. The different formalisms result from different interpretations of the nature of this prior information. As we have pointed out several times before, our criteria of logical consistency and a desire for a unified approach which takes adequate account of everything we know about the problem lead us to consider the extended model AD solution, or the extended GGM model with stochastic restraints, as a better interpretation of the available information. Our next chapter is the final one on estimation techniques, and introduces the logic of Bayesian Analysis as an even better diagnostic tool for analysing group comparisons in mass calibrations.

8. Maximum Likelihood Estimation

8.0 Summary

In this final chapter on parameter estimation methods we adopt the explicitly Bayesian approach of Maximum Likelihood Estimation (MLE) & Maximum a Posteriori estimation (MAP). This approach is analogous to that used in uncertainty estimation as discussed in the first four chapters, insofar as one of the key features is the selection of a *distribution* to describe the observations (the corrected experimental data), and also the parameters, and in this respect is a convenient unification of the preceeding analysis. The true *conditional* nature of probability is pointed out and the basic rules of probability theory are used to generate Bayes's theorem. With the Maximum Likelihood Criteria a *posterior distribution* for the parameters can be established, given the particular observations that were obtained and the available prior information. In this way the posterior distribution of parameters is the prior distribution updated by the new current information obtained during the experiment. Thus *all* known data is included in a unified manner, a desirable feature in mass determination, since we have pointed out that the difference between data obtained in a previous calibration and that obtained in the current one is primarily one of logical relationships insofar as inference and estimation are concerned. We must ensure of course that the full extent of prior knowledge is included which may mean considering the effect of drift on the prior information. This will be explored in more detail in Chapter 10.

It will be shown in this chapter that the MAP estimator generates the same parameter estimate as does the AD method if a Normal Distribution can be used and the prior information is just the known values of some or all of the standards. However, if the prior information comprised a combination of two or more standards (e.g. a sum or difference term) the AD method would provide an adequate estimate while the MAP method would not be possible as the prior vectors / covariance matrices could not be constructed.

The MAP estimator also generates a reduced covariance matrix and does not depend for a solution on the form of the design matrix, as does the LS method. Note in Fig. (8.0.1) below that the system model describing the relationship between parameters and observations is not used in the estimation process, but rather, the distribution function of the data and the prior information is used. Note also in Fig. (8.0.2) that the process can be used in a sequential manner to continually update the parameters as more information is obtained. This latter aspect leads us to expect that

information on the evolution of the mass standards over time can be easily explored with this analytical method.

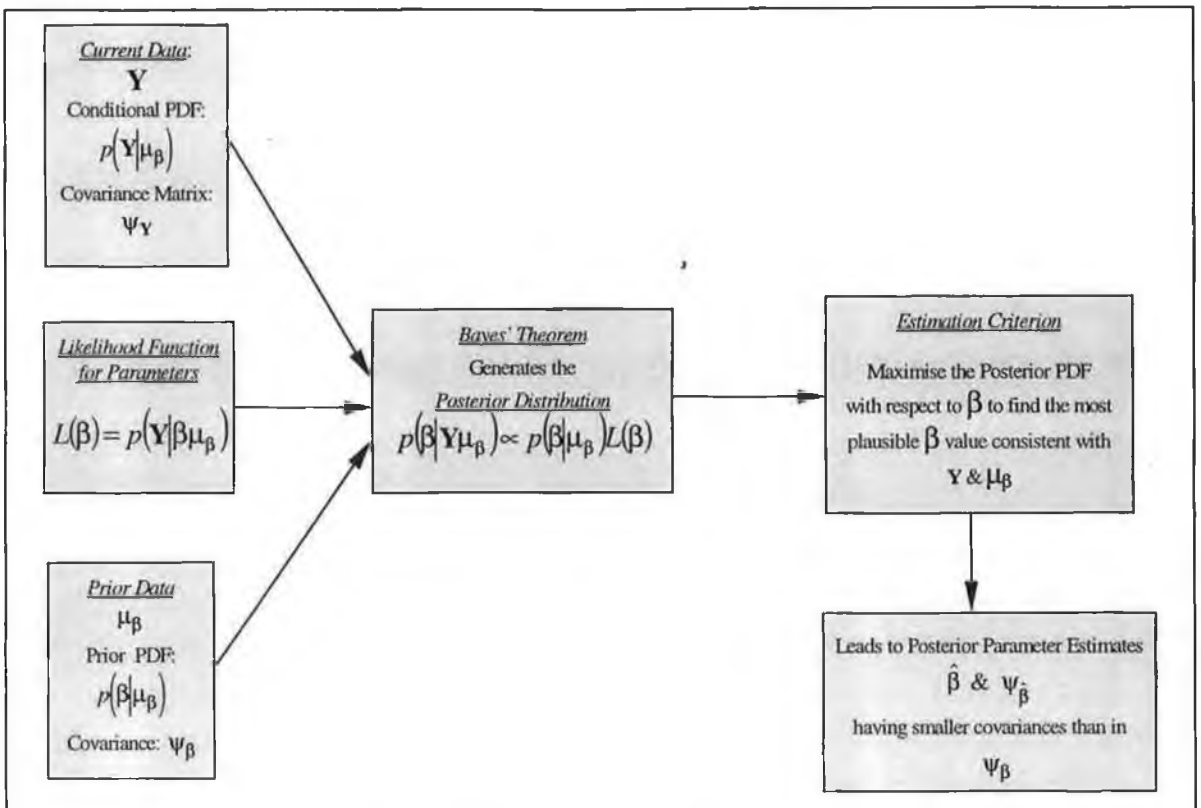


Fig. (8.0.1): Essential Aspects of the Bayesian Estimation method.

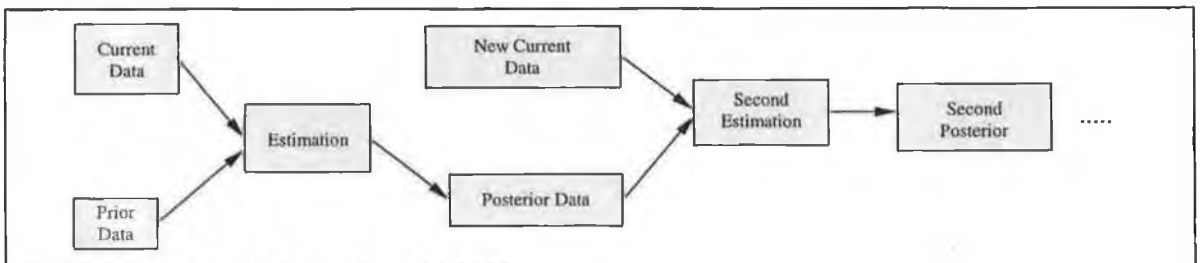


Fig. (8.0.2): Sequential Estimation with MAP

8.1 Introduction

We have so far considered parameter estimation techniques based on Least Squares, the Gauss-Markov theorem, and a generalised estimation technique not relying on either of the first two methods but encompassing them both in its scope. We have seen how these methods can give different results, the essential difference being how they treat the prior information necessary to get a particular solution for the

comparison calibration data. We now wish to look at one more method, based on the completely different approach of Bayesian probability, which we will see, is a very appropriate way of dealing with the information we have. First we must examine the basis of the method.

8.2 Maximum Likelihood (See Beck & Arnold 1977)

The key point here is that, unlike Least Squares techniques which involve a minimisation of the vector norm $\left\|\mathbf{Y}-\mathbf{X}\hat{\beta}\right\|$, or the GGM method which involves finding a best estimator irrespective of any judgements about β, $\mathbf{Y}$ or $\psi_{\mathbf{Y}}$; the methods now being presented are based upon an analysis of Distribution Functions. Thus, information must be available on the type of distribution which best describes the dispersion characteristics of the parameter under scrutiny. Essentially one requires, that for a model

$$\mathbf{Y} = f(\beta) \tag{8.2.1}$$

one can choose from among the possible values for β, the set which maximises the probability of obtaining the set of data, $\mathbf{Y}$, which was in fact observed. Thus one is concerned only with the data set $\mathbf{Y}$ which is known to exist, and not with the wider population of $\mathbf{Y}$ values of which our vector might be a sample—i.e. the space of all data sets which might have been observed, but in fact were not! This policy is in accord with the Consistency Criteria of Chapter1.

To do this one requires the conditional distribution $f(x|\theta)$ which is the joint distribution function for the x values which could be observed for a particular θ value, or a particular distribution of θ values. It is important to realise that no probabilities are absolute: there is always a conditional dependence on some existing or background information. (See discussion in Chapter 1). If β is given then $f(\mathbf{Y}|\beta)$ is a *sampling* distribution which describes the dependence of $\mathbf{Y}$ on fixed β. But it can also be considered for the case of a *fixed data set* $\mathbf{Y}$, in the light of possible values for β. In this context, $f(\mathbf{Y}|\beta)$ is termed the Likelihood Function for β, denoted by $L(\beta)$. This likelihood function can be maximised to give the most plausible β values for the $\mathbf{Y}$ data which was obtained.

Before implementing such a method it is necessary to define a suitable distribution. In our case, the model for $\mathbf{Y}$ is just $\mathbf{Y} = \mathbf{X}\beta$. Here $\mathbf{X}$ is a constant while a probability distribution describes the dispersion characteristics of $\mathbf{Y}$. Following the considerations of Chapter 1 we can assume a Normal Distribution since we will have a mean value and covariance matrix for $\mathbf{Y}$. Thus:

$$\mathbf{cov}[\mathbf{Y}] = \psi_{\mathbf{Y}}, \text{ which is fully known}$$
$$\& \; E[\mathbf{Y}] = \mathbf{X}\beta \tag{8.2.2}$$

We require to find estimators $\hat{\beta}$ such that the Likelihood function is maximised. For a variable x of mean value μ and variance σ^2, the Gaussian distribution may be written as:

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}\exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\} \tag{8.2.3}$$

If $\mathbf{Y}$ is an n x 1 vector, then the distribution function for $\mathbf{Y}$, given by the conditions in (8.2.2) will be:

$$f(\mathbf{Y}|\beta)=(2\pi)^{-\frac{n}{2}}\psi_Y^{-\frac{1}{2}}\exp\left(-\frac{(\mathbf{Y}-\mathbf{X}\beta)^T\psi_Y^{-1}(\mathbf{Y}-\mathbf{X}\beta)}{2}\right) \tag{8.2.4}$$

Before the experiment is carried out, $f(\mathbf{Y}|\beta)$ associated a probability density with each outcome $\mathbf{Y}$, for a fixed parameter vector β. After the data is obtained we need to find the particular β which would maximise the probability density function for the $\mathbf{Y}$ we did get. Hence the Likelihood Function $L(\beta|\mathbf{Y})$ is to be maximised. This will also have the form of $f(\mathbf{Y}|\beta)$ as in (8.2.4) but now β is considered variable and $\mathbf{Y}$ is fixed. Taking the log of (8.2.4) yields:

$$\ln\{L(\beta|\mathbf{Y})\}=-\frac{1}{2}\left[n\ln(2\pi)+\ln|\psi_Y|+S_{ML}\right] \tag{8.2.5a}$$

$$\text{where } S_{ML}=(\mathbf{Y}-\mathbf{X}\beta)^T\psi_Y^{-1}(\mathbf{Y}-\mathbf{X}\beta) \tag{8.2.5b}$$

Maximising this Likelihood function can be achieved by minimising S_{ML} as given in (8.2.5b), as this is the only term which has a β dependence. Hence we require:

$$\nabla_\beta\left\{\mathbf{Y}^T\psi_Y^{-1}\mathbf{Y}-\mathbf{Y}^T\psi_Y^{-1}\mathbf{X}\beta-\beta^T\mathbf{X}^T\psi_Y^{-1}\mathbf{Y}+\beta^T\mathbf{X}^T\psi_Y^{-1}\mathbf{X}\beta\right\}\Big|_{\beta=\hat{\beta}}=0 \tag{8.2.6}$$

The value of $\beta=\hat{\beta}$ which satisfies (8.2.6) will be the Maximum Likelihood Estimator we are seeking. Noting that:

$$\nabla_\beta(\mathbf{AB})=\nabla_\beta(\mathbf{A})\mathbf{B}+\nabla_\beta\left(\mathbf{B}^T\right)\mathbf{A}^T$$

$$\&\ \nabla_\beta\left(\mathbf{A}^T\phi\mathbf{A}\right)=2\left(\nabla_\beta\mathbf{A}^T\phi\mathbf{A}\right);\ \text{while}\ \nabla_\beta\left(\mathbf{B}^T\right)=\mathbf{I}$$

we can evaluate (8.2.6) to get:

$$-2\mathbf{X}^T\psi_Y^{-1}\mathbf{Y}+2\mathbf{X}^T\psi_Y^{-1}\mathbf{X}\hat{\beta}=\mathbf{0} \tag{8.2.7}$$

$$\text{and thus}\quad \boxed{\hat{\beta}_{ML}=\left(\mathbf{X}^T\psi_Y^{-1}\mathbf{X}\right)^{-1}\mathbf{X}^T\psi_Y^{-1}\mathbf{Y}} \tag{8.2.8}$$

Noting that $E[\mathbf{Y}]=\mathbf{X}\beta$ we can see easily that $E\left[\hat{\beta}_{ML}\right]=\beta$ and thus we have an unbiased estimator.

From (8.2.8) $\hat{\beta}_{ML}=\mathbf{L}.\mathbf{Y}$ with $\mathbf{L}=\left(\mathbf{X}^T\psi_Y^{-1}\mathbf{X}\right)^{-1}\mathbf{X}^T\psi_Y^{-1}$. Thus:

$$\psi_{\hat{\beta}_{ML}}=\mathbf{L}\psi_Y\mathbf{L}^T$$

$$\Rightarrow\ \psi_{\hat{\beta}_{ML}}=\left(\mathbf{X}^T\psi_Y^{-1}\mathbf{X}\right)^{-1} \tag{8.2.9}$$

So we can see that, providing the strict assumption holds that **Y** can be characterised by a Normal Distribution, (and from Maximum Entropy considerations we can be confident in this assumption with the information we have), the ML technique generates an estimator the same as that given by WLS, or the AD method. Hence by the GM theorem we can be assured that this is a minimum variance estimator.

However, while in this section we have pointed out the conditional nature of probabilities, and used this information to construct a likelihood function in order to establish a parameter estimate, we have not fully included all known information. Leaving the development like this will not do as it violates our requirement for a unified analysis and in any case would not be solvable since $\left(\mathbf{X}^T\psi_Y^{-1}\mathbf{X}\right)$ is singular in mass calibration problems. We now need to carry the Likelihood technique further to the case where we do know some prior information about the parameters β. We can incorporate this information with the aid of Bayes' theorem.....

8.3 Bayes' Theorem & Maximum a Posteriori Estimation

In this more complete analysis we want to explicitly identify any prior information that exists, and show how our probability functions depend on it. Again we highlight that there is no such thing as an absolute probability: all probabilities are conditional on some background information. Analogous to human experience, we do not discard all of yesterday's information and deal only with the immediately observable: rather we form a synthesis of the totality of our information. It is further necessary to be aware that this 'background' or 'prior' data is to be interpreted primarily as *logically* distinct from the current data. Chronological or causal relationships are *by no means* implied or required by the theory, albeit such may well exist in practice. "A-priori" probabilities, or data/information, are those which are known or available independently of the current experiment. (Jaynes (1996), Sivia (1996)).

In general terms, scientific inference involves the situation of a set of data, *D,* at hand along with various other prior information, the requirement being to associate probabilities with a set of hypotheses in the light of this information. The information may be "subjective" in the sense discussed in Chapter 1, i.e. that it is all that is available at the time but its dispersion characteristics will reflect the degrees of belief/plausibility which can reasonably be attributed to this information.

To proceed, we note the two fundamental rules of probability theory, (see, e.g. Jaynes (1996), Fröhner (1997), Bretthorst (1989) or Cox (1946)) the Product & Sum Rules:

$$\text{Product Rule :}\quad p(AB|C) = p(A|C)p(B|AC) = p(B|C)p(A|BC) \tag{8.3.1a}$$

which investigates the probability that two propositions A & B could both be true given that background information C is true.

$$\text{Sum Rule :}\quad p(A+B|C) = p(A|C) + p(B|C) - p(AB|C) \tag{8.3.1b}$$

which considers the probability of either (A or B) being true given the background information C. These two rules are derived as inescapable consequences of the basic requirements that probability theory be consistent with the fundamental desiderata of rational belief and Aristotelian Logic.

Let us now consider the situation where we have a vector μ_β of prior information, a vector $\mathbf{Y}$ of current data, while our hypothesis takes the form of a vector β of parameters we want to determine. So with the product rule of Eq. (8.3.1a) we have:

$$p\left(\mathbf{Y}\beta|\mu_\beta\right) = p\left(\mathbf{Y}|\mu_\beta\right)p\left(\beta|\mathbf{Y}\mu_\beta\right) = p\left(\beta|\mu_\beta\right)p\left(\mathbf{Y}|\beta\mu_\beta\right) \tag{8.3.2}$$

which considers the probability that data vector $\mathbf{Y}$ and some parameter vector β are both true given some prior information μ_β. We can rewrite this as:

$$p\left(\beta|\mathbf{Y}\mu_\beta\right) = p\left(\beta|\mu_\beta\right)\frac{p\left(\mathbf{Y}|\beta\mu_\beta\right)}{p\left(\mathbf{Y}|\mu_\beta\right)} \tag{8.3.3}$$

and in this formalism is commonly referred to as "Bayes' Theorem". The l.h.s is called a *posterior probability*, meaning that it is logically later in the inference process than the others. It gives us the probability that a particular β could occur given that both $\mathbf{Y}$ and μ_β are known with some specified degree of belief. On the r.h.s $p\left(\beta|\mu_\beta\right)$ depends only on the prior information μ_β and as such is termed a *prior probability*, and deals with the possibility of β existing in the light of only the prior information. The numerator in the final term on the r.h.s of Eq. (8.3.3) is called the Likelihood for β, $L(\beta)$, as in Section 8.2 above. This is not a probability but a term which when multiplied by a normalisation constant and a prior probability would become a probability term. The denominator in the r.h.s. term, being the distribution for Y and independent of β provides this normalisation constant. So, we can say that the posterior probability is proportional to the prior probability multiplied by the likelihood function.

We can now consider our model as follows:

$$\mathbf{Y} = \mathbf{X}\beta \tag{8.3.4a}$$

$$\text{with } p\left(\mathbf{Y}\middle|\mu_\beta\right) \Rightarrow E[\mathbf{Y}] = \mathbf{X}\beta, \text{var}[\mathbf{Y}] = \psi_\mathbf{Y} \tag{8.3.4b}$$

$$\text{while } p\left(\beta\middle|\mu_\beta\right) \Rightarrow E[\beta] = \mu_\beta, \text{var}[\beta] = \psi_\beta \tag{8.3.4c}$$

This is the crucial difference from preceeding chapters: the parameter vector β, is considered to have a prior-known expectation value μ_β and a dispersion matrix ψ_β. We further expect $\mathbf{cov}(\mathbf{Y},\beta) = \mathbf{0}$. We are thus deciding that $p\left(\beta\middle|\mu_\beta\right)$ is a normal distribution which we can write as

$$p\left(\beta\middle|\mu_\beta\right) = (2\pi)^{-p/2}\psi_\beta^{-1/2}\exp\left(-\frac{\left(\beta-\mu_\beta\right)^\mathrm{T}\psi_\beta^{-1}\left(\beta-\mu_\beta\right)}{2}\right) \tag{8.3.5}$$

where β is a p x 1 vector of parameters. Our Likelihood function for β is:

$$L(\beta) = p\left(\mathbf{Y}\middle|\beta\mu_\beta\right) \tag{8.3.6}$$

since $p\left(\mathbf{Y}\middle|\mu_\beta\right)$ is really a constant term describing the probability distribution of the data. Thus we have:

$$L(\beta) = (2\pi)^{-n/2}\psi_\mathbf{Y}^{-1/2}\exp\left(-\frac{(\mathbf{Y}-\mathbf{X}\beta)^\mathrm{T}\psi_\mathbf{Y}^{-1}(\mathbf{Y}-\mathbf{X}\beta)}{2}\right) \tag{8.3.7}$$

which is the conditional probability distribution for the data, $\mathbf{Y}$, from Eq. (8.3.4b). So we can write Bayes' Theorem as:

$$p\left(\beta\middle|\mathbf{Y}\mu_\beta\right) \propto p\left(\beta\middle|\mu_\beta\right)L(\beta) \tag{8.3.8}$$

We need to maximise the l.h.s. of Eq. (8.3.8) in order to find the parameter vector which is the most plausible in the context of the current information $\mathbf{Y}$ and the prior information μ_β. To do this we must then maximise the product on the r.h.s of Eq. (8.3.8), which from Eqs. (8.3.5) & (8.3.7) is :

$$(2\pi)^{-(n+p)/2}\left|\psi_\beta\right|^{-1/2}\left|\psi_\mathbf{Y}\right|^{-1/2}\exp\left[-\frac{(\mathbf{Y}-\mathbf{X}\beta)^\mathrm{T}\psi_\mathbf{Y}^{-1}(\mathbf{Y}-\mathbf{X}\beta)+\left(\beta-\mu_\beta\right)^\mathrm{T}\psi_\beta^{-1}\left(\beta-\mu_\beta\right)}{2}\right] \tag{8.3.9}$$

The problem is now reduced to finding the estimator $\hat{\beta}$ which satisfies this maximisation. The maximum of the r.h.s. of (8.3.9) will occur at the same point as the maximum of its natural log – i.e.:

$$-\frac{1}{2}\left[(n+p)\ln 2\pi + \ln\left|\psi_\beta\right| + \ln\left|\psi_\mathbf{Y}\right| + S_{MAP}\right] \tag{8.3.10a}$$

$$\text{where } S_{MAP} = (\mathbf{Y}-\mathbf{X}\beta)^\mathrm{T}\psi_\mathbf{Y}^{-1}(\mathbf{Y}-\mathbf{X}\beta) + \left(\beta-\mu_\beta\right)^\mathrm{T}\psi_\beta^{-1}\left(\beta-\mu_\beta\right) \tag{8.3.10b}$$

We want to find an estimator $\hat{\beta}$ that maximises (8.3.10a); This can be done by minimising (8.3.10b) with respect to β. Thus we need:

$$\nabla_{\beta}\left[\left(\mathbf{Y}^{\mathrm{T}}-\mathbf{X}^{\mathrm{T}}\beta^{\mathrm{T}}\right)\psi_{\mathrm{Y}}^{-1}\left(\mathbf{Y}-\mathbf{X}\beta\right)+\left(\beta^{\mathrm{T}}-\mu_{\beta}^{\mathrm{T}}\right)\psi_{\beta}^{-1}\left(\beta-\mu_{\beta}\right)\right]_{\beta=\hat{\beta}}=0 \quad (8.3.11)$$

which evaluates to:

$$2\left(-\mathbf{X}^{\mathrm{T}}\psi_{\mathrm{Y}}^{-1}\mathbf{Y}+\mathbf{X}^{\mathrm{T}}\psi_{\mathrm{Y}}^{-1}\mathbf{X}\hat{\beta}+\psi_{\beta}^{-1}\hat{\beta}-\psi_{\beta}^{-1}\mu_{\beta}\right)=0 \quad (8.3.12)$$

on noting that $\nabla_{\beta}(\mathbf{AB})=\nabla_{\beta}(\mathbf{A})\mathbf{B}+\nabla_{\beta}\left(\mathbf{B}^{\mathrm{T}}\right)\mathbf{A}^{\mathrm{T}}$ &

$$\nabla_{\beta}\left(\mathbf{A}^{\mathrm{T}}\phi\mathbf{A}\right)=2\left(\nabla_{\beta}\mathbf{A}^{\mathrm{T}}\phi\mathbf{A}\right);\ \text{while}\ \nabla_{\beta}\left(\mathbf{B}^{\mathrm{T}}\right)=\mathbf{I}$$

$$\text{Thus}\ \left(\mathbf{X}^{\mathrm{T}}\psi_{\mathrm{Y}}^{-1}\mathbf{X}+\psi_{\beta}^{-1}\right)\hat{\beta}\ =\ \mathbf{X}^{\mathrm{T}}\psi_{\mathrm{Y}}^{-1}\mathbf{Y}+\psi_{\beta}^{-1}\mu_{\beta} \quad (8.3.13)$$

$$\Rightarrow\hat{\beta}=\left(\mathbf{X}^{\mathrm{T}}\psi_{\mathrm{Y}}^{-1}\mathbf{X}+\psi_{\beta}^{-1}\right)^{-1}\left(\mathbf{X}^{\mathrm{T}}\psi_{\mathrm{Y}}^{-1}\mathbf{Y}+\psi_{\beta}^{-1}\mu_{\beta}\right) \quad (8.3.14)$$

By adding and subtracting $2\mathbf{X}^{\mathrm{T}}\psi_{\mathrm{Y}}^{-1}\mathbf{X}\mu_{\beta}$ to (8.3.12) we can get another expression for $\hat{\beta}$:

$$2\left(-\mathbf{X}^{\mathrm{T}}\psi_{\mathrm{Y}}^{-1}\mathbf{Y}+\mathbf{X}^{\mathrm{T}}\psi_{\mathrm{Y}}^{-1}\mathbf{X}\hat{\beta}+\psi_{\beta}^{-1}\hat{\beta}-\psi_{\beta}^{-1}\mu_{\beta}+\mathbf{X}^{\mathrm{T}}\psi_{\mathrm{Y}}^{-1}\mathbf{X}\mu_{\beta}-\mu_{\beta}\right)=0$$

$$=2\left(-\mathbf{X}^{\mathrm{T}}\psi_{\mathrm{Y}}^{-1}\left[\mathbf{Y}-\mathbf{X}\mu_{\beta}\right]+\left(\mathbf{X}^{\mathrm{T}}\psi_{\mathrm{Y}}^{-1}\mathbf{X}+\psi_{\beta}^{-1}\right)\hat{\beta}-\left(\mathbf{X}^{\mathrm{T}}\psi_{\mathrm{Y}}^{-1}\mathbf{X}+\psi_{\beta}^{-1}\right)\mu_{\beta}\right)=0$$

$$\Rightarrow\left(\mathbf{X}^{\mathrm{T}}\psi_{\mathrm{Y}}^{-1}\mathbf{X}+\psi_{\beta}^{-1}\right)\hat{\beta}=\mathbf{X}^{\mathrm{T}}\psi_{\mathrm{Y}}^{-1}\left(\mathbf{Y}-\mathbf{X}\mu_{\beta}\right)+\left(\mathbf{X}^{\mathrm{T}}\psi_{\mathrm{Y}}^{-1}\mathbf{X}+\psi_{\beta}^{-1}\right)\mu_{\beta}$$

$$\therefore\ \boxed{\hat{\beta}_{\mathrm{MAP}}=\mu_{\beta}+\left(\mathbf{X}^{\mathrm{T}}\psi_{\mathrm{Y}}^{-1}\mathbf{X}+\psi_{\beta}^{-1}\right)^{-1}\mathbf{X}^{\mathrm{T}}\psi_{\mathrm{Y}}^{-1}\left(\mathbf{Y}-\mathbf{X}\mu_{\beta}\right)} \quad (8.3.15)$$

This estimator has some interesting features: μ_{β} was our prior estimate and it is apparent that the posterior estimate is simply the prior estimate updated by a term which depends upon the new current information. Also $\hat{\beta}_{\mathrm{MAP}}$ is a biased estimator since $E[\beta] = \mu_{\beta}$ and therefore $E\left[\hat{\beta}_{\mathrm{MAP}}\right]=\mu_{\beta}$. In other words the distribution of $\hat{\beta}_{\mathrm{MAP}}$ is centered on the prior information and not the new experimental information. However this is not surprising, neither should it be considered a problem, since the process is focused on the prior information anyway—the new knowledge is considered as updating what is previously known.

Another significant feature is that the existence of $\hat{\beta}_{\mathrm{MAP}}$ depends only on the existence of $\left(\mathbf{X}^{\mathrm{T}}\psi_{\mathrm{Y}}^{-1}\mathbf{X}+\psi_{\beta}^{-1}\right)^{-1}$ and it is thus no longer a requirement that $\left|\mathbf{X}^{\mathrm{T}}\psi_{\mathrm{Y}}^{-1}\mathbf{X}\right|\neq 0$. This is of particular significance in dealing with comparison calibration experiments where $\mathbf{X}^{\mathrm{T}}.\mathbf{X}$ will always be singular. Essentially this method is performing a similar operation to the AD or RLS or GGM models insofar as it is including extra information but the manner in which it does this illustrates clearly the role of the prior information. We shall investigate shortly under what conditions this estimator is identical to AD or MLE. However, we can note that if we have no prior information, μ_{β} is undefined and $\psi_{\beta}\sim\infty$. Thus $\psi_{\beta}^{-1}\sim 0$ and we get:

$$\hat{\beta}_{MAP} = \left(\mathbf{X}^T \psi_Y^{-1} \mathbf{X}\right)^{-1} \mathbf{X}^T \psi_Y^{-1} \mathbf{Y}$$

which of course is simply the MLE or AD estimator. Under further, stricter assumptions that $\psi_Y = \sigma^2.\mathbf{I}$, we are reduced to Least Squares once more. So MAP is the most general estimator of which all the others are but special cases.

In mass calibration, there may not be complete prior information available. In such circumstances the unknown elements of μ_β can be given arbitrarily assigned values while their corresponding variances in ψ_β are given infinite values. Thus if there are no covariances at all in ψ_β, its inverse is easily obtained by inverting its diagonal terms, resulting in zeros for the unknown terms in ψ_β^{-1}. If there are covariances, matters are not so trivial, however by letting $s^2(b_k) \sim 10^{10}$ or some similar large number so that it will have negligible influence on the results. In Chapters 9 & 10 we will investigate the effect on the estimated parameter values of varying relative accuracies between ψ_β and ψ_Y.

8.4 Covariance Matrix of the MAP Estimator

In order to obtain the covariance matrix of the MAP estimate, we will use the equivalent form given in Eq. (8.3.14), restated here as:

$$\hat{\beta} = \mathbf{P}\mathbf{X}^T \psi_Y^{-1} \mathbf{Y} + \mathbf{P}\psi_\beta^{-1} \mu_\beta \tag{8.4.1a}$$

$$\text{where } \mathbf{P} = \left(\mathbf{X}^T \psi_Y^{-1} \mathbf{X} + \psi_\beta^{-1}\right)^{-1} \tag{8.4.1b}$$

Now from (8.3.4b) and (8.3.4c) we know that $\mathbf{cov}(\mathbf{Y}) = \psi_Y$ & $\mathbf{cov}(\mu_\beta) = \psi_\beta$. So from (8.4.1a):

$$\mathbf{cov}\left(\hat{\beta}\right) = \left(\mathbf{P}\mathbf{X}^T \psi_Y^{-1}\right) \psi_Y \left(\mathbf{P}\mathbf{X}^T \psi_Y^{-1}\right)^T + \left(\mathbf{P}\psi_\beta^{-1}\right) \psi_\beta \left(\mathbf{P}\psi_\beta^{-1}\right)^T \tag{8.4.2}$$

$$= \boxed{\mathbf{P}\mathbf{X}^T \psi_Y^{-1} \mathbf{X}\mathbf{P} \; + \; \mathbf{P}\psi_\beta^{-1}\mathbf{P}} \tag{8.4.3}$$

Note that $\mathbf{P}$ is a symmetric matrix. (8.4.3) can be expressed as:

$$\mathbf{P}\left(\mathbf{X}^T \psi_Y^{-1} \mathbf{X}\mathbf{P} \; + \; \psi_\beta^{-1}\mathbf{P}\right)$$

$$= \mathbf{P}\left(\left\{\mathbf{X}^T \psi_Y^{-1} \mathbf{X} \; + \; \psi_\beta^{-1}\right\}\mathbf{P}\right) \tag{8.4.4a}$$

But from (8.4.1a), this is just:

$$\mathbf{P}\left(\mathbf{P}^{-1}\mathbf{P}\right) = \mathbf{P} \tag{8.4.4b}$$

$$\text{so } \boxed{\psi_{\hat{\beta}} = \left\{\mathbf{X}^T \psi_Y^{-1} \mathbf{X} \; + \; \psi_\beta^{-1}\right\}^{-1}} \tag{8.4.5}$$

From this we see that $\psi_{\hat{\beta}}$ is made up of components due to the prior information and also components due to the new information obtained in the current experiment. The new estimate will have a lower covariance than the prior one as a result of the

minimum variance characteristics of the estimator. It is useful to note that this method lends itself naturally to a sequential estimation technique: for example, we can re-write (8.3.15) as:

$$\boxed{\beta_{k+1} = \beta_k + \left(\mathbf{X}^T\psi_Y^{-1}\mathbf{X} + \psi_k^{-1}\right)^{-1}\mathbf{X}^T\psi_Y^{-1}\left(\mathbf{Y} - \mathbf{X}\beta_k\right)} \tag{8.4.6}$$

and we would also have: $\boxed{\psi_{k+1} = \left\{\psi_k^{-1} + \mathbf{X}^T\psi_Y^{-1}\mathbf{X}\right\}^{-1}}$ (8.4.7)

so that the k^{th} estimate is updated to the $(k+1)^{th}$ estimate by means of the new data in $\mathbf{Y}$ & ψ_Y.

This estimator, using either (8.3.15) & (8.4.5) or (8.4.6) & (8.4.7) will form the basis of most of the later investigations in mass calibration reported elsewhere in this work. We will look in particular at properties of the estimator and how it deals with different types of prior/current information; and how it responds to varying relative accuracies between the two.

8.5 Relationship with Other Models

In Section 8.3 we noted that MAP had some similarities with the other methods we have discussed previously, in particular the AD method which also includes prior information. We wish now to consider the circumstances under which both of these methods would give the same solution, and indeed when they would differ. From Eq. (5.5.7) we know that the AD solution can be expressed as:

$$\hat{\beta}_{AD} = \left(\mathbf{X}^T\psi_Y^{-1}\mathbf{X} + \mathbf{A}\psi_R^{-1}\mathbf{A}^T\right)^{-1}\left(\mathbf{X}^T\psi_Y^{-1}\mathbf{Y} + \mathbf{A}\psi_R^{-1}\mathbf{R}\right) \tag{8.5.1}$$

while the MAP solution can be given by Eq. (8.3.14) as:

$$\hat{\beta}_{MAP} = \left(\mathbf{X}^T\psi_Y^{-1}\mathbf{X} + \psi_\beta^{-1}\right)^{-1}\left(\mathbf{X}^T\psi_Y^{-1}\mathbf{Y} + \psi_\beta^{-1}\mu_\beta\right) \tag{8.5.2}$$

where the various symbols have their usual meanings in this thesis. Considering the form of the two equations, we can see that they would produce the same parameter estimate if:

$$\begin{aligned} \psi_\beta^{-1} &= \mathbf{A}\psi_R^{-1}\mathbf{A}^T \\ \textit{and } \psi_\beta^{-1}\mu_\beta &= \mathbf{A}\psi_R^{-1}\mathbf{R} \end{aligned} \tag{8.5.3}$$

According to this $\mathbf{A}$ acts like a transformation matrix which transforms ψ_R^{-1}, the inverse covariance matrix of constraints in AD, into the inverse prior covariance matrix of MAP, ψ_β^{-1}. The crucial point we must remember is that $\mathbf{A}$ contains physical information about the actual prior information that is known, in the form of the m constraints, where usually $m \leq p$ if p is the number of parameters. So because of this,

the "transformation" of (8.5.3) is increasing the dimensions of ψ_R (m x m) up to those of ψ_β (p x p). However, this cannot be adding extra information as we do not know anything else a-priori, so there will have to be extras rows / columns of zeros in ψ_β^{-1} as we have already suggested in Section 8.3. To do this, **A** *must only have one non-zero element on each row, and this element must be unity.* In mathematics, such a matrix is sometimes referred to as a Hermite Canonical matrix. If we have prior information on anything from one up to all of the parameters, **A** will indeed satisfy this requirement and we can expect both AD and MAP to produce the same results. Indeed, if we *do* have complete prior information then **A=I** and the conditions of (8.5.3) are immediately satisfied. An example will help to develop the situation further. Suppose we have four parameters such that

$$\beta = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix};\ \mathbf{X} = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 1 & 0 & -1 & 0 \\ 1 & 0 & 0 & -1 \\ 0 & 1 & -1 & 0 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \end{bmatrix} \quad \& \quad \mathbf{Y} = \begin{bmatrix} y_1 \\ . \\ . \\ . \\ . \\ y_6 \end{bmatrix};\ \psi_Y = s^2\mathbf{I}_6 \tag{8.5.4}$$

The prior information comprises the values of b_1 & b_4 only, so for AD we will have:

$$\mathbf{A}^T\beta = \mathbf{R}$$

$$\Rightarrow \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix} = \begin{bmatrix} m_1 \\ m_4 \end{bmatrix} \text{ while } \psi_R = \begin{bmatrix} s^2(m_1) & 0 \\ 0 & s^2(m_4) \end{bmatrix} \tag{8.5.5}$$

where we have assumed no covariances in the prior information. If we were using MAP we would consider the prior variances of b_2 and b_3 to be infinite which means we can assume any prior value for them (we can use zero for convenience) and then we can write down the prior information as:

$$\mu_\beta = \begin{bmatrix} m_1 \\ 0 \\ 0 \\ m_4 \end{bmatrix} \quad \& \quad \psi_\beta^{-1} = \begin{bmatrix} s^{-2}(m_1) & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & s^{-2}(m_4) \end{bmatrix} \tag{8.5.6}$$

Using (8.5.5) & (8.5.6) it is now easy to show that (8.5.3) is satisfied in this case and thus MAP and AD will indeed produce the same estimates.

Now we will make the situation a little more complicated by supposing there *is* a covariance, c, in the prior information such that for AD we would have:

$$\mathbf{A}^T\beta = \mathbf{R}$$

$$\Rightarrow \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix} = \begin{bmatrix} m_1 \\ m_4 \end{bmatrix} \text{ while } \psi_R = \begin{bmatrix} s^2(m_1) & c \\ c & s^2(m_4) \end{bmatrix} \tag{8.5.7}$$

Along with (8.5.4) we can now easily obtain a solution by AD. However for MAP, we cannot so easily write down ψ_β^{-1} since the off-diagonal element means the inverse matrix is no longer simply the inverse of the diagonal elements. What we can do is assign a numerically large prior variance to parameters b_2 and b_3 such as below:

$$\psi_\beta = \begin{bmatrix} s^2(m_1) & 0 & 0 & c \\ 0 & 10^{10} & 0 & 0 \\ 0 & 0 & 10^{10} & 0 \\ c & 0 & 0 & s^2(m_4) \end{bmatrix} \tag{8.5.8a}$$

The inverse then evaluates to:

$$\psi_\beta^{-1} = \begin{bmatrix} \frac{s^2(m_4)}{s^2(m_1)s^2(m_4)-c^2} & 0 & 0 & \frac{-c}{s^2(m_1)s^2(m_4)-c^2} \\ 0 & 10^{-10} & 0 & 0 \\ 0 & 0 & 10^{-10} & 0 \\ \frac{-c}{s^2(m_1)s^2(m_4)-c^2} & 0 & 0 & \frac{s^2(m_1)}{s^2(m_1)s^2(m_4)-c^2} \end{bmatrix} \tag{8.5.8b}$$

The central two terms will obviously be practically negligible and can be approximated as zero. This does assume that we can choose a value (like 10^{10}) which is sufficiently large compared to the other information to be effectively infinite while still being computationally possible. In dealing with such extreme values there could be problems with numerical accuracy in results. Now evaluating ψ_R^{-1} from (8.5.7) yields:

$$\psi_R^{-1} = \begin{bmatrix} \frac{s^2(m_4)}{s^2(m_1)s^2(m_4)-c^2} & \frac{-c}{s^2(m_1)s^2(m_4)-c^2} \\ \frac{-c}{s^2(m_1)s^2(m_4)-c^2} & \frac{s^2(m_1)}{s^2(m_1)s^2(m_4)-c^2} \end{bmatrix} \tag{8.5.9}$$

With **A** given in (8.5.7) and approximating the two tiny terms in (8.5.8b) as zero, we can once again verify that (8.5.3) is satisfied and so both methods will yield the same results.

Finally let us suppose that the prior information available concerns the sum S, and difference, D, of the two parameters b_1 & b_4. In that case, the prior information for AD will be as follows:

$$\mathbf{A}^T\beta = \mathbf{R}$$

$$\Rightarrow \begin{bmatrix} 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix} = \begin{bmatrix} S \\ D \end{bmatrix} \text{ while } \psi_{\mathbf{R}} = \begin{bmatrix} r_1 & c \\ c & r_4 \end{bmatrix} \tag{8.5.10}$$

where for generality we have assumed there is a covariance between the two pieces of prior information. This information being independent from that represented by the experimental data as in (8.5.4), we can easily proceed with the AD method and obtain a solution. However, **A** as given in (8.5.10) above does not meet the requirements for a Hermite Canonical matrix and so we cannot expect MAP and AD to give the same results in this case. In fact, we cannot expect MAP to give a solution at all since we in fact have no information to construct the prior vector μ_β or covariance matrix ψ_β. We could 'invent' the information with the aid of the AD constraint data and equations (8.5.3), thus ensuring the two methods once again agreed, but the data we would generate would not be physically meaningful in terms of the prior information and so this would not be sensible.

In conclusion then, we have shown that both AD and MAP will give the same solutions in cases where the prior information comprises the values of some or all of the standards involved in the comparison exercise, and in such cases the MAP estimator is probably preferable since its Bayesian basis makes clear the type of analysis that is being carried out. However if we are not able to form a prior vector with at least one known parameter value, or if we need to incorporate other types of prior information, such as that expressed in (8.5.10) above—which does occur in some comparison exercises, particularly those involving primary standards—we must then use the AD approach to adequately incorporate the data. In what follows (Chapters 9 & 10) we will use the MAP method mostly since the experimental case studies we will report are suitable for this, and the separation of prior and current data is more clearly highlighted than with the AD method.

****

9. Parameter Estimation Techniques in Action

9.1 Introduction

This chapter introduces our first case study using actual experimental data and shows an implementation of the Unified Approach to Parameter Estimation and Uncertainty Analysis in mass calibration experiments. We show how the information is presented and how the various vectors/matrices are constructed, in particular noting the various contributions to ψ_Y, the covariance matrix of the experimental data.

We consider the Restrained Least Squares solution and highlight its crucial shortcomings, before proceeding to an in-depth analysis of the Bayesian Estimator. We probe the role of the prior information and show that relative accuracy and Degrees of Belief are important in establishing the posterior estimates; we see how the estimator would cope in the event of inaccurate prior information being used, pointing out its robustness and capabilities for correcting errors.

We also consider in some detail the role of the covariance matrix of the experimental data, and, of particular interest, highlight a theoretical limit on the improvements in accuracy that can be achieved with this estimator. We employ a novel graphical technique to show the range of values and the upper and lower bounds on the posterior parameter values and variances/covariances for a range of values of the experimental covariance matrix.

9.2 Example I

Initially we look at a comparison experiment involving three standards of nominal value 50g. The details of how the data is obtained and processed are explained in Chapter 11. For our purposes here we need simply state the data that is obtained and proceed to use it. Recall that the weighing equation is:

$$y_i \equiv \Delta w_i + \rho_{a_i} . \Delta v_i \tag{9.2.1}$$

where we leave out the corrections for centre of gravity differences and volume expansion coefficients in order to simplify things and also because their effects would not be significant with 50g standards. Along with the Weighing Equation we also have the System Model:

$$y_i = \sum_{j=1}^{p} x_{ij} . b_j \tag{9.2.2}$$

for the i^{th} comparison. Since the x_{ij} terms are either *1*, *0* or *-1*, (9.2.2) indicates which of the j parameters (b_j) are involved in each comparison. In our case there are three parameters and our parameter vector and design matrix are as follows:

$$\beta = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} \; ; \; \mathbf{X} = \begin{bmatrix} 1 & -1 & 0 \\ 1 & 0 & -1 \\ 0 & 1 & -1 \\ -1 & 1 & 0 \\ -1 & 0 & 1 \\ 0 & -1 & 1 \end{bmatrix}$$

Fig. 9.2.1: Parameter vector & Design Matrix

—i.e. all possible combinations are carried out, where half of them are simply reversed repeats of the other half; the purpose of this is to provide enhanced over-determinacy in order to get better information on the statistical nature of the experimental process. Table 9.2.1 gives the Prior Information which is available on the three parameters—to be viewed either stochastically or deterministically according to the chosen model. (Data from Calibration Certificates in PTB (1994), NPL (1990) & South Yorks.(1995))

Table 9.2.1 - Prior Information

Parameter	Value (μg)	Std. Dev.(μg)	Volume (cm^3)	Vol. Std. Dev./cm^3	OIML Class
b_1	-63.0	5.0	6.2202	0.0011	E_1
b_2	+34.0	15.0	6.3621	0.0009	E_2
b_3	+186.0	15.0	6.3468	0.0009	E_2

By way of explanation, the 'Value' quoted in Table 9.2.1 above for each parameter is a deviation from nominal value, expressed as physical mass—in this case a deviation from 50g. Because mass standards are classified according to their maximum permissible error (OIML, 1994), it is conventional to tabulate them in terms of their deviation, rather than absolute value. From the information in Table 9.2.1, we can form the vectors and matrices in Fig. 9.2.2 below.

$$\mu_\beta = \begin{bmatrix} -63 \\ 34 \\ 186 \end{bmatrix} \mu g; \quad \psi_\beta = \begin{bmatrix} 25 & 0 & 0 \\ 0 & 225 & 0 \\ 0 & 0 & 225 \end{bmatrix} \mu g^2$$

$$\mathbf{V} = \begin{bmatrix} 6.2202 \\ 6.3621 \\ 6.3468 \end{bmatrix} cm^3; \quad \psi_V = \begin{bmatrix} 1.21 & 0 & 0 \\ 0 & 0.81 & 0 \\ 0 & 0 & 0.81 \end{bmatrix} \times 10^{-6} \left(cm^3\right)^2$$

Fig. 9.2.2 : Prior Information

The variance of each piece of information has been taken as the square of its standard deviation quoted in Table 9.2.1. In particular note for ψ_β and ψ_V that we are assuming no correlations exist between any of the volumes or between any of the prior values of

the standards. ψ_β in particular might not always be diagonal: should the previous calibration have been a so-called "within-group" calibration involving all three of the present parameters, they would almost certainly have been correlated. However in the present circumstances we can only conclude that there are no correlations, based on the available information from calibration certificates. Here we are using the reasoning of Chapter 1 inasmuch as we must only use the information supplied, not basing decisions on hypothetical data.

The Experimental Information which is available is given in Table 9.2.2. Here the data quoted for the 'Weight-in-Air' difference ($\Delta\overline{W}$) are mean values of six experimental measurements in each case. From this the standard deviations of the data are obtained which are taken as experimental measures of the dispersion characteristics of the comparator used in the comparison.

However, because the standard deviations are smaller than the resolution of the instrument [1μg in this case], an extra dispersion term of $\pm\left(\frac{1}{2}\text{Resolution}\right)$, taken to be uniformly distributed, must be included with the standard deviation quoted for $\Delta\overline{W}$. (This type of reasoning is consistent with the conclusions of maximum entropy analysis as discussed in Chapter 1. See also Lira & Wöger (1997), Yoneda (1996)). The air density data given in each case is an average value for six measurements over the period of interest for each comparison. [See Chapter 11 for details]. From Eq. (3.2.5a) the variance of the air density is evaluated as $1{\cdot}45\times10^{-7}\left[\text{mg.cm}^{-3}\right]^2$ for the instruments used in the experimental work. The volume difference is evaluated from $\Delta\mathbf{V} = \mathbf{XV}$, with **X** & **V** as given above. The $\Delta\overline{m}$ term is calculated using Eq. (9.2.1).

Table 9.2.2- Experimental Information

$\Delta\overline{W}$ (μg)	Std. Dev (μg)	ρ_a (mg/cm^3)	ΔV (cm^3)	$\Delta\overline{m}$ (μg)
66.0	0.154	1.199856	-0.1419	-104.2
-109.0	0.379	1.202400	-0.1266	-261.2
-173.0	0.327	1.202216	0.0153	-154.6
-65.6	0.327	1.199876	0.1419	104.7
109.6	0.170	1.190610	0.1266	260.3
172.4	0.239	1.206947	-0.0153	153.9

$$\mathbf{Y} = \begin{bmatrix} -104.2 \\ -261.2 \\ -154.6 \\ 104.7 \\ 260.3 \\ 153.9 \end{bmatrix} \mu g; \quad \psi_{\Delta W} = \begin{bmatrix} 0.107 & 0 & 0 & 0 & 0 & 0 \\ & 0.227 & 0 & 0 & 0 & 0 \\ & & 0.190 & 0 & 0 & 0 \\ & & & 0.190 & 0 & 0 \\ & & symm & & 0.112 & 0 \\ & & & & & 0.140 \end{bmatrix} \mu g^2$$

$$\rho_a = \begin{bmatrix} 1.200 \\ 1.202 \\ 1.202 \\ 1.200 \\ 1.191 \\ 1.207 \end{bmatrix} mg.cm^{-3}; \quad \psi_{\rho_a} = s^2(\rho_a)\mathbf{I}_6$$

Fig. 9.2.3: Current (Experimental) Information.

Note that $\psi_{\Delta W}$ in Fig.(9.2.3) above is evaluated from the standard deviation data in Table 9.2.2 and also the variance of a rectangular distribution of width $\pm 0 \cdot 5 \mu g$ which is the term due to rounding errors in the comparator display, as explained earlier.

We need to evaluate $\psi_{\mathbf{Y}}$ the covariance matrix of our input data for the estimation techniques. To do this recall Eq. (4.5.17) from Chapter 4:

$$\psi_{\mathbf{Y}} = \psi_{\Delta W} + diag\{\mathbf{XV}\}\,\psi_{\rho}\,diag\{\mathbf{XV}\} + \rho\mathbf{X}\psi_{\mathbf{V}}\mathbf{X}^{\mathrm{T}}\rho^{\mathrm{T}} \qquad (9.2.3)$$

With the data in Figs.(9.2.1) to (9.2.3) we can now calculate Eq. (9.2.3) easily. We need to be careful with units since those of the second and third terms on the r.h.s. of Eq. (9.2.3) will evaluate in units of mg^2 since volumes are measured in cm^3 and air densities in $mg.cm^{-3}$. As we are using μg, and μg^2 as units in this analysis there must be a multiplicative factor of $1\,mg^2 \equiv 1 \times 10^6 \mu g^2$ applied to terms 2 and 3 on the r.h.s. of Eq. (9.2.3). Fig. (9.2.4) below gives the evaluation of term 2 ($\mathbf{t}_2$) & term 3 ($\mathbf{t}_3$) of Eq. (9.2.3) and also the complete $\psi_{\mathbf{Y}}$. Note how $\mathbf{t}_3$ is the one which introduces covariances, and indeed also the largest variance components.

The great convenience of this unified approach is that $\mathbf{Y}$ & $\psi_{\mathbf{Y}}$ are now a complete description of the corrected experimental data: there are no further calculations necessary at this level. The data can now be processed by an estimation technique to give complete covariances, assuming the method chosen is able to do this!

$$\mathbf{t}_2 = \begin{bmatrix} 3 & 0 & 0 & 0 & 0 & 0 \\ & 2.4 & 0 & 0 & 0 & 0 \\ & & 0.035 & 0 & 0 & 0 \\ & & & 3 & 0 & 0 \\ & & \textit{symm} & & 2.4 & 0 \\ & & & & & 0.035 \end{bmatrix} \times 10^{-3}\mu g^2$$

$$\mathbf{t}_3 = \begin{bmatrix} 2.90 & 1.74 & -1.17 & -2.90 & -1.73 & 1.17 \\ & 2.92 & 1.17 & -1.74 & -2.89 & -1.17 \\ & & 2.34 & 1.17 & -1.16 & -2.35 \\ & & & 2.91 & 1.73 & -1.17 \\ & & \textit{symm} & & 2.86 & 1.16 \\ & & & & & 2.36 \end{bmatrix} \mu g^2$$

$$\psi_Y = \begin{bmatrix} 3.01 & 1.74 & -1.17 & -2.90 & -1.73 & 1.17 \\ & 3.15 & 1.17 & -1.74 & -2.89 & -1.17 \\ & & 2.53 & 1.17 & -1.16 & -2.35 \\ & & & 3.10 & 1.73 & -1.17 \\ & & \textit{symm} & & 2.98 & 1.16 \\ & & & & & 2.50 \end{bmatrix} \mu g^2$$

Fig.9.2.4: Last 2 terms of Eq. (9.2.3) & Complete Covariance Matrix

9.3 RLS

Initially we will see how Restrained Least Squares handles the data. Using b_1 as the constraint, we have $\mathbf{A}^T = [1 \quad 0 \quad 0]$ and then with Eqs. (5.3.23) & (5.3.31) we get:

$$\hat{\beta} = \begin{bmatrix} -63.00 \\ 41.91 \\ 196.98 \end{bmatrix} \mu g; \quad \psi_{\hat{\beta}_Y} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 2.95 & 1.75 \\ 0 & 1.75 & 2.94 \end{bmatrix} \mu g^2$$

Fig. 9.3.1: Estimated Parameter Vector & Covariance Matrix using RLS & b_1 as constraint.

Two immediate observations from Fig.(9.3.1) are that (i) parameter b_1 remains unchanged by the estimation process and (ii) its variance/covariance terms are zero. This is as expected since b_1 is treated by the RLS estimator as a deterministic "constant". Of course we know that this is not really so, thus in this approach we treat the variance terms of the constraint as "systematic" uncertainties, i.e. those which cannot be affected by the experiment. From Section 5.4 we know that the "constraint contribution" to the overall covariance matrix is $\psi_{\hat{\beta}_R} = \mathbf{c}_2 \psi_R \mathbf{c}_2^T$ where ψ_R is the "systematic" covariance matrix of the constraint information and $\mathbf{c}_2 = \mathbf{a}_0^{-1}\mathbf{A}\left(\mathbf{A}^T\mathbf{a}_0^{-1}\mathbf{A}\right)^{-1}$. Interestingly, this depends explicitly on the form of **A** & **X**—i.e.

the form of the experimental and constraint design matrices. It is these that govern the constraint contribution to $\psi_{\hat{\beta}}$. In this example $\mathbf{R}$ and $\psi_{\mathbf{R}}$ are just scalars since the constraint information is just one parameter. With $\mathbf{a_0}$ as given in Eq. 5.3.16 we can easily evaluate the components as follows:

$$\mathbf{c_2} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \; ; \; \& \; \therefore \; \psi_{\hat{\beta}_R} = \sigma_{b_1}^2 . \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

Fig. 9.3.2a: "Constraint Contribution" to Final Covariance Matrix

Noting from Table 9.2.1 that $\sigma_{b_1}^2 = 25\mu g^2$, we find, with Fig. (9.3.1), that the final covariance matrix is:

$$\psi_{\hat{\beta}} = \begin{bmatrix} 25.0 & 25.0 & 25.0 \\ 25.0 & 27.95 & 26.75 \\ 25.0 & 26.75 & 27.94 \end{bmatrix} \mu g^2$$

Fig. 9.3.2b: Complete Covariance Matrix with RLS Method

With the parameter estimates in Fig. (9.3.1), the estimated experimental observations and residuals are evaluated as in Fig. (9.3.3).

$$\hat{\mathbf{Y}} = \mathbf{X}.\hat{\beta} = \begin{bmatrix} -104.913 \\ -259.986 \\ -155.073 \\ 104.913 \\ 259.986 \\ 155.073 \end{bmatrix} \mu g \; \& \; \left(\mathbf{Y} - \hat{\mathbf{Y}}\right) = \mathbf{res} = \begin{bmatrix} 0.813185 \\ -1.2143 \\ 0.472518 \\ -0.213185 \\ 0.314297 \\ -1.17252 \end{bmatrix} \mu g$$

Fig. 9.3.3: Estimated Observations & Residuals

The residuals provide a useful measure of the agreement between the estimated data and the original data. In this case the agreement is acceptable since the residuals are mostly of a similar order of magnitude to the standard deviations of the data in Table 9.2.2. If there were systematic errors in the data, such agreement would not be observed.

Before making further comments we will indicate the solutions obtained if b_2 or b_3 were used as constraints.

$$\hat{\beta} = \begin{bmatrix} -70.91 \\ 34.00 \\ 189.07 \end{bmatrix} \mu g \;\&\; \psi_{\hat{\beta}_V} = \begin{bmatrix} 2.95 & 0 & 1.2 \\ 0 & 0 & 0 \\ 1.2 & 0 & 2.39 \end{bmatrix} \mu g^2$$

Fig. 9.3.4a : RLS Solution using b_2 as constraint

$$c_2 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \;;\; \psi_{\hat{\beta}_R} = \sigma^2_{b_2} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \;\&\; \text{hence } \psi_{\hat{\beta}} = \begin{bmatrix} 227.952 & 225.0 & 226.201 \\ 225.0 & 225.0 & 225.0 \\ 226.201 & 225.0 & 227.391 \end{bmatrix} \mu g^2$$

Fig. 9.3.4b: Corresponding Constraint Contribution & Complete Covariance Matrix

$$\hat{\beta} = \begin{bmatrix} -73.9857 \\ 30.9275 \\ 186.0 \end{bmatrix} \mu g \;\&\; \psi_{\hat{\beta}_V} = \begin{bmatrix} 2.94 & 1.19 & 0 \\ 1.19 & 2.39 & 0 \\ 0 & 0 & 0 \end{bmatrix} \mu g^2$$

Fig. 9.3.5a: RLS Solution using b_3 as constraint

$$c_2 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \;;\; \psi_{\hat{\beta}_R} = \sigma^2_{b_3} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \;\&\; \text{hence } \psi_{\hat{\beta}} = \begin{bmatrix} 227.941 & 226.19 & 225.0 \\ 226.19 & 227.391 & 225.0 \\ 225.0 & 225.0 & 225.0 \end{bmatrix} \mu g^2$$

Fig. 9.3.5b: Corresponding Constraint Contribution & Complete Covariance Matrix

It can be quickly verified that $\hat{\mathbf{Y}}$ & **res** are the same as that given in Fig. (9.3.3) for the case where b_1 is used as constraint. In other words, when just one constraint is used, the same apparent agreement is reached with the data irrespective of the value [or known accuracy] of that constraint! Indeed this highlights a serious flaw with RLS, inasmuch as it cannot discriminate against bad constraint data. For example, if a value of $b_1 = -200\mu g$ were used, which would be totally wrong of course, values of $\hat{\beta}$ would be produced which would agree equally as well with the experimental data as does the present prior information but which would be entirely wrong as absolute values for the parameters. Now such an error would quickly become evident in other comparison experiments with other standards, but the point remains that this experiment with this fitting method will fail completely to find a problem.

We will see shortly how the other methods are much more robust in dealing with this situation. Although it should be borne in mind that with only one piece of prior information no estimation technique can totally compensate for errors in this single prior value as there are not enough degrees of freedom to make adjustments. With that in mind let us see what happens if we increase the prior information to two known values. We will see that this is not helpful where RLS is concerned, with this example,

since there are only three parameters in total. Let the prior information be the first two parameters so that we have:

$$\mathbf{R} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} \equiv \begin{bmatrix} -63.0 \\ 34.0 \end{bmatrix} \mu g \text{ then } \mathbf{A} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}; \text{ also } \psi_{\mathbf{R}} = \begin{bmatrix} 25 & 0 \\ 0 & 225 \end{bmatrix} \mu g^2$$

Fig. 9.3.6: Constraint Information for 2 parameters

Then we find:

$$\hat{\beta} = \begin{bmatrix} -63.0 \\ 34.0 \\ 192.29 \end{bmatrix} \mu g \quad \& \quad \psi_{\hat{\beta}_V} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1.9 \end{bmatrix} \mu g^2$$

Fig. 9.3.7a: Estimated Parameter Vector & Covariance Matrix Using 2 Constraints

Here the two constraint values remain unchanged and contribute no variance or covariance terms to the resulting covariance matrix. We can complete the covariance information by evaluating the constraint contribution as before:

$$\mathbf{c}_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0.4 & 0.6 \end{bmatrix} \quad \& \quad \psi_{\hat{\beta}_R} = \begin{bmatrix} 25 & 0 & 10.17 \\ 0 & 225 & 133.4 \\ 10.17 & 133.4 & 83.3 \end{bmatrix} \mu g^2, \text{so that } \psi_{\hat{\beta}} = \begin{bmatrix} 25 & 0 & 10.17 \\ 0 & 225 & 133.4 \\ 10.17 & 133.4 & 85.2 \end{bmatrix} \mu g^2$$

Fig. 9.3.7b: Constraint Contribution & Complete Covariance Matrix.

Again we can see that the prior information has remained unchanged in the process. The following values obtain for $\hat{\mathbf{Y}}$ & **res**:

$$\hat{\mathbf{Y}} = \mathbf{X}.\hat{\beta} = \begin{bmatrix} -97.0 \\ -255.292 \\ -158.292 \\ 97.0 \\ 255.292 \\ 158.292 \end{bmatrix} \mu g \quad \& \quad \mathbf{Y} - \hat{\mathbf{Y}} = \mathbf{res} = \begin{bmatrix} -7.1 \\ -5.9 \\ 3.69 \\ 7.7 \\ 5.0 \\ -4.39 \end{bmatrix} \mu g$$

Fig. 9.3.8: Estimated Observations & Residuals

So we can see from the residual vector that we now have a worse fit with respect to the original observations. Indeed, should one of the two fixed parameters be in fact in error, the posterior estimates deteriorate further. Suppose we have b_1 = *-63μg* as before but now b_2 = *20μg* rather than *34μg*. This is an error, but we are assuming the experimenter is not aware of it. We then obtain the same posterior covariance estimate since we are not changing the prior information in that regard, but the fitted values and residuals will be as in Fig. (9.3.9):

$$\hat{\beta} = \begin{bmatrix} -63.0 \\ 20.0 \\ 183.991 \end{bmatrix} \mu g;\ \hat{\mathbf{Y}} = \begin{bmatrix} -83.0 \\ -246.991 \\ -163.991 \\ 83.0 \\ 246.991 \\ 163.991 \end{bmatrix} \mu g \quad \& \quad \mathbf{res} = \begin{bmatrix} -21.1 \\ -14.2 \\ 9.39 \\ 21.7 \\ 13.3 \\ -10.09 \end{bmatrix} \mu g$$

Fig. 9.3.9: Estimated Values, Observations & Residuals

We can see the residuals have deteriorated further. This is because, by restraining b_1 & b_2 to fixed priors, all the adjustment must now be done on b_3, and any errors in the constraint information will be reflected in a bad fit to the data. Without further work it is not possible to know this since the experimenter may well conclude that the problem lies with the experimental data and not with the prior information. So in order to get good agreement with the experimental data it is best, with this estimation technique, to fix as little as possible of the data. However, because it is not possible to include variance/covariance information about this prior information, the technique is always at a disadvantage.

9.4 Bayesian Estimation

Now let us consider the same set of data treated by either MAP or AD. Since both will produce the same results for our present data as explained in Sec. 8.5, we will focus on MAP since its form is a little easier to analyse. The distinguishing feature of this method is that we can include all our known information about the parameters in the estimation, as was explained in Chapter 8. This of course includes variance/covariance information too. Using the values for μ_β, ψ_β, $\mathbf{X}$, $\mathbf{Y}$ & $\psi_\mathbf{Y}$ as given in Figs (9.2.1) to (9.2.4) above we can now evaluate the posterior estimates for the parameters using Eqs. (8.3.15) & (8.4.5):

$$\hat{\beta}_{MAP} = \begin{bmatrix} -64.69 \\ 40.07 \\ 195.13 \end{bmatrix} \mu g \quad \& \quad \psi_{\hat{\beta}_{MAP}} = \begin{bmatrix} 20.53 & 20.11 & 20.11 \\ 20.11 & 22.60 & 21.40 \\ 20.11 & 21.40 & 22.60 \end{bmatrix} \mu g^2$$

Fig. 9.4.1: Estimated Parameter Vector & Covariance Matrix using MAP

Where we can see that all three of the parameters have been updated, the covariance matrix is complete, and it is **smaller** than any of the combined covariance matrices achieved with the RLS estimators! (Compare Figs. (9.3.2b), (9.3.4b), (9.3.5b) for example.) Table 9.4.1 below gives a comparison between prior and posterior values for the parameters:

Table 9.4.1 - Prior / Posterior Values of the Parameters

Parameter	Prior Value μg	Prior Std. Dev. μg	Posterior Value μg	Posterior Std. Dev. μg
b_1	-63.0	5.0	-64.7	4.5
b_2	+34.0	25.0	+40.0	4.75
b_3	+186.0	25.0	+195.1	4.75

From this table, we can see that the standard deviations of b_2 & b_3 have been reduced by the largest amount in the posterior estimates. Clearly the influence of b_1, having the smallest standard deviation, is greatest. The fitted observations and residuals are:

$$\hat{\mathbf{Y}} = \begin{bmatrix} -104.762 \\ -259.819 \\ -155.057 \\ 104.762 \\ 259.819 \\ 155.057 \end{bmatrix} \mu g \quad \& \quad \mathbf{res} = \begin{bmatrix} 0.662 \\ -1.380 \\ 0.456 \\ -0.062 \\ 0.480 \\ -1.156 \end{bmatrix} \mu g$$

Fig. 9.4.2 : Estimated Observations & Residuals using MAP

So the Residuals are still acceptably small and indeed comparable to those obtained by RLS (See Fig.(9.3.3) above). But the smaller covariance matrix, and also the fact that it is complete by one calculation, make this method more desirable.

Now if there is only one piece of prior information, there is only one possible solution and all estimators will produce it. Thus if this information is in error, so also will be the result, albeit a good fit with the experimental data may well be possible. If we choose b_1 as the prior information, we then have:

$$\mu_\beta = \begin{bmatrix} -63.0 \\ 0 \\ 0 \end{bmatrix} \mu g \quad \& \quad \psi_\beta^{-1} = \begin{bmatrix} 1/25 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \mu g^2$$

Fig. 9.4.3: Prior Information, using 1 parameter

According to this ψ_β is not defined since ψ_β^{-1} is singular, but this is of no consequence since we do not need it. This is simply how we deal with a lack of information on parameters b_2 & b_3. The value of zero assigned to them in μ_β is also entirely arbitrary, since the *de facto* infinite variance assigned to them ensures that these arbitrary, unknown prior values will have no influence on the result. We obtain the same estimated values as does RLS, but with complete covariance matrix:

$$\hat{\beta} = \begin{bmatrix} -63.00 \\ 41.91 \\ 196.98 \end{bmatrix} \mu g; \quad \& \quad \psi_{\hat{\beta}} = \begin{bmatrix} 25 & 25 & 25 \\ 25 & 27.95 & 26.75 \\ 25 & 26.75 & 27.94 \end{bmatrix} \mu g^2$$

Fig. 9.4.4a: MAP Solution using one prior value (b_1)

$$\hat{\beta} = \begin{bmatrix} -70.91 \\ 34.00 \\ 189.07 \end{bmatrix} \mu g; \quad \& \quad \psi_{\hat{\beta}} = \begin{bmatrix} 227.95 & 225 & 226.20 \\ 225 & 225 & 225 \\ 226.20 & 225 & 227.39 \end{bmatrix} \mu g^2$$

Fig. 9.4.4b: MAP Solution using b_2 as prior information

$$\hat{\beta} = \begin{bmatrix} -73.99 \\ 30.93 \\ 186.00 \end{bmatrix} \mu g; \quad \& \quad \psi_{\hat{\beta}} = \begin{bmatrix} 227.94 & 226.19 & 225 \\ 226.19 & 227.39 & 225 \\ 225 & 225 & 225 \end{bmatrix} \mu g^2$$

Fig. 9.4.4c: MAP Solution using b_3 as prior information

Comments: It is clear that 'Degrees of Belief' about the prior information play an important part in establishing the estimates: prior information of lower accuracy is adjusted much more than prior information of higher accuracy. Thus in Fig. (9.4.1) the posterior variance of b_2 & b_3 is reduced from $225\mu g^2 \rightarrow\sim 23\mu g^2$ and their values are also adjusted more significantly than b_1, whose variance too is not adjusted so much. Fig. (9.4.4a) is an extreme case where there is no prior information on b_2 & b_3 and thus they are assigned infinite variance. In this case b_1 is not adjusted while b_2 & b_3 are adjusted even more than in Fig. (9.4.1) and their variance is reduced from $\sim \infty\ \mu g^2 \rightarrow\sim 28\mu g^2$. In Figs. (9.4.4b) & (9.4.4c), b_1 is treated as having infinite variance and is thus adjusted much more than in Fig. (9.4.1). Note that the variance cannot be reduced below that of the single piece of prior information.

It would appear from Table 9.4.1 that b_3's prior value is the most in error, or in need of updating, based on the current experimental information. If, for example, we used $b_3 = 195\mu g$ as a single piece of prior information instead of 186μg as in Fig. (9.4.4c), we would find b_1 adjusted much less. In fact

$$\hat{\beta} = \begin{bmatrix} -64.98 \\ 32.92 \\ 195.0 \end{bmatrix} \mu g$$

would result. But this can only be seen by comparison with Fig. (9.4.1) where all three prior values were used. Consider the cases where only two prior values are used:

$$\hat{\beta} = \begin{bmatrix} -63.78 \\ 41.00 \\ 196.10 \end{bmatrix} \mu g; \quad \& \quad \psi_{\hat{\beta}} = \begin{bmatrix} 22.59 & 22.24 & 22.35 \\ 22.23 & 24.86 & 23.79 \\ 22.35 & 23.79 & 25.11 \end{bmatrix} \mu g^2$$

Fig. 9.4.5a: MAP Result using b_1 & b_2 as Prior data

$$\hat{\beta} = \begin{bmatrix} -64.08 \\ 40.75 \\ 195.77 \end{bmatrix} \mu g; \quad \& \quad \psi_{\hat{\beta}} = \begin{bmatrix} 22.53 & 22.35 & 22.24 \\ 22.35 & 25.12 & 23.79 \\ 22.24 & 23.79 & 24.85 \end{bmatrix} \mu g^2$$

Fig. 9.4.5b: MAP Result using b_1 & b_3 as Prior data

$$\hat{\beta} = \begin{bmatrix} -72.45 \\ 32.47 \\ 187.53 \end{bmatrix} \mu g; \quad \& \quad \psi_{\hat{\beta}} = \begin{bmatrix} 114.85 & 112.50 & 112.50 \\ 112.50 & 113.10 & 111.90 \\ 112.50 & 111.90 & 113.10 \end{bmatrix} \mu g^2$$

Fig. 9.4.5c: MAP Result using b_2 & b_3 as Prior data

These results show that whenever b_1 (prior variance $25\mu g^2$) is included in the prior information (Figs 9.4.5a & b above), the posterior estimate for b_1 is only changed a little while b_2 & b_3 (prior variance $225\mu g^2$) are adjusted more significantly. But when b_1 is not included in the prior information, it is adjusted much more itself and b_2 & b_3 to a lesser extent.

This shows us the influence of prior information depends upon its relative accuracy: more accurate information will constrain the corresponding posterior estimate much more—indeed the case of a single piece of prior information is an extreme example of this. This leads us to consider the case of incorrect prior information—could this cause in-error posterior estimates to be produced, and if so, would we have any indication that this has occurred?

1) Consider the case of an error on b_3, such that:

$$\mu_\beta = \begin{bmatrix} -63.0 \\ 34.0 \\ 160.0 \end{bmatrix} \mu g \quad \& \quad \psi_\beta = \begin{bmatrix} 25 & 0 & 0 \\ 0 & 25 & 0 \\ 0 & 0 & 25 \end{bmatrix} \mu g^2$$

Fig. 9.4.6: Prior Information with an error on b_3

so that we are considering all the prior information to be of equal accuracy. Using the same experimental information as before we now obtain these posterior estimates:

$$\hat{\beta} = \begin{bmatrix} -77.08 \\ 27.16 \\ 180.91 \end{bmatrix} \mu g; \quad \& \quad \psi_{\hat{\beta}} = \begin{bmatrix} 9.31 & 7.84 & 7.84 \\ 7.84 & 9.15 & 8.00 \\ 7.84 & 8.00 & 9.15 \end{bmatrix} \mu g^2$$

Fig. 9.4.7a: Resulting Estimated Values

So b_3 has been adjusted significantly but the other two have also been altered by a relatively large amount compared with their prior standard deviations. We do have a reduced covariance matrix, but can we know that the data is valid? The Difference Vector (difference between Prior and Estimated Parameter values, $\left(\mu_\beta - \hat{\beta}\right)$) and Residual vector are shown below:

$$\left(\mu_\beta - \hat{\beta}\right) = \begin{bmatrix} 14.09 \\ 6.83 \\ -20.92 \end{bmatrix} \mu g; \text{ while } \mathbf{res} = \begin{bmatrix} 0.155 \\ -3.196 \\ -0.851 \\ 0.445 \\ 2.296 \\ 0.151 \end{bmatrix} \mu g$$

Fig. 9.4.7b: Difference Vector and Experimental Residuals

2) Now if we maintain the prior value of b_3 ($160\mu g$) but allow it a much larger prior variance as shown below, we then find the results shown in Figs. (9.4.7c) & (9.4.7d).

$$\psi_\beta = \begin{bmatrix} 25 & 0 & 0 \\ 0 & 25 & 0 \\ 0 & 0 & 900 \end{bmatrix} \mu g^2$$

$$\hat{\beta} = \begin{bmatrix} -67.18 \\ 37.28 \\ 192.47 \end{bmatrix} \mu g; \quad \& \quad \psi_{\hat{\beta}} = \begin{bmatrix} 13.03 & 11.63 & 12.17 \\ 11.63 & 13.02 & 12.43 \\ 12.17 & 12.43 & 14.20 \end{bmatrix} \mu g^2$$

Fig. 9.4.7c: Alternative Estimate with larger b_3 prior variance

$$\left(\mu_\beta - \hat{\beta}\right) = \begin{bmatrix} 4.18 \\ -3.28 \\ -32.47 \end{bmatrix} \mu g; \text{ while } \mathbf{res} = \begin{bmatrix} 0.363 \\ -1.55 \\ 0.587 \\ 0.237 \\ 0.650 \\ -1.287 \end{bmatrix} \mu g$$

Fig. 9.4.7d: Difference Vector & Resulting Residuals

In this case b_3 is adjusted much more and the other parameters much less—reflecting the (incorrect) prior value of b_3 being given a "smaller" degree of belief. This results in slightly better residuals—i.e. estimated values which are in better agreement with the experimental data. Compare the **res** vectors in Figs. (9.4.7b) & (9.4.7d).

However, it is still the case that the covariance matrix of the experimental information is much smaller than the prior information—i.e. $\psi_Y << \psi_\beta$. (Compare ψ_Y in Fig. (9.2.4) with ψ_β in Fig. (9.4.6)). Let us consider the case where both are of a

similar order of magnitude and we retain the "error" on the prior value of b_3. Fig (9.4.8) gives the prior information while Figs. (9.4.9a) & (9.4.9b) give the resulting estimated values.

$$\mu_\beta = \begin{bmatrix} -63.0 \\ 34.0 \\ 160.0 \end{bmatrix} \mu g \quad \& \quad \psi_\beta = \begin{bmatrix} 4 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 4 \end{bmatrix} \mu g^2$$

Fig. 9.4.8: Prior Information with smaller covariance ($\psi_\beta \approx \psi_Y$)

$$\hat{\beta} = \begin{bmatrix} -73.78 \\ 28.21 \\ 176.58 \end{bmatrix} \mu g; \quad \& \quad \psi_{\hat{\beta}} = \begin{bmatrix} 2.08 & 0.96 & 0.96 \\ 0.96 & 1.98 & 1.06 \\ 0.96 & 1.06 & 1.98 \end{bmatrix} \mu g^2$$

Fig. 9.4.9a: Estimated Values & Covariance Matrix with incorrect prior b_3 & $\psi_\beta \approx \psi_Y$

$$\left(\mu_\beta - \hat{\beta}\right) = \begin{bmatrix} 10.79 \\ 5.79 \\ -16.58 \end{bmatrix} \mu g; \text{ while } \mathbf{res} = \begin{bmatrix} -7.2 \\ -38.2 \\ -28.6 \\ 7.7 \\ 37.3 \\ 27.9 \end{bmatrix} \mu g$$

Fig. 9.4.9b: Difference & Residual vectors

Now the Difference Vector indicates less adjustment to the prior values while the residuals are large implying very poor agreement with the experimental data. This indicates that the incorrect value for b_3 is now having a more significant effect upon the result. However, if we let $s^2(b_3)$ increase as shown below

$$\psi_\beta = \begin{bmatrix} 4 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 900 \end{bmatrix} \mu g^2$$

we get the data in Figs. (9.4.9c) & (9.4.9d) which indicate a big adjustment for b_3 while the others are adjusted much less Also the residuals are very much smaller indicating that improved agreement with the experimental information now exists.

$$\hat{\beta} = \begin{bmatrix} -65.959 \\ 36.814 \\ 192.688 \end{bmatrix} \mu g; \quad \& \quad \psi_{\hat{\beta}} = \begin{bmatrix} 2.53 & 1.45 & 1.89 \\ 1.45 & 2.53 & 2.09 \\ 1.89 & 2.09 & 3.90 \end{bmatrix} \mu g^2$$

Fig. 9.4.9c: Estimation Data with $s^2(b_3)$ large

$$\left(\mu_\beta - \hat{\beta}\right) = \begin{bmatrix} 2.959 \\ -2.814 \\ -32.688 \end{bmatrix} \mu g; \text{ while } \mathbf{res} = \begin{bmatrix} -1.327 \\ -2.553 \\ 1.274 \\ 1.927 \\ 1.653 \\ -1.974 \end{bmatrix} \mu g$$

Fig. 9.4.9d: Corresponding Difference & Residual Vectors

So we can conclude that the relative accuracy of the prior and current information is important as well as the relative accuracy among the respective elements of the prior information. Incorrect prior information will have minimal influence on the posterior estimates if all the prior information is of lower accuracy than the current information. However, even then, it will exert some influence if it is of equal or greater accuracy than the other elements of μ_β. On the other hand, if the prior information *is* of similar accuracy to the experimental information, any errors in the prior information can have devastating effects on the posterior data. Of course all the same remarks apply to the reciprocal situation of errors in the experimental data.

Thus the MAP Estimator is remarkably robust inasmuch as it can handle both "good" (consistent) data and can deal very well with inconsistent data too. In cases where it cannot correct for problems, it will nevertheless highlight them via significantly adjusted posterior estimates or large experimental residuals. It may not be possible, directly from such data, to decide whether prior or current data is at fault–this may require supplementary investigations–but nevertheless the existence of a problem will be clearly highlighted. It should be noted that none of this analysis is possible with the rigid RLS method!

9.5 Significance of the Covariance Matrix, ψ_Y

We have seen that ψ_Y contains two terms due to the buoyancy correction, $\rho\mathbf{XV}$. (See Eq. (9.2.3) for example). Now it has been commonplace in mass calibration to leave these covariance terms out of the estimation process and include them afterwards via some other calculation (e.g. Schwartz, (1991)), or else to ignore them altogether, (Bich, (1989a), (1989b), (1993a)), considering them to be of no significance to the final uncertainty quoted. Indeed, $\psi_Y = \sigma^2.\mathbf{I}$ is often used as a result. Sometimes, uncertainties due to the volumes of the standards are ignored too (e.g. Lewis et al (1990)).

However in our Unified Approach the dispersion characteristics of all influence quantities must be included and since ψ_Y appears in the estimation equation, this may have some effect upon the results.

Let the dispersion matrix be given by $\psi_Y = \psi_{\Delta W}$, where $\psi_{\Delta W}$ is as given in Fig. (9.2.3), the two terms due to the buoyancy correction being ignored, ($\mathbf{t_2}$ & $\mathbf{t_3}$ in Fig. (9.2.4)). Carrying out the Bayesian estimation now results in:

$$\hat{\beta} = \begin{bmatrix} -64.71 \\ 40.24 \\ 195.18 \end{bmatrix} \mu g; \quad \& \quad \psi_{\hat{\beta}} = \begin{bmatrix} 20.46 & 20.45 & 20.45 \\ 20.45 & 20.49 & 20.46 \\ 20.45 & 20.46 & 20.49 \end{bmatrix} \mu g^2$$

Fig. 9.5.1: Estimated Values & Covariance Matrix when using $\psi_Y = \psi_{\Delta W}$

These values are compared directly with Fig. (9.4.1) in Table 9.5.1 below where we can see that the difference between the estimated values in each case is small while there is some reduction in the variance and covariance terms. This is expected since ψ_Y is now much smaller. However because ψ_β is so much bigger than ψ_Y, the effects of adjusting ψ_Y are not manifest very clearly. Later we will find that if ψ_β and ψ_Y are of a comparable order of magnitude a reduction in ψ_Y has a more marked effect. What we wish to do first is consider the range of variations in the posterior estimates that are possible with the given initial conditions (i.e. prior information).

Table 9.5.1 - Comparison of Estimated Values & Variances for ψ_Y Diagonal / Non-Diagonal

	Variances $(\mu g)^2$			Values (μg)		
Parameter	Fig. (9.4.1)	Fig. (9.5.1)	% Difference	Fig. (9.4.1)	Fig. (9.5.1)	Difference
b_1	20.53	20.46	-0.3%	-64.69	-64.71	-0.02
b_2	22.60	20.49	-9.3%	40.073	40.24	0.17
b_3	22.60	20.49	-9.3%	195.13	195.18	0.05

Consider the case where we transform ψ_Y according to $\psi_Y \rightarrow \nu.\mathbf{I}$ where ν is a scalar multiplier. Using all the same data as before for ψ_β and $\mathbf{X}$ we obtain for $\psi_{\hat{\beta}}$:

$$\psi_{\hat{\beta}} = \begin{bmatrix} \frac{25(450+\nu)}{550+\nu} & \frac{11250}{550+\nu} & \frac{11250}{550+\nu} \\ & \frac{225\left(67500+1000\nu+\nu^2\right)}{(550+\nu)(1350+\nu)} & \frac{101250(150+\nu)}{(550+\nu)(1350+\nu)} \\ & & \frac{225\left(67500+1000\nu+\nu^2\right)}{(550+\nu)(1350+\nu)} \end{bmatrix}$$

Fig. 9.5.2a: Theoretical Form of $\psi_{\hat{\beta}}$ for $\psi_{\mathbf{Y}} = \nu.\mathbf{I}$

This is interesting as it shows that–for this particular ψ_{β} data–the two limiting cases are as given below in Figs (9.5.2b) & (9.5.2c):

$$\psi_{\hat{\beta}} = \begin{bmatrix} 20.45 & 20.45 & 20.45 \\ 20.45 & 20.45 & 20.45 \\ 20.45 & 20.45 & 20.45 \end{bmatrix} \mu g^2$$

Fig. 9.5.2b: $\psi_{\hat{\beta}}$ for $\nu \rightarrow 0$ (Current data infinitely accurate)

$$\psi_{\hat{\beta}} = \begin{bmatrix} 25 & 0 & 0 \\ 0 & 225 & 0 \\ 0 & 0 & 225 \end{bmatrix} \mu g^2$$

Fig. 9.5.2c: $\psi_{\hat{\beta}}$ for $\nu \rightarrow \infty$ (Current data absolutely useless and inaccurate, no information!)

The latter figure shows that we are left with just the prior information as before. This is useful as it provides a benchmark with which to compare Figs. (9.5.1) & (9.4.1). In fact we can see that $\psi_{\hat{\beta}}$ in Fig. (9.5.1), obtained for $\psi_{\mathbf{Y}} = \psi_{\Delta\mathbf{W}}$, is very close to the theoretical limit for improvements in accuracy obtainable by the estimation method. Clearly when the variance/covariance information due to the buoyancy correction is included, a larger $\psi_{\mathbf{Y}}$ results and so $\psi_{\hat{\beta}}$ will always be bigger than Fig (9.5.2b)

In considering the theoretical basis for what we have just seen, we need to recall Eq. (8.4.3) for the posterior covariance matrix:

$$\psi_{\hat{\beta}} = \left(\mathbf{X}^{\mathrm{T}}\psi_{\mathbf{Y}}^{-1}\mathbf{X} + \psi_{\beta}^{-1}\right)^{-1}$$

From this we can see that the upper limit for a very large $\psi_{\mathbf{Y}}$ occurs when $\left(\mathbf{X}^{\mathrm{T}}\psi_{\mathbf{Y}}^{-1}\mathbf{X}\right) \rightarrow \mathbf{0}$ and then the posterior covariance matrix $\psi_{\hat{\beta}} \rightarrow \psi_{\beta}$ as suggested by Fig. (9.5.2c) above. The lower limit occurs for very accurate experimental information where $\psi_{\mathbf{Y}}$ becomes very small, in which case $\left(\mathbf{X}^{\mathrm{T}}\psi_{\mathbf{Y}}^{-1}\mathbf{X}\right) \rightarrow \infty$ and we find

that the posterior covariance matrix $\psi_{\hat{\beta}} \to \left(\mathbf{X}^{\mathrm{T}}\psi_{\mathbf{Y}}^{-1}\mathbf{X}\right)^{-1}$. However, we cannot directly compute this owing to the singularity in $\mathbf{X}^{\mathrm{T}}\mathbf{X}$, meaning that we must utilise some sort of numerical method in order to approach this limit.This is what Fig. (9.5.2a) represents and which will be developed further in the following pages

What we have just done clarifies some very important aspects of this Bayesian estimation technique. We have pointed out that there *is* a lower limit on the posterior variances for a given set of prior information (initial conditions); an issue to be discussed further in what follows. We have additionally shown that the prior information represents what we might call a "worst case" with regard to the posterior covariance matrix. What this means is that we effectively have a system where we can add new stochastic information (from the experiment) *without adding noise or 'uncertainty'*; a point which again highlights the utility of this estimator.

Using the transformation $\psi_{\mathbf{Y}} \to \nu.\mathbf{I}$ is of course a simplification, which makes computation easier. However, this is not a critical over-simplification since we are primarily interested in what happens at the limits of $\nu \to 0$ and $\mathrm{v} \to \infty$. We could also consider the transformation $\psi_{\mathbf{Y}} \to \nu.\psi_{\mathbf{Y}}$ which would use the actual experimental information. (The algebraic form for this is given in Fig. 9.5.3 below). If we were interested in exploring the role of the buoyancy correction variance/covariance terms, we could use $\psi_{\mathbf{Y}} \to \psi_{\Delta\mathrm{W}} + \nu(\mathbf{t}_2 + \mathbf{t}_3)$ In Fig. (9.5.4) below, we have shown a simulation of $s^2\left(\hat{b}_1\right)$ for ν in the range $\nu = 10^{-5} \to \mathrm{v} = 10^{10}$—which for all practical purposes is the range $(0,\infty)$. The graph shows the results using both $\nu\mathbf{I}$ and $\nu\psi_{\mathbf{Y}}$. The other possibility we mentioned is in fact identical to the curve for $\psi_{\mathbf{Y}} \to \nu.\psi_{\mathbf{Y}}$ because $\psi_{\Delta\mathrm{W}}$ is so much smaller than the terms due to the buoyancy correction.

$$\psi_{\hat{\beta}} = \begin{bmatrix} \dfrac{25(47.9038+\nu)}{58.5488+\nu} & \dfrac{1192.04(189.07+\nu)}{(58.5488+\nu)(188.195+\nu)} & \dfrac{1203.17(187.32+\nu)}{(58.5488+\nu)(188.195+\nu)} \\ \dfrac{225(8.30508+\nu)(120.612+\nu)}{(58.5488+\nu)(188.195+\nu)} & \dfrac{15782.6(14.28+\nu)}{(58.5488+\nu)(188.195+\nu)} & \\ & \dfrac{225(8.33814+\nu)(120.134+\nu)}{(58.5488+\nu)(188.195+\nu)} & \end{bmatrix}$$

Fig. 9.5.3: Theoretical Form of $\psi_{\hat{\beta}}$ for $\psi_{\mathbf{Y}} \to \nu.\psi_{\mathbf{Y}}$

In Fig. (9.5.4) we can see that as $\mathrm{v} \to \infty$ and the current data's accuracy decreases, the posterior variance reverts to that of the prior data, while as $\nu \to 0$ the current data exerts an ever larger influence on the estimation process and we see a lower limit on

the posterior variance for infinitely accurate experimental data[1]. We also notice that (as we would expect) the curve for $\psi_Y \to v\psi_Y$ is everywhere larger than the one for $\psi_Y \to v\mathbf{I}$ except at the limits where they converge. Fig. (9.5.5) below illustrates the same features for the posterior variances of parameters b_2 and b_3. We will only use the transformation $\psi_Y \to v\mathbf{I}$ from this point, as it is primarily the limiting values we are interested in.

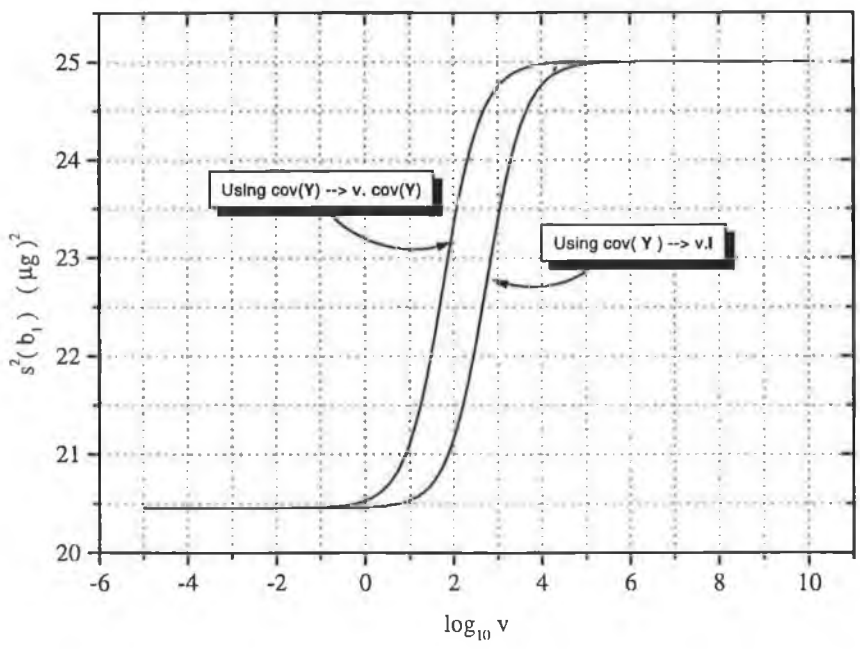


Fig. 9.5.4: $s^2\left(\hat{b}_1\right)$ for $\psi_Y \to v\mathbf{I}$ & $\psi_Y \to v\psi_Y$ for values of v in the range $\sim (0, \infty)$

Fig. (9.5.6) shows a similar simulation for the posterior *covariance* between parameters b_1 & b_2 and b_1 & b_3. There is no covariance in the prior information and so we see the graph approaches zero as $v \to \infty$. On the other hand, as $v \to 0$ the experimental information exerts a larger influence and the effect of the correlation intrinsic to the mass comparison process becomes more pronounced in the posterior estimate.

The distribution of covariance information is not always straight-forward as can be seen in Fig. (9.5.7) where we show the simulation for the covariance of b_2 & b_3. We need to recall that a covariance must always be considered in relation to *two*

[1] The curve we have shown is in fact a Sigmoidal-type function and can be easily modelled as a Boltzmann equation of the form $y = \frac{A_1 - A_2}{1 + e^{x - x_0}} + A_2$ where A_1 & A_2 are the limits of the function.

parameters and thus the <u>total</u> variance information[2] will always reduce (with this estimation technique) as the experiment becomes more accurate, albeit the correlation between individual parameters may increase. Recall also that we did not design the comparison matrix **X** specifically to minimise covariances in the posterior estimates.

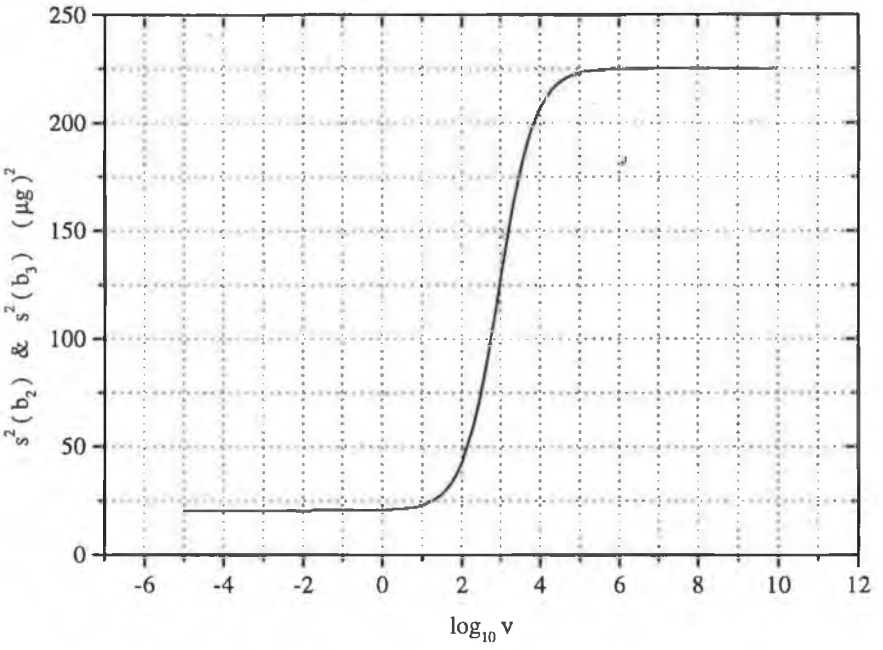


Fig. 9.5.5: $s^2\left(\hat{b}_2\right)$ & $s^2\left(\hat{b}_3\right)$ for $\psi_Y \rightarrow v.\mathbf{I}$ for v in the range ~ $(0,\infty)$

[2] For example, the total variance information for the sum of two parameters b_1 & b_2 depends on both variance and covariance according to $\text{var}(b_1+b_2)=s^2(b_1)+s^2(b_2)+s(b_1,b_2)$

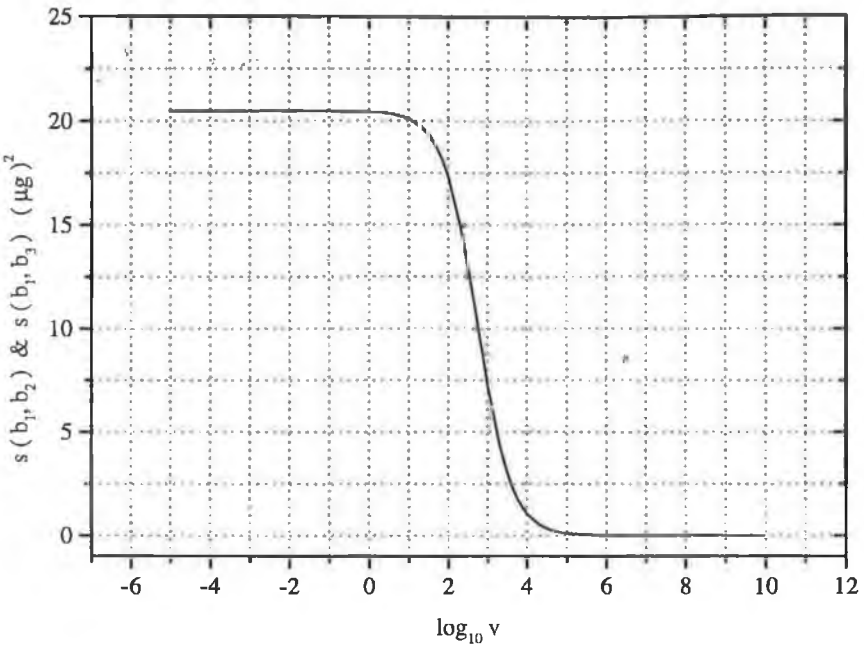


Fig. 9.5.6: $s\left(\hat{b}_1, \hat{b}_2\right)$ & $s\left(\hat{b}_1, \hat{b}_3\right)$ for $\psi_Y \to \nu.\mathbf{I}$ with ν in the range $\sim (0, \infty)$

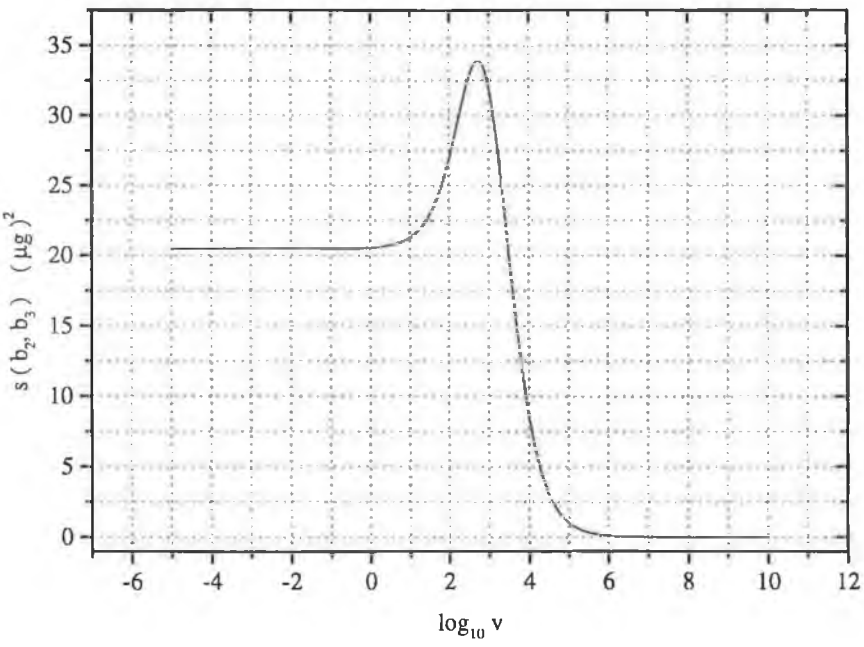


Fig. 9.5.7: $s\left(\hat{b}_2, \hat{b}_3\right)$ for $\psi_Y \to \nu.\mathbf{I}$ with ν in the range $\sim (0, \infty)$

While considering the estimator's capabilities in regard to the covariance matrix, it is also interesting to look at the possible range in estimated parameter values which can occur when $\psi_Y \to \nu.\mathbf{I}$ for ν in the range $(0,\infty)$. When $\nu \to 0$ and the current data is exerting maximum influence, the largest adjustment to the prior values takes place, based on the information obtained in the comparison experiment; while at the other end of the scale as $\nu \to \infty$ the posterior estimates are unchanged from the prior information. Figs. (9.5.9) to (9.5.11) illustrate this information for parameters b_1 to b_3. As we have already discussed in Section 9.4, the adjustment that can be carried out does depend significantly on the relative accuracy of the prior information. So while Fig. (9.5.8) shows the general form of $\hat{\beta}$ for the given prior covariance matrix ψ_β, a different prior covariance matrix would result in different lower bounds on the posterior estimates (i.e. as $\nu \to 0$). The upper bound as $\nu \to \infty$ would of course remain unchanged as the prior information.

$$\begin{bmatrix} \dfrac{-63(565.158+\nu)}{(550+\nu)} \\ \dfrac{34(688.912+\nu)(1277.93+\nu)}{(550+\nu)(1350+\nu)} \\ \dfrac{186(572.786+\nu)(1361.21+\nu)}{(550+\nu)(1350+\nu)} \end{bmatrix}$$

Fig. 9.5.8: General form of $\hat{\beta}$ for the given data and $\psi_Y \to \nu.\mathbf{I}$

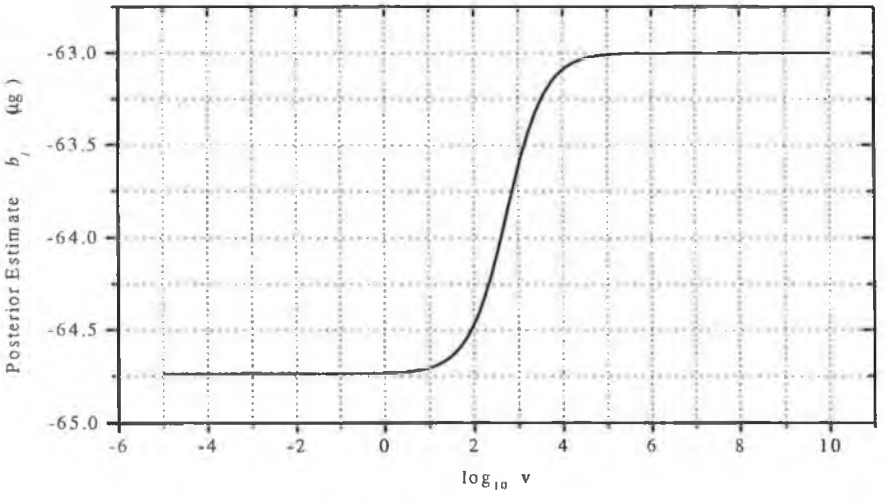


Fig. 9.5.9: $\hat{b}_1$ values for ν in the range $(0,\infty)$ when $\psi_Y \to \nu.\mathbf{I}$

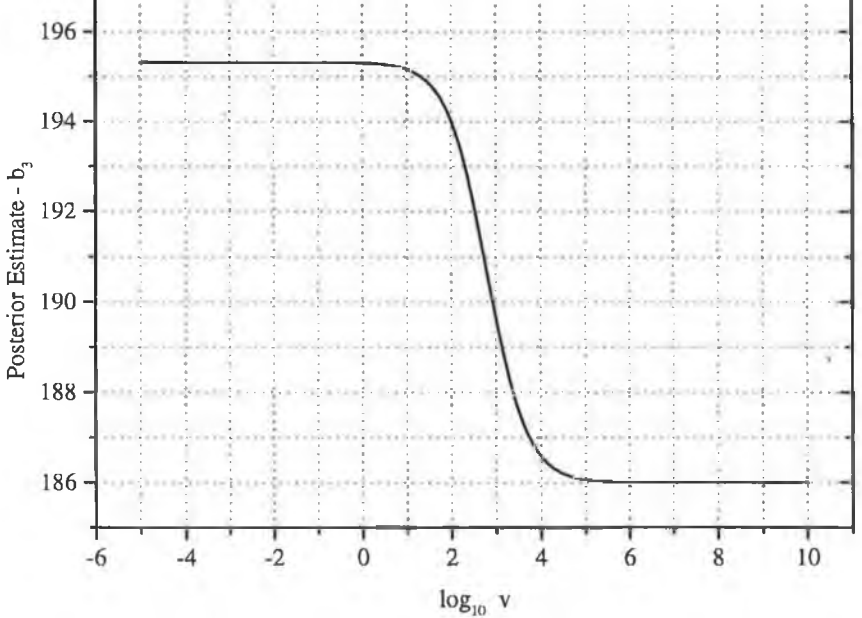


Fig. 9.5.11: $\hat{b}_3$ values for ν in the range $(0,\infty)$ when $\Psi_{\mathbf{Y}} \to \nu.\mathbf{I}$

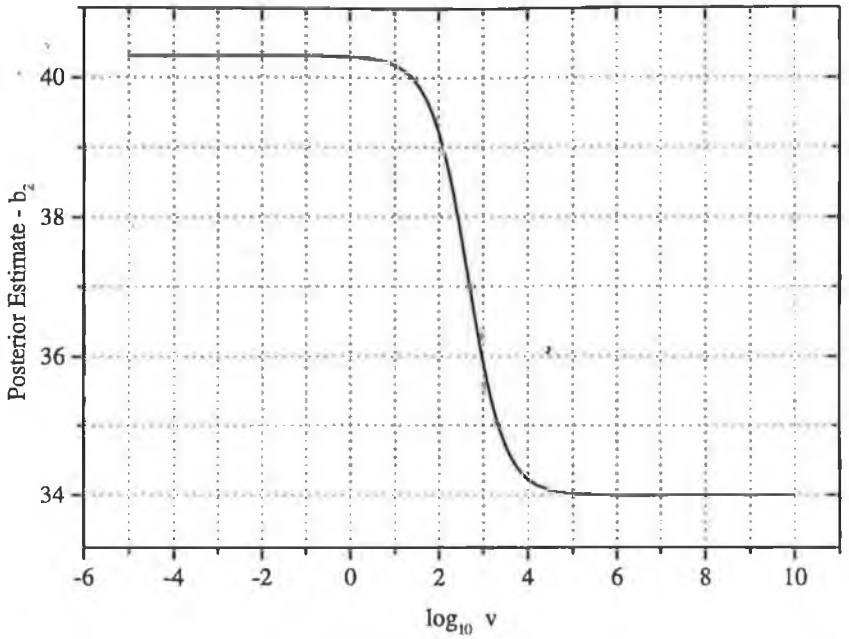


Fig. 9.5.10: $\hat{b}_2$ values for ν in the range $(0,\infty)$ when $\psi_Y \to \nu.\mathbf{I}$

We might note, as shown earlier, (Eqs. 8.4.6 & 8.4.7), that the MAP estimator can be written recursively, so that posterior estimates β_{k+1} & $\psi_{\beta_{k+1}}$ are produced from prior knowledge β_k & ψ_{β_k} and new current data $\mathbf{Y}$ & $\psi_{\mathbf{Y}}$. As we have seen, theoretically and experimentally, the posterior estimates always have lower covariance matrices than does the prior information. Since these posterior estimates will form the new prior information for subsequent estimations, as shown in Fig. (8.0.2) for example, we can expect the next posterior estimate to have still lower covariance. So, specifically in mass calibration, does this mean that every time we re-calibrate our standards we achieve lower covariance matrices each time? Of course we will not! It is a fair assumption that $\psi_{\mathbf{Y}}$ will not change all that much between one calibration and the next since probably the same equipment will be used. In that case, the posterior covariance matrix for successive calibrations will tend to converge to a lower limit after which no further improvements will be possible. In fact this lower limit is none other than the one shown in the variance / covariance plots of Figs (9.5.4), to (9.5.7), for the transformation $\psi_{\mathbf{Y}} \rightarrow \nu.\mathbf{I}$ when $\nu \rightarrow 0$. We can appreciate this by recalling that $\nu \rightarrow 0$ corresponds to the experimental data being infinitely accurate. Now while this is a highly idealised proposition, it nevertheless corresponds to a situation where we could not learn anything new about the mass standards involved. Certainly in subsequent calibrations, it is highly likely that the standards will have drifted somewhat, an issue we address in more detail in the next chapter, and therefore there will of course be new information to learn about the standards themselves, but we can be sure that we will not evaluate any posterior estimates to higher accuracy than that which occurs when $\nu \rightarrow 0$ in our simulation. We should remark however, that this analysis would assume we use the same parameters in subsequent evaluations. Should we introduce new parameters and change the design scheme the scope of the problem is changed and new information of higher accuracy may well be obtainable.

So far we have shown, that for our example data, the difference in final covariance matrix from leaving out the buoyancy correction variance/covariance terms in the analysis (Fig. (9.5.1)), is not enormous. We then considered the possible variation which could occur for all possible values of the input covariances and this led us to establish upper and lower bounds for the achievable accuracy. However, it will be recalled from Section 9.4, where we discussed the technique's robustness in dealing with incorrect information, that the relative accuracy of the prior/current information was significant in this regard. We now want to see what happens to our covariance analysis when $\psi_{\beta} \cong \psi_{\mathbf{Y}}$. To do this we will assume $\psi_{\beta} = diag[4,4,4]$ which is of a similar order of magnitude to $\psi_{\mathbf{Y}}$ as given in Fig. (9.2.4) earlier. In this case,

using ψ_Y as given in Fig. (9.2.4) results in the estimated values of Fig. (9.5.12a), while if we let $\psi_Y = \psi_{\Delta W}$, which eliminates in particular all the off-diagonal terms, we find the results given in Fig. (9.5.12b):

$$\hat{\beta} = \begin{bmatrix} -67.52 \\ 35.08 \\ 189.45 \end{bmatrix} \mu g \quad \& \quad \psi_{\hat{\beta}} = \begin{bmatrix} 2.084 & 0.957 & 0.959 \\ 0.957 & 1.982 & 1.060 \\ 0.959 & 1.060 & 1.980 \end{bmatrix} \mu g^2$$

Fig. 9.5.12a: MAP results when $\psi_\beta = diag[4,4,4]$

$$\hat{\beta} = \begin{bmatrix} -69.25 \\ 35.66 \\ 190.59 \end{bmatrix} \mu g \quad \& \quad \psi_{\hat{\beta}} = \begin{bmatrix} 1.349 & 1.326 & 1.325 \\ 1.326 & 1.350 & 1.324 \\ 1.325 & 1.324 & 1.350 \end{bmatrix} \mu g^2$$

Fig. 9.5.12b: MAP results when $\psi_\beta = diag[4,4,4]$ & $\psi_Y = \psi_{\Delta W}$

Table 9.5.2 - Comparison of Estimated Values & Variances for ψ_Y Diagonal / Non-Diagonal - Now $\psi_\beta \sim \psi_Y$

	Variances $(\mu g)^2$			Values (μg)		
Parameter	Fig. (9.5.12b)	Fig. (9.5.12a)	% Difference	Fig. (9.5.12b)	Fig. (9.5.12a)	Difference
b_1	1.349	2.084	~ 35%	-69.25	-67.52	1.73
b_2	1.350	1.982	~ 32%	35.66	35.08	-0.581
b_3	1.350	1.980	~ 32%	190.59	189.45	-1.14

Comparing with Table 9.5.1, we can see that while there is not much difference in either case between the two pairs of fitted values, there is now a big difference between the variances in the latter case, with a large reduction occurring as a result of ignoring the variance/covariance terms of the buoyancy correction, highlighting the dangers of doing this in cases where the prior and current information are of similar accuracies. Clearly it is good practice–not to mention required by the consistency criteria outlined in this thesis–to always include the full extent of all available information.

For completeness, let us now examine the situation for $\psi_Y \to v.\mathbf{I}$ when we scale the multiplier v over a wide range of values from close to zero to very large. We again obtain sigmoidal-type plots similar to those we have seen already but now of course the lower and upper limits in each case are different. Figs. (9.5.13) to (9.5.17) shown below illustrate the relevant data.

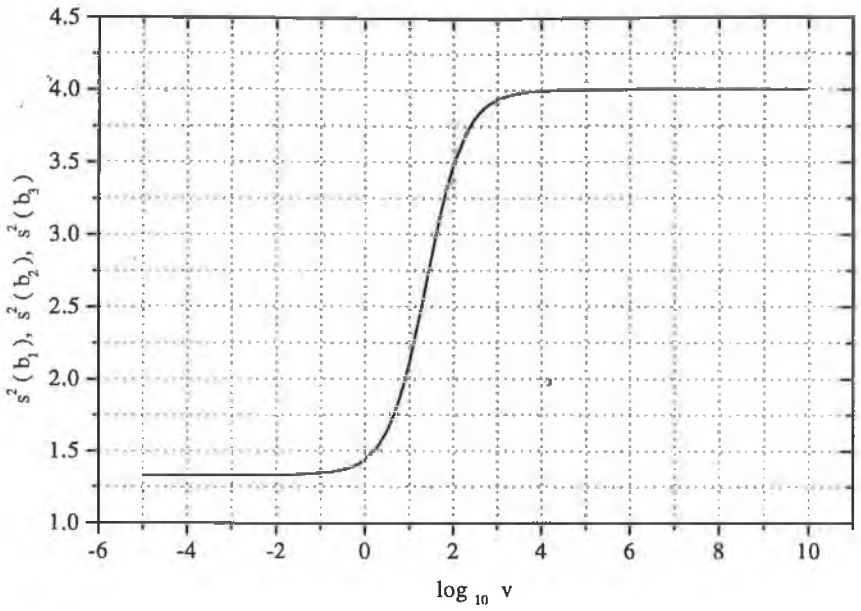


Fig. 9.5.13:Estimated Posterior Variance for b_1, b_2, b_3 when $\psi_\beta = diag\{4,4,4\}$ and $\psi_{\mathbf{Y}} \rightarrow v.\mathbf{I}$

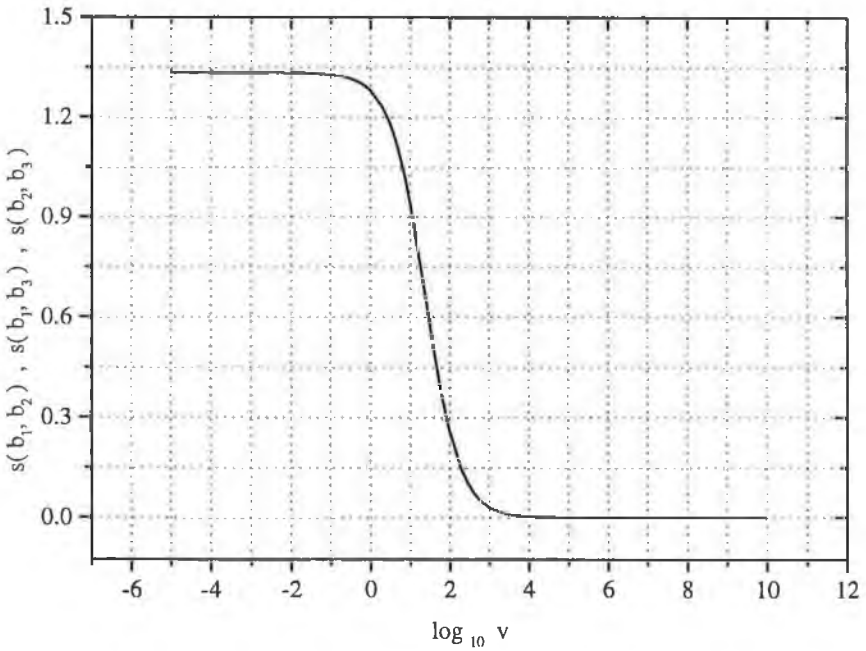


Fig. 9.5.14:Estimated Posterior Covariances for b_1, b_2, b_3 . when $\psi_\beta = diag\{4,4,4\}$ and $\psi_{\mathbf{Y}} \rightarrow v.\mathbf{I}$

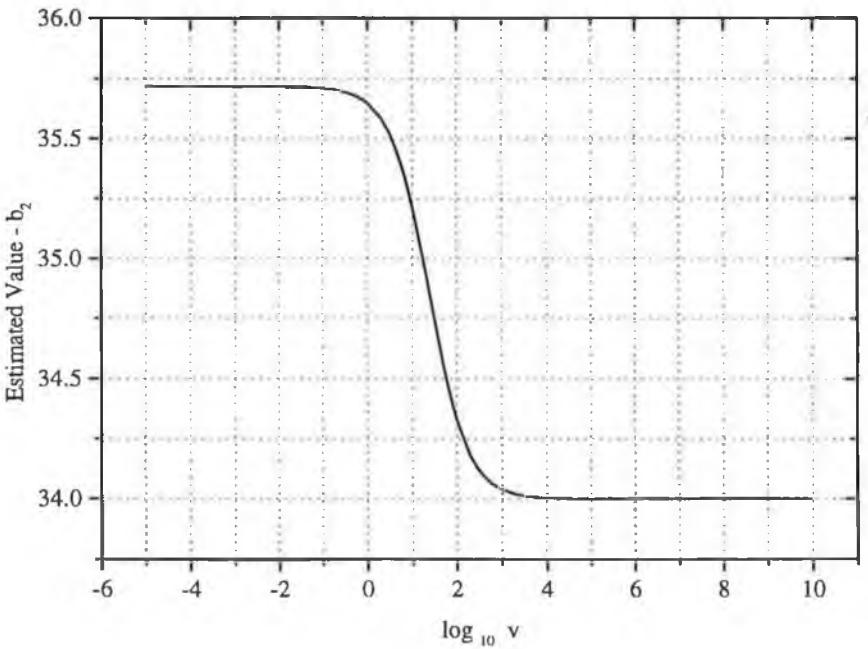


Fig. 9.5.16:Posterior Estimate for b_2, when $\psi_\beta = diag\{4,4,4\}$ and $\Psi_Y \to v.\mathbf{I}$

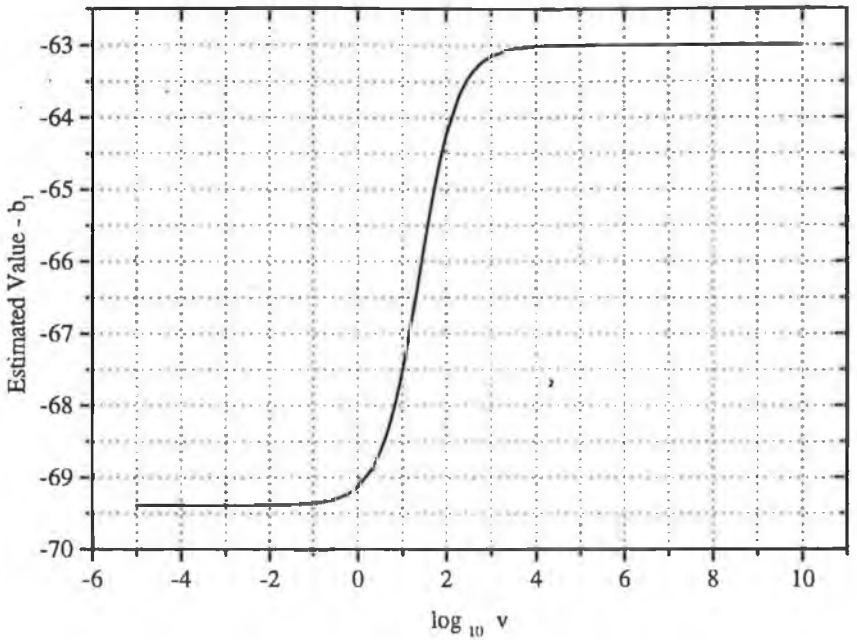


Fig. 9.5.15:Posterior Estimate for b_1, when $\psi_\beta = diag\{4,4,4\}$ and $\psi_Y \rightarrow v.\mathbf{I}$

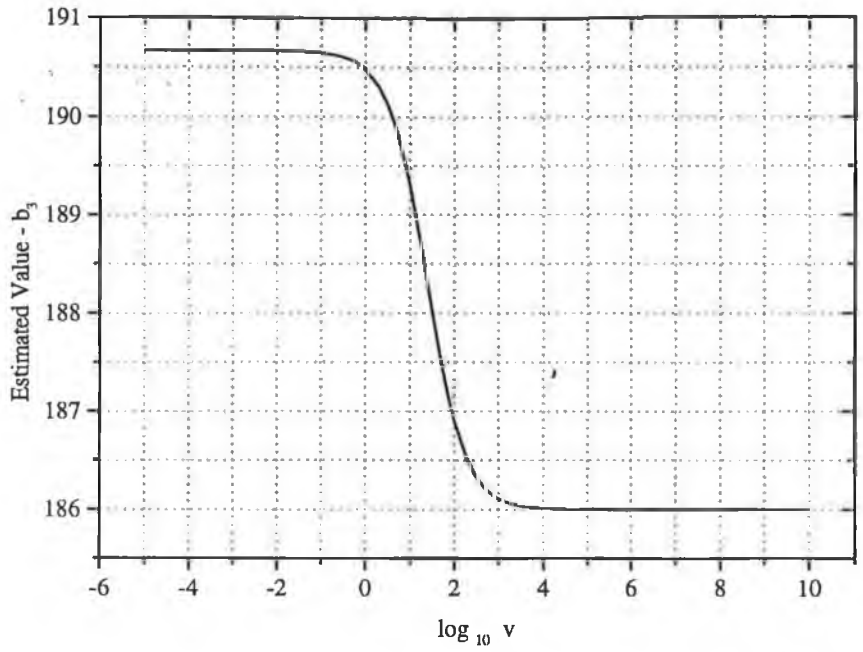


Fig. 9.5.17:Posterior Estimate for b_3, when $\psi_\beta = diag\{4,4,4\}$ and $\psi_Y \to \nu.\mathbf{I}$

It is interesting to look at the residuals in the case of this transformation $\psi_Y \to \nu.\mathbf{I}$. We find large residuals when the current data is considered much less accurate, i.e. as $v \to \infty$ and $\therefore \hat{\beta} \to \mu_\beta$. In such a case we find the residual vector of Fig. (9.5.18).

$$\mathbf{res} = \begin{bmatrix} -7.210 \\ -12.200 \\ -2.599 \\ -7.699 \\ 11.300 \\ 1.899 \end{bmatrix} \mu g$$

Fig. 9.5.18: **res** for $\hat{\beta} \to \mu_\beta$

which is large and implies poor experimental agreement—as we would expect since the current data is having minimal influence now.

On the other hand, as $\nu \to 0$ and $\hat{\beta} \to \begin{bmatrix} -69.38 \\ 35.72 \\ 190.67 \end{bmatrix}$ μg we find the residual vector of Fig. (9.5.19):

$$\mathbf{res} \to \begin{bmatrix} 0.999 \\ -1.150 \\ 0.349 \\ -0.399 \\ 0.250 \\ -1.050 \end{bmatrix} \mu g$$

Fig. 9.5.19: res for $\nu \to 0$

—i.e. the experimental information is now considered maximally accurate and hence exerts greatest influence on the result, leading to good experimental agreement. These latter residuals can probably be considered as a fairly intrinsic estimate of the internal consistency of the data, or the extent of any systematic errors present, since there is minimal influence on them from any other source.

****

10. Further Examples

Included in this chapter are two further case studies involving experimental comparison data, treated by the MAP estimator, with further investigation of its characteristics. In particular we give increased consideration to the idea of inaccurate prior information and note how the estimator copes with this situation. We emphasise once again the role of *relative accuracies* among the various sets of information and show how the estimator can deal with inaccurate information if the accompanying variance/covariance information permits this. We consider the *robustness* of the estimator in terms of the stability of the solutions in the presence of variations in parts of the initial conditions. One likely cause of incorrect prior information is the phenomenon of drift and we investigate how our estimation techniques can respond to this situation. We highlight a fundamental problem in current mass metrology where independent information on drift may be hard to obtain. In this respect the estimator must be treated realistically in the light of the available physical information.

10.1 Example II

We now take a calibration example involving eight parameters and ten observations. The available prior information on all the parameters is as given in Table 10.1.1 below. The prior deviations-from-nominal, standard deviations and volumes are taken directly from the available calibration certificates, while the volume standard deviations are obtained from an assumed density uncertainty of $\pm 2 kg.m^{-3}$, taken to be uniformly distributed.

Table 10.1.1: Prior Information

Parameter	Nominal Value (g)	Prior Value (deviation) (mg)	Std. Dev (mg)	Volume (cm^3)	Vol. Std. Dev (cm^3)
b_1	1000.0	2.0	0.25	125.9763	0.0188
b_2	1000.0	0.9	0.75	119.0	0.01635
b_3	500.0	-0.9	0.125	62.99	0.00916
b_4	500.0	0.1	0.375	59.52	0.008179
b_5	200.0	0.45	0.05	25.20	0.0045
b_6	200.0	0.04	0.15	23.81	0.003271
b_7	100.0	1.07	0.05	12.74	0.00187
b_8	100.0	-0.49	0.075	11.90	0.0016

The prior vectors and matrices, μ_β, ψ_β, ψ_V & **V** can easily be constructed from this information. The supplied prior information does not include data on possible covariances among the parameters so we must assume that ψ_β is diagonal under these circumstances. The Design Matrix **X** is given in Fig. (10.1.1) below. The experimental data is given in Table (10.1.2), from which we can construct the observation vector **Y** and also $\psi_{\Delta W}$ and ρ_a. In Table (10.1.2) the data $\Delta\overline{W}$ and ρ_a are mean values from 6 experimental measurements. The standard deviations in column 2 are thus those of mean values of 6 measurements. This data is used directly to compute $\psi_{\Delta W}$ as shown in Fig. (10.1.2). The calibration data for the environment monitoring instruments used in this case leads to the following standard uncertainties:

$$s(t) = 0 \cdot 1^{\circ}\mathrm{C}$$
$$s(P) = 1 \text{ Torr} = 133.3 \text{ Pa}$$
$$s(h) = 5\%$$

Thus using Eq. (3.2.4) we evaluate the air density variance to be:

$$s^2(\rho_a) \cong 3 \times 10^{-6} \left(mg.cm^{-3}\right)^2$$

$$\mathbf{X} = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & -1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & -1 & -1 & -1 & 0 \\ 0 & 0 & 0 & 1 & -1 & -1 & 0 & -1 \\ 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & -1 & -1 \\ 0 & 0 & 0 & 0 & 0 & 1 & -1 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 \end{bmatrix}$$

Fig. 10.1.1: Design Matrix

Once again we evaluate the components of ψ_Y using Eq. (4.5.17):

$$\psi_Y = \psi_{\Delta w} + diag\{\mathbf{XV}\}\psi_\rho diag\{\mathbf{XV}\} + \rho\mathbf{X}\psi_V\mathbf{X}^T\rho^T$$

The first of these, $\psi_{\Delta W}$, we have already referred to, while the other two terms are shown in Figs. (10.1.3) and (10.1.4).The term $\mathbf{t_2}$ is the variance/covariance contribution due to ψ_ρ, while $\mathbf{t_3}$ is that due to ψ_V. The total experimental covariance matrix ψ_Y is given in Fig. (10.1.5).

Table 10.1.2 : Experimental Information

$\Delta\overline{W}$ (mg)	Std. Dev (mg)	ρ_a (mg.cm^{-3})	ΔV (cm^3)	$\mathbf{Y} \equiv \Delta\overline{m}$ (mg)
-7.18333	0.044096	1.1814	6.9763	1.0660
-1.29647	0.031623	1.1818	3.4663	2.8000
5.86162	0.030732	1.1809	-3.51	1.7166
-5.27272	0.025	1.1809	3.47	-1.1750
-3.8867	0.058333	1.1896	1.24	-2.4116
1.84465	0.008333	1.1892	-1.39	0.1916
-1.35358	0.021972	1.2244	1.39	0.3483
-0.851816	0.022161	1.2236	0.56	-0.1660
0.481667	0.020777	1.2249	-0.83	-0.5350
0.526504	0.008851	1.2244	0.84	1.5550

$$\psi_{\Delta w} = \begin{bmatrix} 19.444 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ & 10.0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ & & 9.444 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ & & & 6.25 & 0 & 0 & 0 & 0 & 0 & 0 \\ & & & & 34.027 & 0 & 0 & 0 & 0 & 0 \\ & & & & & 0.694 & 0 & 0 & 0 & 0 \\ & & & & & & 4.827 & 0 & 0 & 0 \\ & & & \text{symm} & & & & 4.911 & 0 & 0 \\ & & & & & & & & 4.316 & 0 \\ & & & & & & & & & 0.783 \end{bmatrix}$$

Fig. 10.1.2: $\psi_{\Delta w} \times 10^{-4}$ mg^2

$$\mathbf{t_2} = \begin{bmatrix} 1.46006 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ & 0.360457 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ & & 0.369603 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ & & & 0.361227 & 0 & 0 & 0 & 0 & 0 & 0 \\ & & & & 0.046128 & 0 & 0 & 0 & 0 & 0 \\ & & & & & 0.057963 & 0 & 0 & 0 & 0 \\ & & & & & & 0.057963 & 0 & 0 & 0 \\ & & \text{symm} & & & & & 0.009408 & 0 & 0 \\ & & & & & & & & 0.020667 & 0 \\ & & & & & & & & & 0.021168 \end{bmatrix}$$

Fig. 10.1.3: $\mathbf{t_2}$ $\left(\equiv diag\{\mathbf{X.V}\}.\psi_{\rho_a}.diag\{\mathbf{X.V}\}\right) \times 10^{-4}$ mg^2

$$\mathbf{t}_3 = \begin{bmatrix} 8.4189 & 4.68837 & -3.73054 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ & 6.7961 & 2.10455 & -0.23725 & -1.17952 & -0.94021 & 0 & 0 & 0 & 0 \\ & & 5.8319 & -0.237069 & -1.17863 & -0.939494 & 0 & 0 & 0 & 0 \\ & & & 2.10294 & 1.17863 & -0.939494 & 0 & 0 & 0 & 0 \\ & & & & 1.57789 & 0.340936 & -0.03933 & -0.144104 & -0.104914 & -0.050979 \\ & & & & & 1.32369 & -0.03931 & -0.157151 & -0.117989 & 0.037857 \\ & & & & & & 0.361296 & 0.200756 & -0.160475 & 0 \\ & & & \text{symm} & & & & 0.291953 & 0.091426 & -0.013484 \\ & & & & & & & & 0.252064 & -0.013498 \\ & & & & & & & & & 0.091448 \end{bmatrix}$$

Fig. 10.1.4: $\mathbf{t}_3 \left(\equiv \rho.\mathbf{X}.\psi_v.\mathbf{X}^T.\rho\right) \times 10^{-4}$ mg^2

$$\psi_Y = \begin{bmatrix} 29.3123 & 4.68837 & -3.73054 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ & 17.1566 & 2.10455 & -0.23725 & -1.17952 & -0.94021 & 0 & 0 & 0 & 0 \\ & & 15.6348 & -0.237069 & -1.17863 & -0.939494 & 0 & 0 & 0 & 0 \\ & & & 8.71417 & 1.17863 & -0.939494 & 0 & 0 & 0 & 0 \\ & & & & 35.6507 & 0.340936 & -0.03933 & -0.144104 & -0.104914 & -0.050979 \\ & & & & & 2.07498 & -0.03931 & -0.157151 & -0.117989 & 0.037857 \\ & & & & & & 5.24593 & 0.200756 & -0.160475 & 0 \\ & & & \text{symm} & & & & 5.21136 & 0.091426 & -0.013484 \\ & & & & & & & & 4.5894 & -0.013498 \\ & & & & & & & & & 0.89595 \end{bmatrix}$$

Fig. 10.1.5: $\psi_Y \times 10^{-4}$ mg^2

Using this data, we now apply the MAP Estimator of Eqs. (8.3.15) & (8.4.5) to find updated parameter estimates as given in Table 10.1.3 below. This table shows the prior data, posterior estimated values, the Difference Vector between these two, the prior and posterior standard deviations, and the combined standard deviation of the Difference Vector. Fig. (10.1.6) gives the estimated covariance matrix while experimental residuals are shown in Table 10.1.4. As can be seen, the posterior estimates have lower standard deviations than the prior information. Of interest is the combined standard deviation of the Difference Vector, $s\left(\mu_\beta - \hat{\beta}\right)$ (last column of Table 10.1.3): this is at all points larger than the respective elements of the difference vector (3rd column of Table 10.1.3). This is an important point, and one we would expect: since each value is just an estimate based on whatever information is available at the time, each should be subject to updating. However, since the standard deviation is taken as a measure of the possible dispersion in the values, one would expect that any updated value should lie within this bound. Note: this does assume, however, that the measurand is in fact constant over time; in the case of mass standards, drift *is* possible and indeed observed (see Girard, (1994) for example, also Sutton & Clarkson (1993/94), while Davis (1990) provides a detailed discussion of the stability of

Reference and Working Standards in one situation). Also, Bich (1992), (1993b) has pointed out that this drift can be modelled via a Kalman Filter type approach, but that is a different issue, and although it can easily be built into our model, for the present we will make the assumption that the measurands/parameters are constant in time. (For applications of the Kalman Filter and MLE to atomic clock parameterisation see Tyron & Jones, (1983) & Jones & Tyron (1983)). In the following section of this chapter we will make further comment on drift and how we might deal with it.

The standards were not calibrated as a group before, indeed they are taken from three different sets of quite different densities, and so the *prior* standard deviations of nominally equally pairs (e.g. b_1 & b_2 or b_3 & b_4) are quite different. However, after the analysis, the estimated standard deviations are much more uniform.

Table 10.1.3: Comparison of Prior , Posterior Data, after carrying out MAP Estimation (data in *mg*)

μ_β	$\hat{\beta}$	$\left(\mu_\beta - \hat{\beta}\right)$	$diag\left(\psi_{\hat{\beta}}\right)^{1/2}$	$diag\left(\psi_\beta\right)^{1/2}$	$s\left(\mu_\beta - \hat{\beta}\right)$
2.0	2.08008	-0.0800812	0.133485	0.25	0.283405
0.9	1.00608	-0.10608	0.134876	0.75	0.762031
-0.9	-0.934075	0.0340754	0.0673833	0.125	0.142005
0.1	0.220219	-0.120219	0.0671767	0.375	0.380969
0.45	0.429602	0.0203979	0.0276786	0.05	0.0571498
0.04	0.072579	-0.032579	0.0291039	0.15	0.152797
1.07	1.077	-0.0070027	0.0153515	0.05	0.0523036
-0.49	-0.475644	-0.0143557	0.0153016	0.075	0.076545

$$\psi_{\hat{\beta}} = \begin{bmatrix} 178.155 & 170.768 & 84.3603 & 84.2958 & 32.3125 & 34.3595 & 16.4158 & 16.8544 \\ & 181.926 & 85.5725 & 85.507 & 32.7766 & 34.853 & 16.6516 & 17.0965 \\ & & 45.3896 & 41.4927 & 15.9388 & 16.9485 & 8.12745 & 8.28316 \\ & & & 45.1132 & 16.8989 & 17.9694 & 8.55494 & 8.84545 \\ & & & & 7.6588 & 6.34044 & 3.22181 & 3.18763 \\ & & & & & 8.46778 & 3.46465 & 3.4287 \\ & & \text{symm} & & & & 2.36719 & 1.91358 \\ & & & & & & & 2.32835 \end{bmatrix}$$

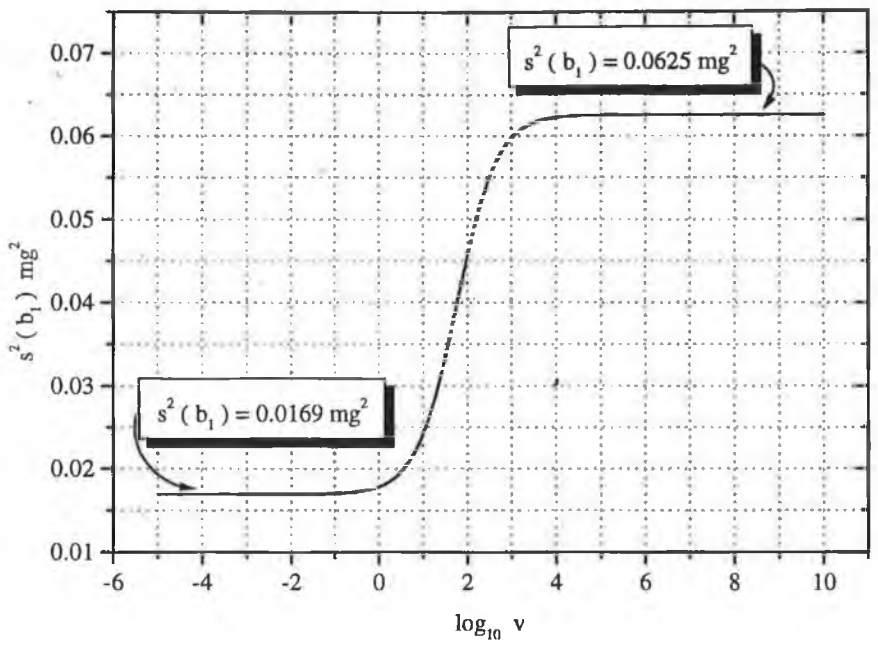
Fig. 10.1.6: Estimated Covariance Matrix $\psi_{\hat{\beta}}$ $(mg^2 \times 10^{-4})$

Table 10.1.4: Observations, Fitted Observations, Residuals and Measurement Std. Dev. (*mg*)

$\mathbf{Y}$	$\hat{\mathbf{Y}}$	$\left(\mathbf{Y}-\hat{\mathbf{Y}}\right)$	$diag(\psi_{\mathbf{Y}})^{1/2}$
1.0666	1.074	-0.00740112	0.05415
2.8	2.79394	0.00606211	0.04142
1.71666	1.71994	-0.00327677	0.03954
-1.175	-1.15429	-0.0207059	0.02952
-2.4166	-2.51326	0.101659	0.05971
0.191666	0.193682	-0.00201586	0.01440
0.348333	0.357023	-0.00869009	0.02291
-0.1666	-0.171756	0.00515636	0.02283
-0.535	-0.528779	-0.00622055	0.02078
1.555	1.55265	0.002353	0.08850

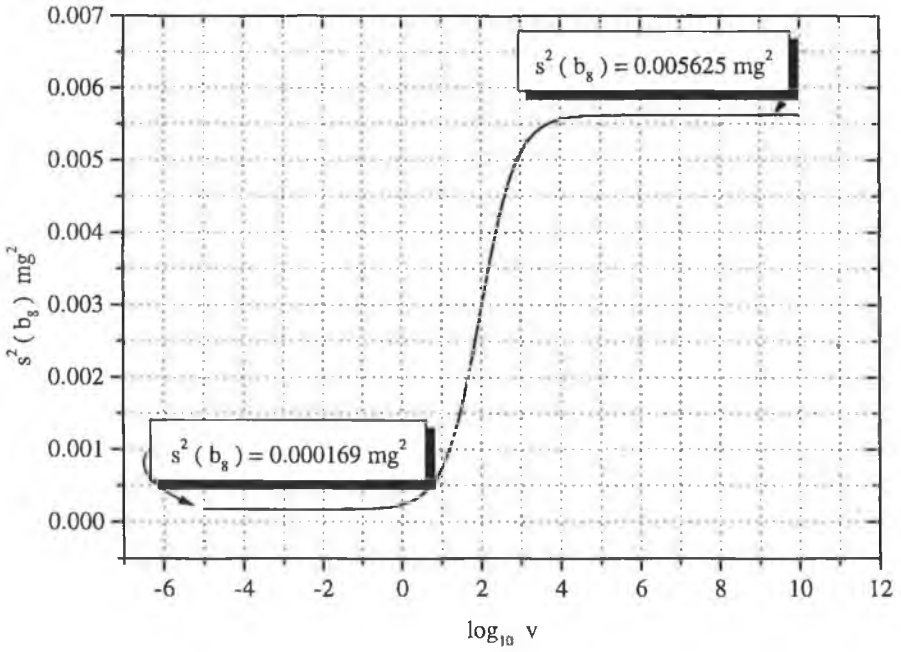
10.2 Analysis of the Estimator's Capability

We now consider the effects of making the transformation $\psi_{\mathbf{Y}} \rightarrow \nu.\psi_{\mathbf{Y}}$ where ν is a scalar multiplier which we can vary in the range $(0,\infty)$. This is just a computationally convenient means of varying the influence of the current data between the two extremes of near-total and near-zero control over the posterior estimates, as explained in the last chapter. Doing this allows us to see what type of adjustment to the prior data is possible and to compare that with what has been achieved with the given data. We will see how this can sometimes highlight problems in the data that might not otherwise surface. The graphs shown below are of the same general sigmoidal type as those obtained in the analysis of Chapter 9, as we might expect, illustrating lower and upper bounds in each case. Figs. (10.2.1) & (10.2.2) give the posterior variance estimate for just b_1 & b_8 by way of example; while Figs (10.2.3). & (10.2.4). give the posterior covariance matrix, $\psi_{\hat{\beta}}$ for the two cases, $\nu \rightarrow 0$ & $\nu \rightarrow \infty$, respectively.



$\nu \to 0$ $\nu \to \infty$

Fig. 10.2.1: Variation in Estimated Variance of b_1 for $\Psi_Y \to \nu.\Psi_Y$



$\nu \to 0$ $\nu \to \infty$

Fig. 10.2.2: Variation in Estimated Variance of b_g for $\Psi_Y \to \nu.\Psi_Y$

$$\Psi_{\hat{\beta}} = \begin{bmatrix} 169.173 & 169.173 & 84.5865 & 84.5865 & 33.8346 & 33.8346 & 16.9173 & 16.9173 \\ & 169.173 & 84.5865 & 84.5865 & 33.8346 & 33.8346 & 16.9173 & 16.9173 \\ & & 42.2932 & 42.2932 & 16.9173 & 16.9173 & 8.4856 & 8.4856 \\ & & & 42.2932 & 16.9173 & 16.9173 & 8.4856 & 8.4856 \\ & & & & 6.76692 & 6.76692 & 3.38346 & 3.38346 \\ & & symm & & & 6.76692 & 3.38346 & 3.38346 \\ & & & & & & 1.69173 & 1.69173 \\ & & & & & & & 1.69173 \end{bmatrix}$$

Fig. 10.2.3: Covariance Matrix, $\Psi_{\hat{\beta}}$, when $\nu \to 0$ for the transformation $\Psi_{\mathbf{Y}} \to \nu.\Psi_{\mathbf{Y}}$

($mg^2 \times 10^{-4}$)

$$\Psi_{\hat{\beta}} = \begin{bmatrix} 625 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ & 5625 & 0 & 0 & 0 & 0 & 0 & 0 \\ & & 156.25 & 0 & 0 & 0 & 0 & 0 \\ & & & 1406.25 & 0 & 0 & 0 & 0 \\ & & & & 25 & 0 & 0 & 0 \\ & & symm & & & 225 & 0 & 0 \\ & & & & & & 25 & 0 \\ & & & & & & & 56.25 \end{bmatrix}$$

Fig. 10.2.4: Covariance Matrix, $\Psi_{\hat{\beta}}$, when $\nu \to \infty$ for the transformation

$\Psi_{\mathbf{Y}} \to \nu.\Psi_{\mathbf{Y}}$ ($mg^2 \times 10^{-4}$)

Fig. (10.2.3) is the lower limit and presents some interesting features: from it we can quickly calculate that the following hold:

$$\frac{s^2(b_i)}{s^2(b_j)} = \left(\frac{m_i}{m_j}\right)^2 ; \quad \frac{s^2(b_i)}{s(b_i,b_j)} = \frac{m_i}{m_j} \quad \& \quad \frac{s(b_i,b_j)}{s(b_i,b_k)} = \frac{m_j}{m_k}$$

—where m_i, m_j, m_k are the nominal mass values of the i^{th}, j^{th}, k^{th} parameters. Such relationships often appear when doing a "simple" fitting with just one piece of prior information [or using RLS], when a fraction of the 'reference' uncertainty equal to the ratio of nominal masses is always transmitted to the various standards. Using the MAP estimator with full prior information, this does not automatically manifest itself unless it was already the case in the initial conditions; although an examination of Fig. (10.1.6) shows that the posterior covariance matrix for the 'real' data of this experiment does approximate these relationships. However, Fig. (10.2.3) shows that in the limit, this estimator will converge to exactly this situation to which other estimators are tied. In Fig. (10.2.4) we see how the prior variance/covariance information remains unchanged when $\nu \to \infty$ and the current information is effectively removed.

Figs (10.2.5) & (10.2.6) below illustrate the range of possible values that can be assumed by the parameters b_1 & b_8, under the same conditions as those discussed

above (i.e. $\psi_Y \rightarrow v.\psi_Y$ with $v \in (0,\infty)$). Fig. (10.2.7) shows the complete parameter vector in the two limiting cases. From this information we can see the maximum amount of adjustment to the parameters that is possible with the given initial conditions.

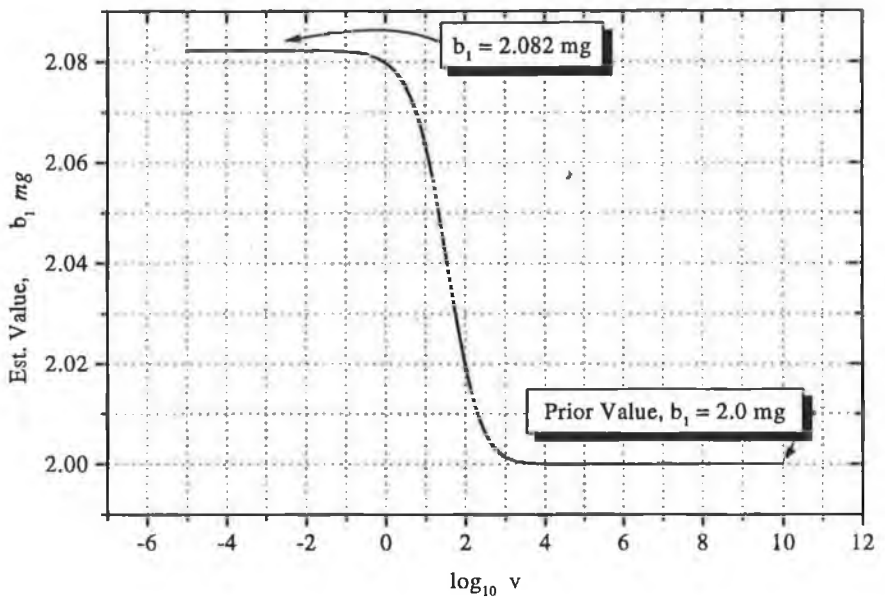


Fig. 10.2.5: Estimated Value for b_1 with v in range $(0,\infty)$

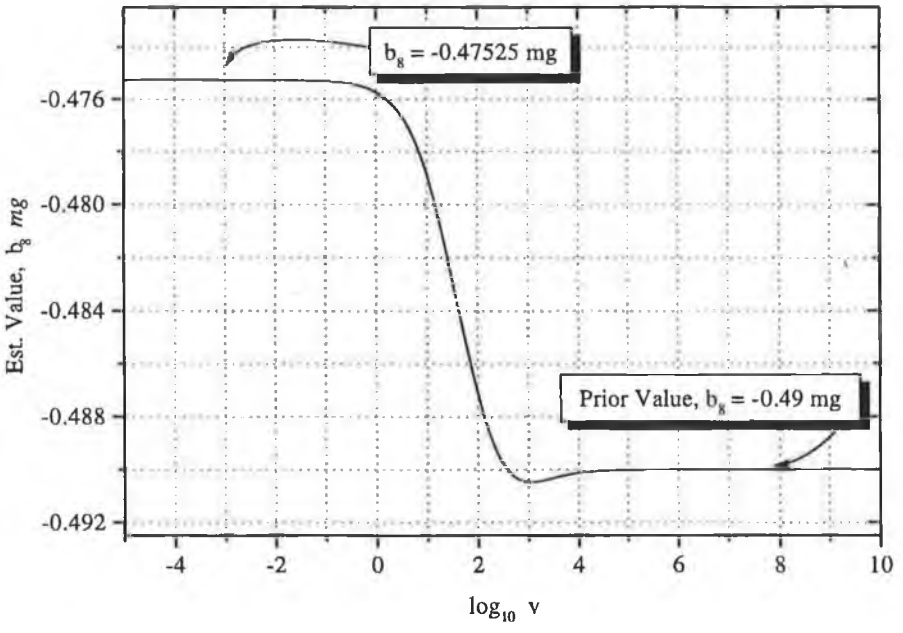


Fig. 10.2.6: Estimated Value for b_8 with v in range $(0,\infty)$

$$\hat{\beta} = \begin{bmatrix} 2.082 \\ 1.007 \\ -0.934 \\ 0.221 \\ 0.429 \\ 0.073 \\ 1.077 \\ -0.475 \end{bmatrix} \text{ mg ;} \qquad \hat{\beta} = \begin{bmatrix} 2.0 \\ 0.9 \\ -0.9 \\ 0.1 \\ 0.45 \\ 0.04 \\ 1.07 \\ -0.49 \end{bmatrix} \text{ mg.}$$

(i) (ii)

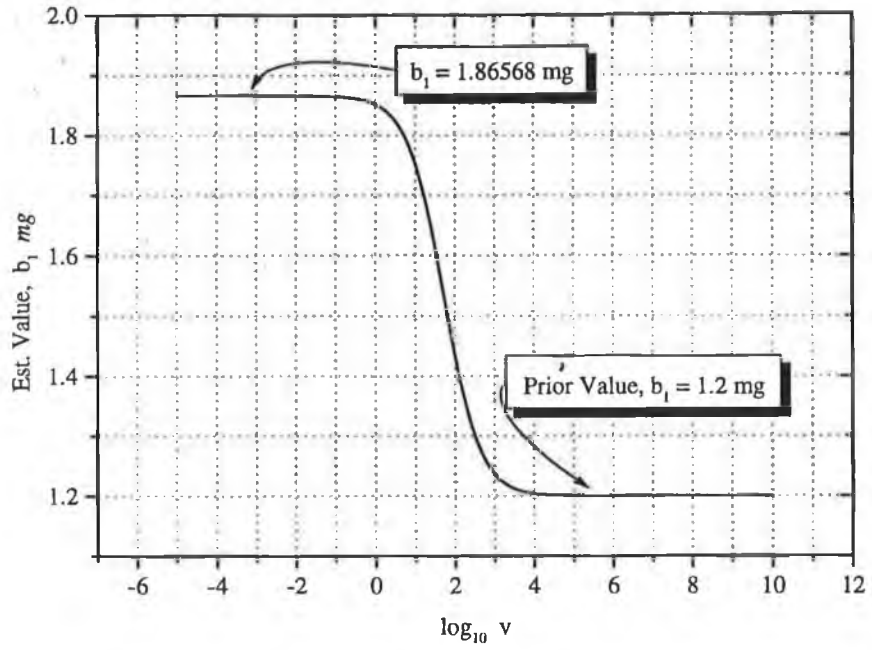
Fig. 10.2.7: Estimated Parameter Vector, $\hat{\beta}$

Next we consider the performance of the MAP estimator in the case of errors in the parameters in order to examine its robustness. In Table 10.2.1 below we simulate an error on the prior knowledge of b_1 so that its value is $1.2mg$ rather than $2.0mg$. We can see that in the updated data, b_1 is adjusted significantly more than the others, and indeed more than its prior, posterior or combined difference standard deviation. This of course suggests a systematic error and can easily be interpreted as such since none of the other parameters are adjusted so much, suggesting that it is b_1 (prior) which was in error. Aside: Should this situation have occurred it may perhaps have been due to *drift*, meaning that the prior value of $1.2mg$ is no longer a good representation of the value of b_1. We will discuss this situation a little more in Section 10.3.

Table 10.2.1: Estimated Parameter Values (b_1 prior in error) (data in *mg*)

μ_β	$\hat{\beta}$	$(\mu_\beta - \hat{\beta})$	$diag(\psi_{\hat{\beta}})^{1/2}$	$diag(\psi_\beta)^{1/2}$	$s(\mu_\beta - \hat{\beta})$
1.2	1.85201	**-0.625006**	0.133485	0.25	**0.283405**
0.9	0.787489	0.112511	0.134876	0.75	0.762031
-0.9	-1.04206	0.142061	0.0673833	0.125	0.142005
0.1	0.112318	-0.0123183	0.0671767	0.375	0.380969
0.45	0.388251	0.0617486	0.0276786	0.05	0.0571498
0.04	0.0286084	0.0113916	0.0291039	0.15	0.152797
1.07	1.05598	0.0140163	0.0153515	0.05	0.0523036
-0.49	-0.497208	0.00720808	0.0153016	0.075	0.076545

Fig. (10.2.8) shows the range of values b_1 can be assigned should $\psi_Y \rightarrow v.\psi_Y$ and v be scaled as before. From this we can see that the best adjustment that can be made to b_1, when $v \rightarrow 0$ and the current information is exerting maximum influence, leads to an estimated value of 1.86568 *mg*, quite close to the "correct" prior value! Thus our data in Table (10.2.1) is quite close to the theoretically best value.



$v \to 0$ $\qquad\qquad$ $v \to \infty$

Fig. 10.2.8: $\hat{b}_1$ for $b_1 = 1.2mg$ & $\psi_Y \to v.\psi_Y$

Table 10.2.2: Fitted Observations and residuals for Table 10.2.1 Parameter Values (data in *mg*)

Y	$\hat{Y}$	res	$s(Y)$	$s(\Delta W)$
1.066	1.06452	0.00208232	0.054151	0.0440957
2.8	2.78175	0.0182513	0.041499	0.0316228
1.71666	1.71723	-0.000571033	0.0394783	0.0307316
-1.175	-1.15438	-0.020621	0.0296299	0.025
-2.4116	-2.5149	0.103304	0.0596984	0.0583333
0.191666	0.192667	-0.00100065	0.014548	0.00833267
0.348333	0.359643	-0.01131	0.0229062	0.021972
-0.1666	-0.170524	0.00392426	0.0228306	0.0221608
-0.535	-0.530167	-0.00483272	0.0214229	0.0207766
1.555	1.55319	0.00180818	0.00946546	0.00885061

In Table (10.2.2) above the posterior estimated observations and residuals are presented, from which we can see that the agreement with the experimental data is still very good. In understanding this, it is helpful to note that the standard deviations

of the prior data are about an order of magnitude bigger than those of the experimental information (compare column 5 of Table 10.2.1 with column 4 of Table 10.2.2). Thus the experimental information is influencing the results more than the prior information, leading to greater robustness in the face of possible errors in the prior information.

So let us now reduce the prior standard deviations by a factor of 10—i.e. reduce the variances by a factor of 100. The new posterior results in Table 10.2.3 now indicate that b_1 is adjusted much less but the difference between prior and posterior values is now outside the combined standard deviations in all cases. Furthermore, in Table 10.2.4 the corresponding residuals are some one to two orders of magnitude *larger* than they were before—so in fact we do not have a good fit to the experimental data.

Table 10.2.3: Estimated Parameter Values (b_1 prior in error and smaller prior Std. Devs (data in *mg*)

μ_β	$\hat{\beta}$	$\left(\mu_\beta - \hat{\beta}\right)$	$diag\left(\Psi_{\hat{\beta}}\right)^{1/2}$	$diag\left(\Psi_\beta\right)^{1/2}$	$s\left(\mu_\beta - \hat{\beta}\right)$
1.2	1.4347	**-0.234701**	0.021379	0.025	**0.0328947**
0.9	0.756997	0.143003	0.0334731	0.075	0.0821307
-0.9	-0.962738	0.062738	0.0110564	0.0125	0.0166882
0.1	0.109827	-0.00982681	0.0154644	0.0375	0.0405635
0.45	0.437496	0.0125036	0.00467686	0.005	0.0068463
0.04	0.0131579	0.0268421	0.00975751	0.015	0.0178944
1.07	1.06641	0.00359329	0.00430808	0.005	0.0065999
-0.49	-0.498196	0.00819559	0.00542494	0.0075	0.0092563

Table 10.2.4: Fitted Observations and residuals for Table 10.2.3 Parameter Values (data in *mg*)

Y	$\hat{Y}$	res	$s(Y)$	$s(\Delta W)$
1.066	0.677705	0.388895	0.054151	0.0440957
2.8	2.28761	0.512388	0.041499	0.0316228
1.71666	1.60991	0.106752	0.0394783	0.0307316
-1.175	-1.07256	-0.102435	0.0296299	0.025
-2.4116	-2.4798	0.068199	0.0596984	0.0583333
0.191666	0.157368	0.0342979	0.014548	0.00833267
0.348333	0.424339	-0.0760056	0.0229062	0.021972
-0.1666	-0.130715	-0.0358853	0.0228306	0.0221608
-0.535	-0.555053	0.0200533	0.0214229	0.0207766
1.555	1.5646	-0.0096023	0.00946546	0.00885061

This tells us that what we learned from the experiment disagrees with our prior knowledge but we may not automatically know which is in error. Should the prior knowledge have been subject to drift, this should have been accounted for in μ_β & ψ_β. As mentioned before, good quantitative information on drift in mass standards is difficult to acquire, but where information is available it should be taken into account. (Aside: Again we refer to a further discussion of drift in Section 10.3.)

In fact, possibilities like drift would need a hard look should the example just cited occur in real data: if the prior information is "definitely" reliable and the experimental procedure well understood and invariably performs satisfactorily, the possibility of physical change to the artefacts would need considering. This is not to rule out the possibility of problems in either the measuring method or the prior information not heretofore imagined! (For example drift *during* the measurement can be a problem—See Sutton & Clarkson (1993/94). Surface contamination via adsorption is relevant in the context of drift and has been discussed widely in the literature, for example Cumpson & Seah (1994), (1994/95), Kochsiek (1982), Seah et al (1994), Schwartz (1994a), (1994b), Schwartz & Glaeser (1994c). Effects due to cleaning the standards to overcome this drift are discussed by Pinot (1994/95) and Pinot (1997); cleaning of standards is also discussed by Girard (1990). In Cumpson & Seah (1996) surface contamination and cleaning of platinum-iridium standards is considered in detail.

On the other hand, let us now suppose that the prior value of b_1 is in fact a little suspect. In that case we may have $s(b_1) = 0.5mg$, the others remaining as originally given. With this situation, Table 10.2.5 results where we can see b_1 adjusted by a large amount, indeed it returns very close to the original estimate in Table 10.1.3, while in Table 10.2.6 we see the new residuals are now much smaller again.

Table 10.2.5: Estimated Parameter Values (b_1 prior in error but with larger Std. Dev.) (data in *mg*)

μ_β	$\hat{\beta}$	$(\mu_\beta - \hat{\beta})$	$diag(\psi_{\hat{\beta}})^{1/2}$	$diag(\psi_\beta)^{1/2}$	$s(\mu_\beta - \hat{\beta})$
1.2	2.02933	**-0.829335**	0.150547	0.5	**0.522173**
0.9	0.957444	-0.0574438	0.150476	0.75	0.764946
-0.9	-0.958102	0.058102	0.0750124	0.125	0.14578
0.1	0.196211	-0.096211	0.0748154	0.375	0.38239
0.45	0.420402	0.0295984	0.0304204	0.05	0.0585269
0.04	0.0627956	-0.0227956	0.0320493	0.15	0.153386
1.07	1.07233	-0.00232604	0.0166381	0.05	0.0526956
-0.49	-0.480442	-0.00955782	0.0166571	0.075	0.0768275

Table 10.2.6 : Fitted Observations and residuals for Table 10.2.5 Parameter Values (data in *mg*)

Y	$\hat{Y}$	res	$s(Y)$	$s(\Delta W)$
1.066	1.07189	-0.00529107	0.054151	0.0440957
2.8	2.79123	0.00877418	0.041499	0.0316228
1.71666	1.71933	-0.00267475	0.0394783	0.0307316
-1.175	-1.15431	-0.020687	0.0296299	0.025
-2.4116	-2.51363	0.102025	0.0596984	0.0583333
0.191666	0.193456	-0.00178998	0.014548	0.00833267
0.348333	0.357606	-0.00927303	0.0229062	0.021972
-0.1666	-0.171482	0.00488222	0.0228306	0.0221608
-0.535	-0.529088	-0.00591176	0.0214229	0.0207766
1.555	1.55277	0.00223178	0.00946546	0.00885061

So the robustness of MAP has once again been illustrated, although we have seen that under some circumstances it can only point to problems/missing information without removing the difficulty. RLS by contrast could not do any of this as it treats the constraints as fixed and so can only adjust whatever is left.

10.3 More on the Influence of the Prior Information

In our discussions in this chapter we have several times mentioned the problem of *drift* on mass standards. Essentially, we need to know if we are estimating a dynamic quantity or a static one. We should note that this is a separate question to the issue of whether our estimates are stochastic or deterministic, since we are now thinking of the underlying measurand, rather than our estimate of it. If the Difference Vector, $\left(\mu_\beta - \hat{\beta}\right)$, is substantially larger than the combined standard deviation of the difference, $s\left(\mu_\beta - \hat{\beta}\right)$, we are led to the conclusion that the prior information was in poor agreement with the new experimental information. In the methods described in this thesis, the prior information certainly plays an important role and can influence the posterior estimates to varying degrees, depending on relative accuracies, as has been explored in detail already. We have also considered possible errors in the prior information in this context in order to probe their influence on the results, and we have seen that in many cases the estimator is very robust. However, if the standards have physically changed in the calibration interval, so that the prior values no longer

properly describe the true situation, how can we be sure the posterior estimates will not be detrimentally influenced by the prior information? It is this possibility we now wish to investigate.

Drift on mass standards is generally a result of surface contamination and so will be a function of surface area. It is easy to see that surface area, A is related to mass m, by $A \propto m^{2/3}$ for a standard having cylindrical geometry. For example, for a minimum surface area cylinder, for which height is equal to diameter, we find A proportional to m and density ρ according to:

$$A = \left(\frac{3\pi}{2}\right)\left(\frac{4m}{\pi\rho}\right)^{\frac{2}{3}} \qquad (10.3.1)$$

while for standard OIML-classified shapes (OIML 1994)) which are geometrically more complex, a quick calculation shows that approximately the same proportionality with mass remains. Thus a 500g standard will have $0.5^{2/3}$ the surface area of a 1000g standard, and might be expected to suffer $0.5^{2/3}$ of the contamination experienced by the 1000g standard. In the previous section we considered an error of *0.8 mg* on parameter b_1. In order to generate some synthetic data for analysis we will now suppose this is due to drift—i.e. the prior value of *2.0mg* is in fact updated to take account of drift since the last calibration, while a value of *1.2mg* would be used if no drift error was suspected. Assuming the other standards to have been affected to the same degree, will lead to the "drift error" of Table 10.3.1 below, where the $(\text{mass ratio})^{2/3}$ approximate proportionality mentioned above has been used.

Table 10.3.1: Simulated Drift Error

Parameter Nominal Value	Drift Error
1000g	+0.8mg
500g	+0.5mg
200g	+0.27mg
100g	+0.17mg

If we assume that the prior data of Table 10.1.1 is indeed correct at the time of measurement, then to simulate an error due to mass-additive drift (typical for seldom-used Reference standards. Frequently used Working Standards on the other hand would probably drift downwards due to wear.), we need to subtract the drift error of Table 10.3.1 above from the prior information, leading to the prior and estimated values of Table 10.3.2 below, where ψ_γ and ψ_β, have not been changed from the original values. The estimated observations and residuals are shown in Table 10.3.3.

Table 10.3.2: Comparison of Prior and Posterior Data, after carrying out MAP Estimation (data in *mg*)

μ_β	$\hat{\beta}$	$\left(\mu_\beta - \hat{\beta}\right)$	$diag\left(\psi_{\hat{\beta}}\right)^{1/2}$	$diag\left(\psi_\beta\right)^{1/2}$	$s\left(\mu_\beta - \hat{\beta}\right)$
1.2	0.980735	0.219265	0.133485	0.25	0.283405
0.1	-0.0973078	0.197308	0.134876	0.75	0.762031
-1.4	-1.48421	-0.0842149	0.0673833	0.125	0.142005
-0.3	-0.334454	0.034454	0.0671767	0.375	0.380969
0.18	0.206088	-0.0260877	0.0276786	0.05	0.0571498
-0.23	-0.148253	-0.0817469	0.0291039	0.15	0.152797
0.9	0.964285	-0.0642848	0.0153515	0.05	0.0523036
-0.66	-0.587362	-0.0726381	0.0153016	0.075	0.076545

Table 10.3.3: Observations, Fitted Observations, Residuals and Measurement Std. Dev. (*mg*)

$\mathbf{Y}$	$\hat{\mathbf{Y}}$	$\left(\mathbf{Y} - \hat{\mathbf{Y}}\right)$	$diag\left(\psi_Y\right)^{1/2}$
1.0666	1.07804	-0.0120427	0.05415
2.8	2.7994	0.000596272	0.04142
1.71666	1.721364	-0.00476107	0.03954
-1.175	-1.14976	-0.025239	0.02952
-2.4166	-2.50633	0.0947344	0.05971
0.191666	0.195073	-0.00347339	0.01440
0.348333	0.354341	-0.00604082	0.02291
-0.1666	-0.170835	0.00483518	0.02283
-0.535	-0.525176	-0.009824	0.02078
1.555	1.55165	0.00335322	0.08850

From these two tables we can see that we still have a good fit to the experimental data and also a satisfactorily small Difference Vector when compared with the standard deviation column in the right of Table 10.3.2 above. *However, we can also see that the estimated parameter values in the second column of Table 10.3.2 above are nothing like those we obtained earlier!* So in spite of the fact that the estimation technique has performed well with the information supplied, it has not been able to uncover the drift error at all. This once again highlights a critical weakness in calibration experiments which only supply difference information about the parameters of interest and means that the external prior information must provide absolute values for the parameters. If this prior data is absolutely wrong, we have a fundamental problem which *no estimator* could hope to circumvent. Thus in any mass calibration experiment, it is essential that at least some of the standards have been

recently calibrated to satisfactorily high accuracy for the current requirements. This may well seem a self-evident statement, but is worth emphasizing since the capabilities of the MAP estimator are such that it would be possible to "calibrate" a group of standards with respect to themselves only, once an initial calibration had been performed with external standards. The very real possibility of drift in the physical value of the standards renders this a dangerous idea indeed.

Therefore, let us amend the situation in our simulation by only letting some of the standards be subject to drift. Looking at the prior information in Table 10.1.1 shows that standard b_1 has higher accuracy than b_2, and similarly for the pairs b_3 & b_4, b_5 & b_6, b_7 & b_8. So we will adopt the realistic situation that standards b_1,b_3,b_5 & b_7 are Reference Standards while the other four are Test Standards. In this case, we will assume the same drift error applies to these latter four only as applied to all eight in the first attempt above. The prior and posterior values are shown in Table 10.3.4 where the four "in-error" parameters have been highlighted.

Table 10.3.4: Comparison of Prior and Posterior Data, after carrying out MAP Estimation (data in *mg*)

μ_β	$\hat{\beta}$	$(\mu_\beta - \hat{\beta})$	$diag(\psi_{\hat{\beta}})^{1/2}$	$diag(\psi_\beta)^{1/2}$	$s(\mu_\beta - \hat{\beta})$
2.0	1.93936	0.0606411	0.133485	0.25	0.283405
0.1	**0.862356**	**-0.762356**	**0.134876**	**0.75**	**0.762031**
-0.9	-1.00357	0.103566	0.0673833	0.125	0.142005
-0.3	**0.146919**	**-0.446919**	**0.0671767**	**0.375**	**0.380969**
0.45	0.403082	0.0469178	0.0276786	0.05	0.0571498
-0.23	**0.0420053**	**-0.272005**	**0.0291039**	**0.15**	**0.152797**
1.07	1.06216	0.00783698	0.0153515	0.05	0.0523036
-0.66	**-0.491863**	**-0.168137**	**0.0153016**	**0.075**	**0.076545**

In this situation we see that the four parameters with an uncorrected simulated drift have had their prior values nearly perfectly corrected by the estimation process, and we are returned to posterior values very close to those obtained in the first estimation with "correct" prior data. (Table 10.1.3). Satisfactory agreement with the experimental data has also been obtained, as shown in Table 10.3.5 below. Of course we have seen this situation already: what we have is one set of prior values (b_2,b_4,b_6 & b_8) having a larger prior variance than the other four. Therefore, if they are in error, they will easily be adjusted by the estimator, as we have seen in previous sections. If they have a larger variance and are *not* in error they will not be adjusted significantly but will simply have their posterior variance reduced by the estimator.

Table 10.3.5: Observations, Fitted Observations, Residuals and Measurement Std. Dev. (*mg*)

$\mathbf{Y}$	$\hat{\mathbf{Y}}$	$\left(\mathbf{Y}-\hat{\mathbf{Y}}\right)$	$diag(\psi_{\mathbf{Y}})^{1/2}$
1.0666	1.077	-0.011003	0.05415
2.8	2.79601	0.00399356	0.04142
1.71666	1.719	-0.00240348	0.03954
-1.175	-1.15048	-0.0245151	0.02952
-2.4166	-2.51082	0.0992168	0.05971
0.191666	0.193694	-0.00209389	0.01440
0.348333	0.361077	-0.0127769	0.02291
-0.1666	-0.167218	0.00121801	0.02283
-0.535	-0.528295	-0.00670505	0.02078
1.555	1.55403	0.000974214	0.08850

To illustrate the situation more clearly, let us set ψ_β as in Fig. (10.3.1) below where we have attached a scaling parameter v to the prior variance of each of the parameters (b_2,b_4,b_6 & b_8).

$$\psi_\beta = \begin{bmatrix} 0.0625 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ & (0.5625)v & 0 & 0 & 0 & 0 & 0 & 0 \\ & & 0.015625 & 0 & 0 & 0 & 0 & 0 \\ & & & (0.140625)v & 0 & 0 & 0 & 0 \\ & & & & 0.0025 & 0 & 0 & 0 \\ & & symm & & & (0.0225)v & 0 & 0 \\ & & & & & & 0.0025 & 0 \\ & & & & & & & (0.005625)v \end{bmatrix}$$

Fig. 10.3.1: Alternative Prior Covariance Matrix, ψ_β

By scaling v over a wide range of values it is now possible to adjust the influence of the four parameters in question. We do this for two cases: one in which the simulated drift error is present on μ_β, and one for which it is not—i.e. the correct prior data, as in Table 10.1.1, is used—and we let v vary in the range $\sim(0,\infty)$ for the two cases. The result, shown in Fig. (10.3.2), for just parameter b_2, illustrates the situation very clearly: prior data given a low degree of belief will be adjusted significantly by the estimator, *if the evidence demands it*, and the posterior estimate will be assigned a lower variance of course. If the prior information in question is in good agreement with the rest of the evidence (other prior data and current data) little adjustment to its value will result while the large prior variance will be reduced. If on the other hand, as we have remarked several times before, the prior data in question is assigned a very *high* degree of belief, little or no adjustment to its value will be possible, irrespective of what might be required.

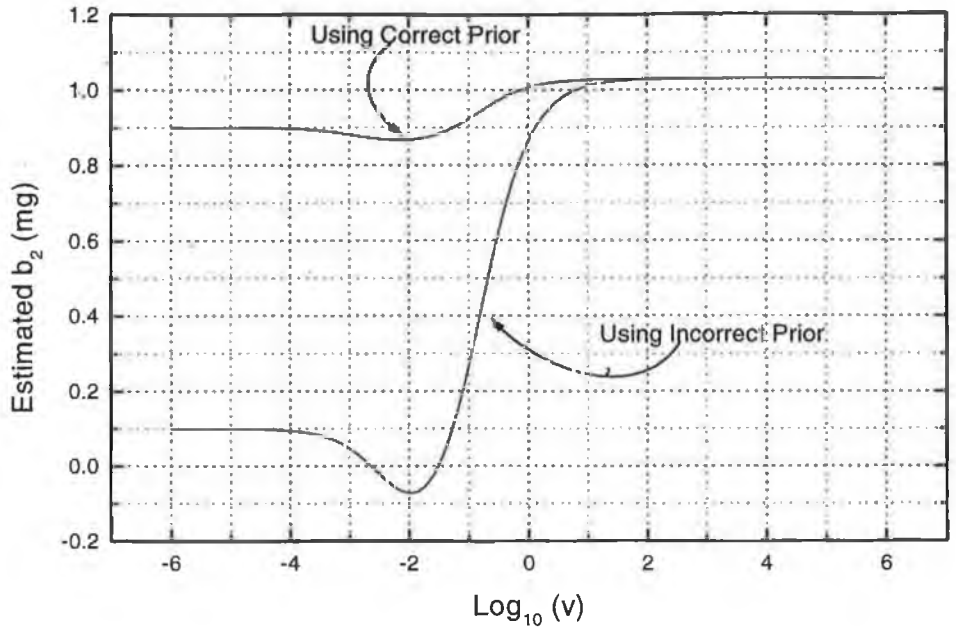


Fig. 10.3.2: Posterior Estimate for b_2 using Ψ_β as given in Table (10.3.4). Two cases are given: "Using Incorrect Prior" refers to the simulated drift error on parameters (b_2,b_4,b_6 & b_8) as discussed in the text; the other curve shows the situation using the normal prior data as given in Table 10.1.1.

Thus, providing some of our prior information is known to be accurate (which will be reflected in its variance), inaccurate, or invalidated prior information can be easily dealt with providing its variance is made sufficiently large to allow any necessary adjustments. However, it must be pointed out that there is a fundamental problem here: how do we decide which standards are the most accurate? Suppose that all standards involved in a comparison have very similar prior variances. Then it may be more difficult to decide to increase the prior variance of some, rather than others, the difference between the various options possibly affecting the posterior estimates significantly, if there were indeed errors in the prior information. (If there were no errors, the difference would be minimal.) To answer this dilemma we must turn to our criteria of logical reasoning: we must consider which is physically most likely to be subject to error (drift), based on the treatment it has received since the last calibration—and indeed, the time interval since the last calibration. We must also consider any other calibrations any of the standards have been involved with, which might increase (or decrease!) the justification for considering the prior information to be still accurate. In mass metrology there is a natural hierarchy in reference standards which is generally used to answer this question, but the point remains that there could be potential difficulties when working at one level rather than between levels in the calibration chain. This issue also has repercussions at the very top of the scale at the

level of the international primary standards, and is fundamentally insoluble without some external information—hence the research thrust towards the non-artifact kilogram standard! See for example Becker et al, Bego (1995), Davis (1989), Frantzus (1996), Olson (1991), Quinn (1991), Robinson & Kibble (1996), (1997), Steiner (1996), Steiner *et al* (1997), Taylor (1991).

10.4 Example III

In this final example the same 10 x 8 design scheme as used in the previous example is employed (reproduced in Fig. (10.4.1) below). The same nominal values of parameters—i.e. 1000g to 100g are used, the important difference here being that prior information is only available for the first two standards (the 1000g standards). The remaining eight are new standards which have never been calibrated before. This is reflected in the central two columns of Table 10.4.1—the prior data. This example is also similar to Example I insofar as one of the standards (b_1) has a much higher prior degree of belief than the other. Parameter b_2 had not previously been calibrated for several years and we will utilise this situation to explore possibilities for correcting drift on the prior information in the context of the MAP estimator.

$$\mathbf{X} = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & -1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & -1 & -1 & -1 & 0 \\ 0 & 0 & 0 & 1 & -1 & -1 & 0 & -1 \\ 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & -1 & -1 \\ 0 & 0 & 0 & 0 & 0 & 1 & -1 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 \end{bmatrix}$$

Fig. 10.4.1: Design Matrix

Table 10.4.1: Prior Information

Parameter	Nominal Value (g)	Prior Value (deviation) (μg)	Std. Dev (μg)	Volume (cm^3)	'Vol. Std. Dev (cm^3)
b_1	1000.0	-960.0	75.0	124.219	0.01
b_2	1000.0	+2723.0	250.0	126.936	0.02
b_3	500.0	-	-	62.124	0.005
b_4	500.0	-	-	62.125	0.005
b_5	200.0	-	-	24.849	0.003
b_6	200.0	-	-	24.849	0.003
b_7	100.0	-	-	12.4261	0.0011
b_8	100.0	-	-	12.4254	0.0011

Fig. (10.4.2) below illustrates the relevant prior vectors and covariance matrices. Observe that parameters b_2 to b_8 have prior values of *0* μg according to μ_β in Fig. (10.4.2), which might seem strange since according to Table 10.4.1 we are not warranted in assigning any prior values to these parameters. However, note also that we have not specified ψ_β, but rather ψ_β^{-1} in Fig. (10.4.2). From this we can see that the prior variance of parameters b_2 to b_8 would be ∞ and hence we can assign any prior value we like as it will have no effect upon the posterior estimate. (Of course we can only easily do this because ψ_β is diagonal: for example, if we had known covariances between parameters b_1 & b_2, things would not be computationally so easy as we would be faced with the problem of sparse matrices to deal with.)

$$\mu_\beta = \begin{bmatrix} -960.0 \\ 2723.0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \mu g \; ; \; \psi_\beta^{-1} = \begin{bmatrix} 177.77 & & & & & & & \\ & 16 & & & & & & \\ & & 0 & & \mathbf{0} & & & \\ & & & 0 & & & & \\ & & & & 0 & & & \\ & \mathbf{0} & & & & 0 & & \\ & & & & & & 0 & \\ & & & & & & & 0 \end{bmatrix} \mu g^2 \times 10^{-6}$$

$$\mathbf{V} = \begin{bmatrix} 124.219 \\ 126.936 \\ 62.124 \\ 62.125 \\ 24.849 \\ 24.849 \\ 12.426 \\ 12.425 \end{bmatrix} cm^3 \; ; \; \psi_V = \begin{bmatrix} 100 & & & & & & & \\ & 400 & & & \mathbf{0} & & & \\ & & 25 & & & & & \\ & & & 25 & & & & \\ & & & & 9 & & & \\ & \mathbf{0} & & & & 9 & & \\ & & & & & & 1.21 & \\ & & & & & & & 1.21 \end{bmatrix} \left(cm^3\right)^2 \times 10^{-6}$$

Fig. 10.4.2: Prior Information

The experimental information for the 10 comparisons carried out on the 8 parameters is given in Table 10.4.2 below The diagonal elements of $\psi_{\Delta W}$ (Fig. (10.4.3)) are constructed from column 2 of Table 10.4.2 plus a $\left(\frac{1}{3}\right)\mu g^2$ term arising from a uniform distribution of $\pm 1 \mu g$ representing the comparator display uncertainty (rounding/digitisation error).(See Lira & Wöger (1997)) An air density variance of $s^2(\rho_a) = 1{\cdot}5 \times 10^{-7} (mg.cm^{-3})^2$ is used, following the considerations leading to Eq. (3.2.5a). Eq. (4.5.17) is used to construct ψ_Y. Recall that air density matrix ρ in Eq. (4.5.17) is formed with the 3rd column of Table 10.4.2 on its diagonal and that we have $\psi_\rho = s^2(\rho_a).\mathbf{I}_{10}$ in this case. With this information Fig. (10.4.4) gives the second term on the r.h.s. of Eq. (4.5.17); this is the contribution to ψ_Y due to the air density

term in the Weighing Equation. Fig. (10.4.5) gives the third term of Eq. (4.5.17) which is the contribution due to the volume term in the Weighing Equation. Observe that this latter term contributes by far the largest to the overall covariance matrix, ψ_Y, which is shown in Fig. (10.4.6).

Table 10.4.2: Experimental Information

$\Delta\overline{W}$ (μg)	Variance. $(\mu g)^2$	ρ_a ($mg.cm^{-3}$)	ΔV (cm^3)	$Y \equiv \Delta\overline{m}$ (μg)
-508.166	0.3611	1.197208	-2.717	-3760.98
-125.833	0.0944	1.194606	-0.030	-161.67
392.833	0.2277	1.193389	2.687	3599.47
-32.5	0.18333	1.211381	-0.001	-33.7
62.166	0.16111	1.198981	-0.0001	62.0
86.66	0.2444	1.186522	0.0016	88.56
20.20	0.1777	1.211495	0	20.20
17.5	0.1166	1.209898	-0.0025	14.47
-1.0	0.1333	1.209051	-0.0025	-4.02
-6.166	0.4277	1.199339	0.0007	-5.32

$$\psi_{\Delta W} = \begin{bmatrix} 0.6944 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ & 0.4277 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ & & 0.5611 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ & & & 0.5166 & 0 & 0 & 0 & 0 & 0 & 0 \\ & & & & 0.4944 & 0 & 0 & 0 & 0 & 0 \\ & & & & & 0.5777 & 0 & 0 & 0 & 0 \\ & & & & & & 0.5111 & 0 & 0 & 0 \\ & & & symm & & & & 0.45 & 0 & 0 \\ & & & & & & & & 0.4666 & 0 \\ & & & & & & & & & 0.7611 \end{bmatrix}$$

Fig. 10.4.3: Covariance matrix of the observations $(\mu g)^2$

$$t_2 = \begin{bmatrix} 1.10731 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ & 0.000135 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ & & 1.082995 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ & & & 1.5*10^{-7} & 0 & 0 & 0 & 0 & 0 & 0 \\ & & & & 1.5*10^{-9} & 0 & 0 & 0 & 0 & 0 \\ & & & & & 3.84*10^{-7} & 0 & 0 & 0 & 0 \\ & & & & & & 0 & 0 & 0 & 0 \\ & & & symm & & & & 9.375*10^{-7} & 0 & 0 \\ & & & & & & & & 9.375*10^{-7} & 0 \\ & & & & & & & & & 7.35*10^{-8} \end{bmatrix}$$

Fig. 10.4.4: $t_2 \left(\equiv diag\{\mathbf{XV}\}\psi_\rho diag\{\mathbf{XV}\}\right)\mu g^2$

$$t_3 = \begin{bmatrix} 716.653 & 143.019 & -571.493 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ & 214.062 & 71.281 & 0 & -35.808 & -35.435 & 0 & 0 & 0 & 0 \\ & & 640.880 & 0 & -35.771 & -35.399 & 0 & 0 & 0 & 0 \\ & & & 73.372 & 36.311 & -35.933 & 0 & 0 & 0 & 0 \\ & & & & 63.554 & 25.607 & 0 & -11.300 & -11.293 & -1.740 \\ & & & & & 62.240 & 0 & -11.183 & -11.175 & 1.722 \\ & & & & & & 26.419 & 13.192 & -13.182 & 0 \\ & & & symm & & & & 16.717 & 3.540 & 0 \\ & & & & & & & & 16.694 & 0 \\ & & & & & & & & & 3.481 \end{bmatrix}$$

Fig. 10.4.5: $\mathbf{t}_3 \left(\equiv \rho \mathbf{X} \psi_v \mathbf{X}^T \rho\right) \mu g^2$

$$\psi_v = \begin{bmatrix} 718.455 & 143.019 & -571.493 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ & 214.49 & 71.281 & 0 & -35.808 & -35.435 & 0 & 0 & 0 & 0 \\ & & 642.523 & 0 & -35.771 & -35.399 & 0 & 0 & 0 & 0 \\ & & & 73.889 & 36.311 & -35.933 & 0 & 0 & 0 & 0 \\ & & & & 64.0488 & 25.607 & 0 & -11.300 & -11.293 & -1.740 \\ & & & & & 62.8177 & 0 & -11.183 & -11.175 & 1.722 \\ & & & & & & 26.9293 & 13.192 & -13.182 & 0 \\ & & & symm & & & & 17.1672 & 3.540 & 0 \\ & & & & & & & & 17.1605 & 0 \\ & & & & & & & & & 4.2413 \end{bmatrix}$$

Fig. 10.4.6: Complete Covariance Matrix of the input data $(\mu g)^2$

Now, using this information we can evaluate the parameter estimates and their respective covariances using Eqs. (8.3.15) & (8.4.5). These are shown in Table 10.4.3 along with the relevant prior data, for comparison purposes; while Fig.(10.4.7) shows the complete covariance matrix. From Table 10.4.3 we can see that both prior parameters b_1 & b_2 have been updated and assigned lower variances, in the case of b_2, a significantly lower variance. This latter point illustrating that the most accurate prior information exerts the greatest influence, a point that has been made several times before. Note also that the adjustment in each case lies comfortably within the bounds of the combined standard deviation of the difference. (Compare columns 3 & 6 in Table 10.4.3). Table 10.4.4 illustrates the fitted observations and residuals.

Table 10.4.3: Comparison of Prior and Posterior Data, after carrying out MAP Estimation (data in μg)

μ_β	$\hat{\beta}$	$\left(\mu_\beta - \hat{\beta}\right)$	$s\left(\hat{\beta}\right)$	$s\left(\mu_\beta\right)$	$s\left(\mu_\beta - \hat{\beta}\right)$
-960.0	-966.385	6.38515	71.8705	75.0	103.88
2723.0	2793.95	-70.9462	75.866	250.0	261.26
-	-418.804	-	36.8421	-	-
-	-386.367	-	36.8386	-	-
-	-180.183	-	14.9852	-	-
-	-199.719	-	14.9851	-	-
-	-100.591	-	7.40093	-	-
-	-94.6731	-	7.40071	-	-

$$\Psi_{\hat{\beta}} = \begin{bmatrix} 5156.37 & 5106.99 & 2576.83 & 2576.83 & 1030.73 & 1030.73 & 515.367 & 515.367 \\ & 5755.65 & 2618.51 & 2618.51 & 1047.39 & 1047.39 & 523.695 & 523.695 \\ & & 1357.34 & 1320.84 & 528.423 & 528.423 & 264.332 & 264.122 \\ & & & 1357.08 & 528.579 & 528.579 & 264.099 & 264.43 \\ & & & & 224.557 & 211.198 & 105.679 & 105.637 \\ & & & & & 224.553 & 105.68 & 105.638 \\ & & & & & & 54.7737 & 52.7796 \\ & & & & & & & 54.7705 \end{bmatrix}$$

Fig. 10.4.7: Complete Covariance Matrix of the Parameter Estimates $(\mu g)^2$.

Table 10.4.4: Observations, Fitted Observations, Residuals and Measurement Std. Dev. (μg)

$\mathbf{Y}$	$\hat{\mathbf{Y}}$	$\left(\mathbf{Y}-\hat{\mathbf{Y}}\right)$	$diag\left(\Psi_{\mathbf{Y}}\right)^{1/2}$
-3760.98	-3760.33	-0.648696	26.804
-161.67	-161.214	-0.455955	14.645
3599.47	3599.12	0.352741	25.348
-33.7	-32.4374	-1.262640	8.596
62.0	61.6888	0.311221	8.003
88.56	88.2079	0.352122	7.926
20.2	19.5357	0.664279	5.189
14.47	15.0815	-0.611459	4.143
-4.02	-4.45426	0.434262	4.142
-5.32	-5.91826	0.598264	2.059

10.5 Correcting the Prior Information

In Section 10.3, while discussing Example II, we noted how we could adjust the plausibility of the prior information in order to take account of possible drift or other errors. In this example, we see that parameter b_2 has a much larger prior variance than does b_1, and indeed it is adjusted by a comparatively larger amount. By increasing its prior variance still further, we would approach the situation where this prior information about b_2 exerts no influence on the posterior estimates and we could then see the maximum adjustment possible. As we have remarked before, if there was no disagreement between prior and current data, the adjustment would be minimal. To the extent that we give b_2 a finite prior variance, we are attaching value to this prior information and allowing it to influence the posterior estimates

However, if we know before carrying out the estimation that b_2 *has* drifted, should we update the prior value first? Or would this be pre-empting the estimator which should be able to highlight the drift *a posteriori*? (Providing of course that there is other, independent and indeed accurate prior information also !) We can answer this question by yet again stating that *any known information must be included in the analysis.* How much value we attach to it is decided by the variance we assign it. Whether such information on drift obtains from a careful empirical analysis or the studied opinion of the "Expert Observer" it still must have some validity. So the natural question for us to ask concerns the possibility of such 'prior adjustment' vitiating the posterior estimate. In other words, accurate correction for drift should actually help the estimator, but would over-enthusiastic correction hinder it?

To investigate, we shall replace $s^2(b_2)$ in ψ_β with $v \times s^2(b_2)$ where v is once again a multiplicative scalar, and compare the posterior estimate for a range of v values in situations of no drift correction, a moderate drift correction, and a much larger drift correction. The situation is presented in Fig. (10.5.1) below.

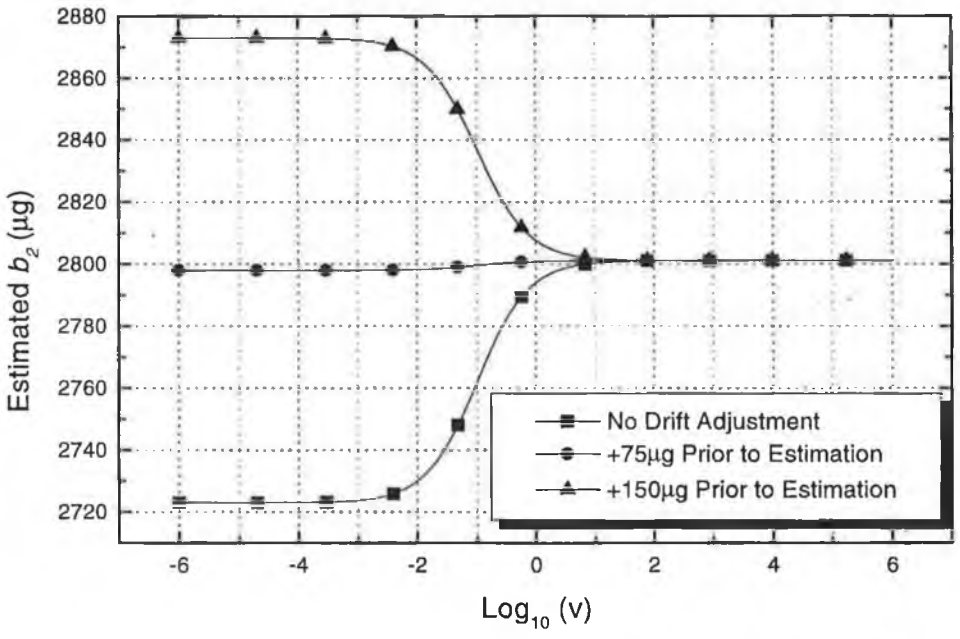


Fig. 10.5.1: Posterior Estimates of b_2 for 3 different 'prior correction of drift' situations. The x-axis parameter v controls the prior variance of b_2, as discussed in the text.

The conclusion we can draw from this information is that the most important way to tackle the possibility of drift or other errors in parts of the prior information is to ensure that its degree of belief is lower than that of the other prior information, as we have remarked before. Then any other adjustments to include suspected drift

cannot adversely affect the posterior estimates. We can see from Fig. (10.5.1) above that a correction of $+75\mu g$ results in very good agreement with the other information since very little adjustment takes place irrespective of the prior variance. The other two cases of no adjustment and what seems to be too much adjustment are largely corrected by the estimator providing $s^2(b_2) \geq s^2(b_1)$. This is quite intuitive as it simply tells us that the more "accurate" information exerts a greater influence, but it is important to draw attention to it, since it highlights the logical nature of this estimation technique and shows how it does implement criteria of plausible reasoning with whatever information is supplied. The imperative rests with the experimenter to supply physically relevant data and to be aware of the limitations of the mathematical tools which can only operate on the supplied information.

****

11. EXPERIMENTAL SYSTEM

In this chapter we will consider the hardware and software used in the experimental work which was carried out at the National Metrology Laboratory, Forbairt, Dublin, and the data reduction carried out prior to the parameter estimation process. We will highlight some interesting experimental results and point out the effects of various systematic problems on the data which lead to inconsistencies among group comparisons and later problems in the parameter estimations. Some sample data is given to illustrate the methods used.

At the heart of the calibration system are the mass comparators used in the comparison experiments. In this case, commercially available, automated instruments based on the electromagnetic force compensation principle are used. These are equipped with micro-positioning turntables for alternately placing each of the two weights (or combinations of weights) involved in the comparison on the load pan. A self-centring mechanism and a lever arrangement involving flexure strips (see, e.g. Quinn *et al* (1986/87) ensures high reproducibility can be achieved. A standard RS232 serial interface is provided and a simple instruction set applies to all the instruments allowing for easy computer control to be implemented. Table 11.1 below shows the instruments used in this study along with their respective ranges and accuracies.

Table 11.1

Instrument	Range	Readability	Rated Std. Dev.
Sartorius C50s	10 g → 50 g	1 μg	10g : 4 μg 20g : 5 μg 50g : 6 μg
Sartorius C1000s	100 g → 1000 g	2 μg	100g, 200g : 2 μg 500g, 1000g : 5 μg
Sartorius C10000s	2 kg → 10 kg	0.1 mg	0.1 mg
Sartorius C20000	10 kg → 20kg	1.0 mg	1.5 mg

Secondly, we need instrumentation to monitor air temperature, barometric pressure and relative humidity in order to calculate air density and the systematic buoyancy correction. Again the instruments used are equipped with serial interfaces and can be interrogated by the controlling computer. Table 11.2 lists the equipment used. The 'system accuracy' listed in Table 11.2 below is a standard uncertainty obtained from instrument calibration certificates.

Table 11.2

Instrument	Range	Resolution	System Accuracy
Hart Scientific Model 1502 Resistance Bridge with Model 1514 PRT Probe	$-100°C \rightarrow +300°C$	$1\ mK$	$60\ mK$
Druck Model DPI 141 Resonant Sensor Barometer	$800 \rightarrow 1100 mBar$	$1\ Pa$ ($0.01\ mBar$)	$15\ Pa$
Vaisala Model HMP 233 Capacitative Humidity Probe	$5\% \rightarrow 95\%, RH$	0.1 %	2 %

Extensive software has been specially written as part of this research work to allow these instruments to be controlled by a computer.[1] Here its principal features and mode of operation will be described. By means of an eight-port serial interface card (Brain Boxes Lynx 8-Port RS232), it is possible for the computer to control several instruments at once. So the comparator in use and the three environment-monitoring instruments are all connected to the one computer, which is physically located some distance from them in order to reduce unwanted interference and heat generation. The software can then log the weight-in-air differences measured by the comparator and the temperature, pressure and humidity at the time of measurement. It is possible to select the start-time for an experiment and its duration in order to allow experiments of arbitrary length to be carried out at arbitrary times. This is particularly advantageous as it is then possible to commence a comparison experiment at times when the laboratory is 'quiet', e.g. night-time. The 'length', or duration, of a comparison is not set by time, but by the number of repeat comparisons to be carried out by the automated comparator. It is also possible to carry out several 'batches' of comparisons with arbitrary intervals between each, without the need for an operator to initiate each one. The software graphically displays the measurements in real-time so that it is possible to see at a glance the current mean value and standard deviation, and also the degree of stability or drift in the measurements. The availability of these features has allowed many of the characteristics and dependencies of the measurement process, described later, to be observed. Fig. (11.1) below illustrates schematically how the various components interconnect. Note that temperature and humidity are measured *inside* the enclosure of the comparator, and thus in the same micro-climate as the standards themselves. Pressure is measured at approximately the same elevation as the standards on an adjoining bench.

[1] Unpublished software documentation and source code describes the details of this. See also the published work *Software Applications in Mass Metrology* reproduced in Appendix 5.

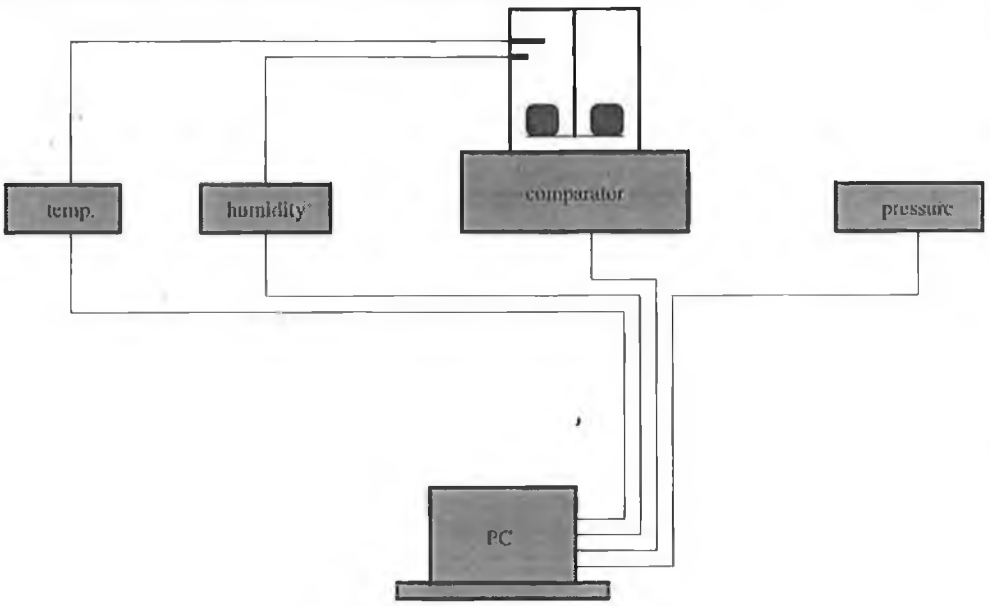


Fig. 11.1: Experimental System

We now wish to examine how this system deals with the data. Fig. (11.2) below shows a typical "raw data" file. Currently, the software only displays the weight-in-air values graphically, it does not split up the other components. In the file fragment in Fig. (11.2) there are 11 columns: the first is the time in seconds from the start of the experiment. This is needed for graphing the data. If there are several sets of comparisons in the file, each separated by a time interval, this will also be reflected in the time column, and thus the 'gaps' will appear in the graph. The second column shows the Weight-in-Air difference (in μg) of the two standards or sets of standards, as evaluated by the comparator. The remaining 9 columns give the climate data in three sets of three—one set for each of temperature, pressure and humidity. The reason for this is as follows: the comparator operates an A-B-B-A comparison sequence and returns data to the PC in two sets: A_1, B_1 & A_2, B_2. The PC queries the climate-monitoring equipment before the start of the sequence, in the middle of the sequence when the first pair of A, B readings are returned, and again at the end of the sequence. Thus the third climate readings of set n and the first of set n+1 will be nearly identical since they are only separated by milliseconds.

This information must now be processed in order to generate meaningful data with which to estimate the parameters. This is done by means of a set of routines which format the raw data into a set of files containing Weight in Air data, temperature, pressure and humidity data. Each file so generated also contains the time information, so that the data sets can be individually graphed. The temperature, pressure and humidity files are obtained by taking means of the three values in the raw data file—i.e. an average for each measurement cycle is used. A density calculation

program forms an air density file from the temperature, pressure and humidity files, by implementing the '81 / 91 Equation for Air Density (Davis (1992)). Finally a buoyancy calculation program takes the Weight in Air file, the air density file, and the appropriate volume difference for the comparison and evaluates the true mass difference for the standards. Fig. (11.3) shows the format of this data processing arrangement. The software routines are described in Appendix 2.

1433.	-118.0	20.624	20.624	20.631	1018.250	1020.450	1020.620	37.410	37.510	37.520
2009.	-120.0	20.630	20.634	20.637	1020.680	1020.680	1020.760	37.520	37.540	37.540
2584.	-122.0	20.637	20.639	20.644	1020.770	1020.770	1020.870	37.550	37.540	37.530
3160.	-124.0	20.644	20.648	20.653	1020.930	1020.930	1020.920	37.520	37.490	37.480
3735.	-121.0	20.653	20.656	20.660	1020.990	1020.990	1021.040	37.480	37.460	37.440
4310.	-122.0	20.660	20.662	20.664	1021.090	1021.100	1021.190	37.450	37.420	37.390
4885.	-123.0	20.664	20.668	20.672	1021.200	1021.210	1021.210	37.390	37.360	37.310
5461.	-122.0	20.671	20.676	20.680	1021.300	1021.380	1021.400	37.310	37.300	37.270
6037.	-122.0	20.679	20.681	20.683	1021.400	1021.400	1021.350	37.270	37.240	37.190
6613.	-122.0	20.683	20.685	20.688	1021.400	1021.410	1021.450	37.210	37.160	37.100
7188.	-122.0	20.689	20.693	20.697	1021.530	1021.540	1021.580	37.110	37.080	37.060
7764.	-121.0	20.697	20.698	20.698	1021.580	1021.570	1021.620	37.050	37.020	37.000
8340.	-120.0	20.697	20.699	20.701	1021.660	1021.660	1021.680	37.000	36.960	36.940
8916.	-121.0	20.701	20.705	20.707	1021.660	1021.660	1021.810	36.930	36.900	36.880
9492.	-120.0	20.707	20.708	20.707	1021.880	1021.890	1021.930	36.870	36.870	36.860

Fig. 11.2: Example data File from the data acquisition program

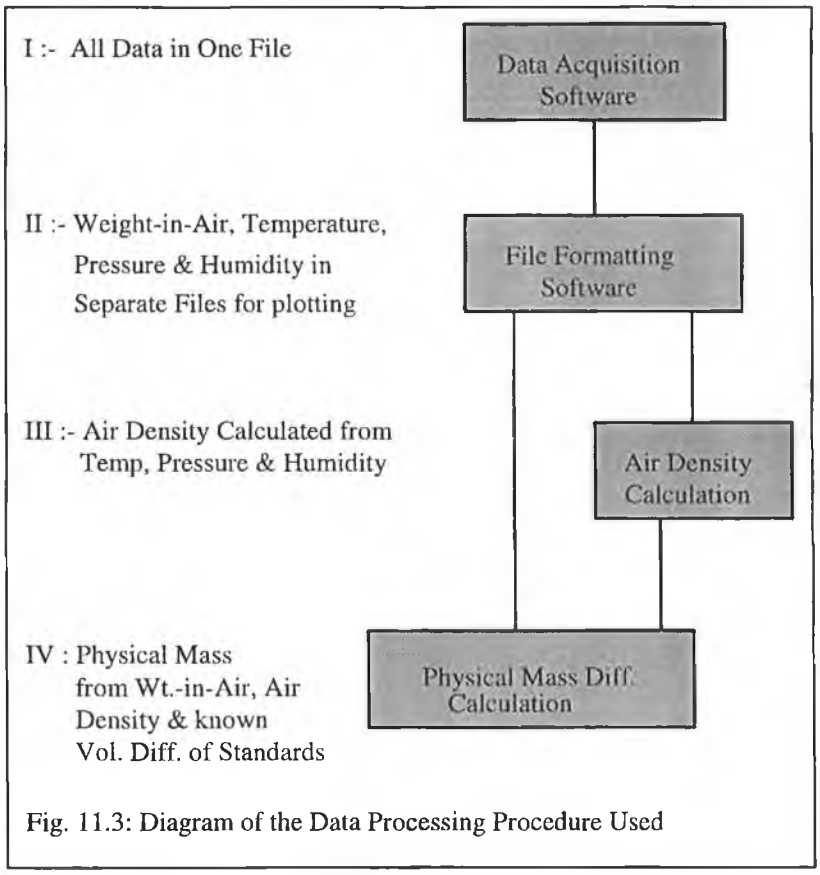


Fig. 11.3: Diagram of the Data Processing Procedure Used

Fig. (11.2) above illustrated the data available from Stage I of the process. Figs. (11.4) to (11.7) below show sample plots of the data available after stage II. Usually the measurement system was operated for a minimum of 12 hours, and often longer. While this amount of data is not in principle needed to establish a mean and variance for later work, it has been found that the measurement system needs to be operated for sufficiently long to allow various initial systematic effects to be minimised, and also to confirm that it has properly stabilised. There is scope for automating this process, via some form of Statistical Process Control to ensure that valid data is obtained. Fig. (11.8) shows the corresponding air density graph, calculated from the data in Figs (11.5) to (11.7) by stage III; while Fig (11.9) is the true mass difference evaluated by stage IV.

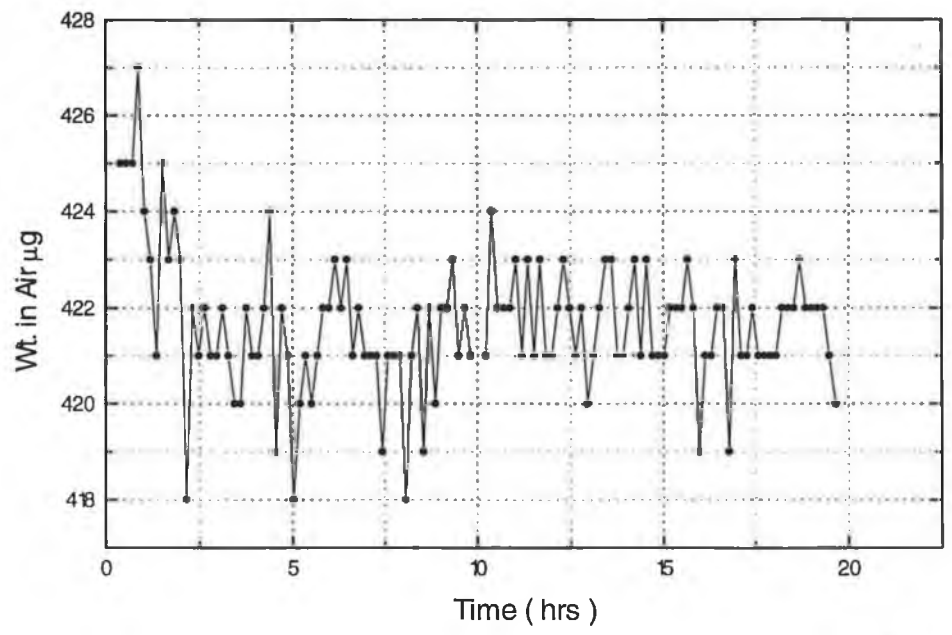


Fig. 11.4: Typical Example of Weight-in-Air Data from Mass Comparator

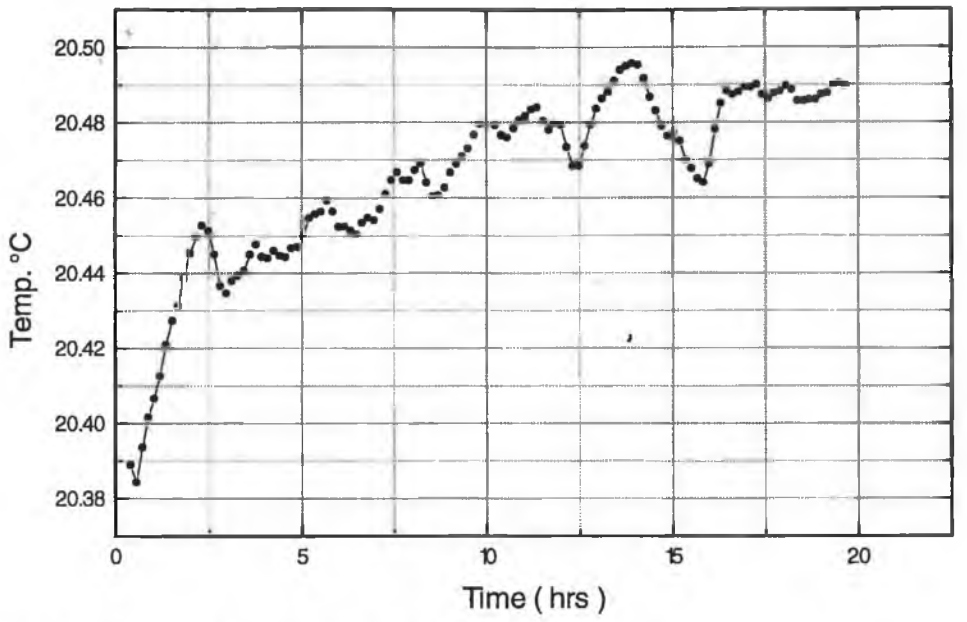


Fig. 11.5: Corresponding Temperature Graph for Fig. 11.4

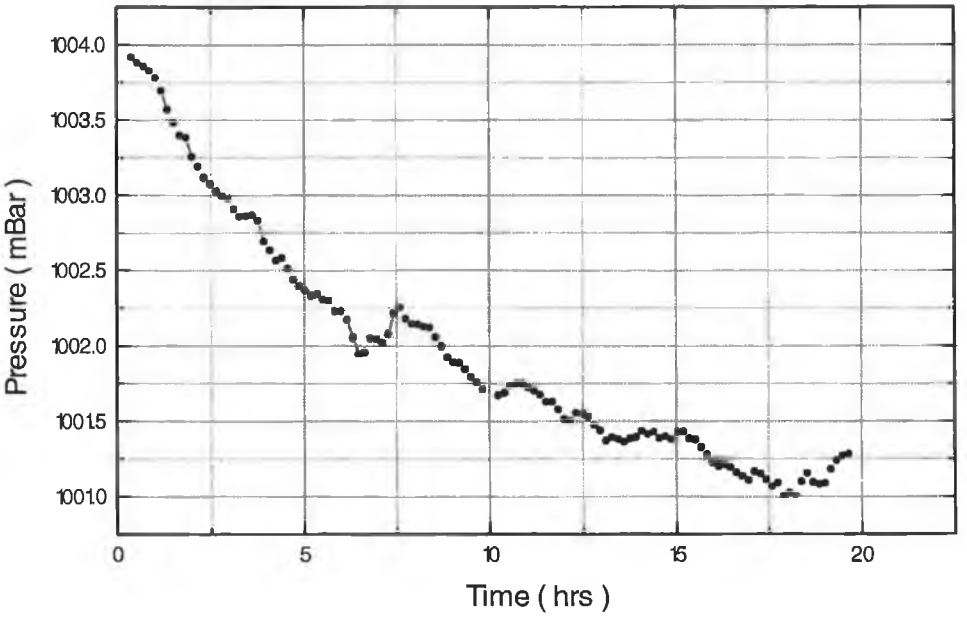


Fig. 11.6: Corresponding Pressure Graph for Fig. 11.4

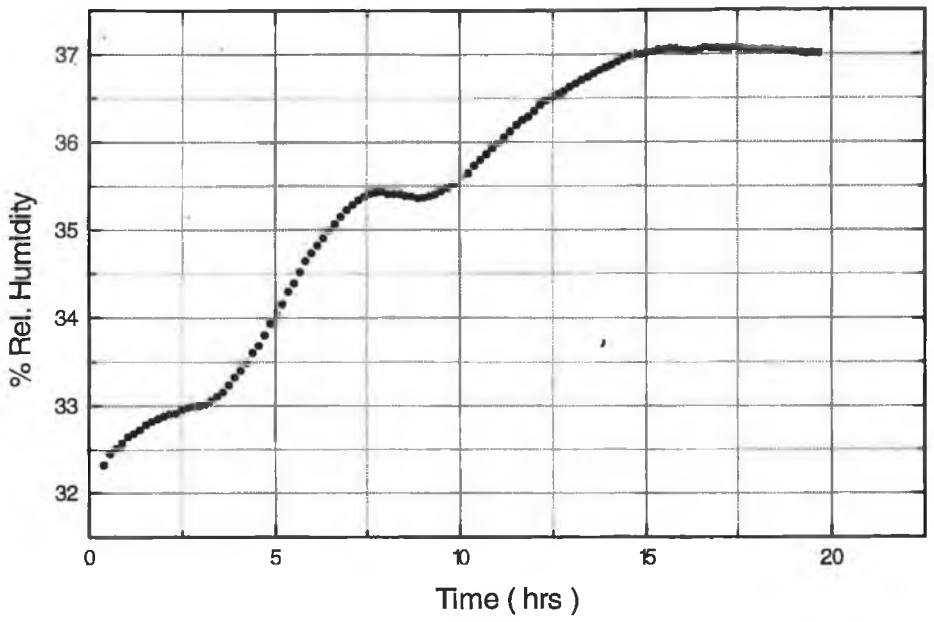


Fig. 11.7: Corresponding Rel. Humidity Graph for Fig. 11.4

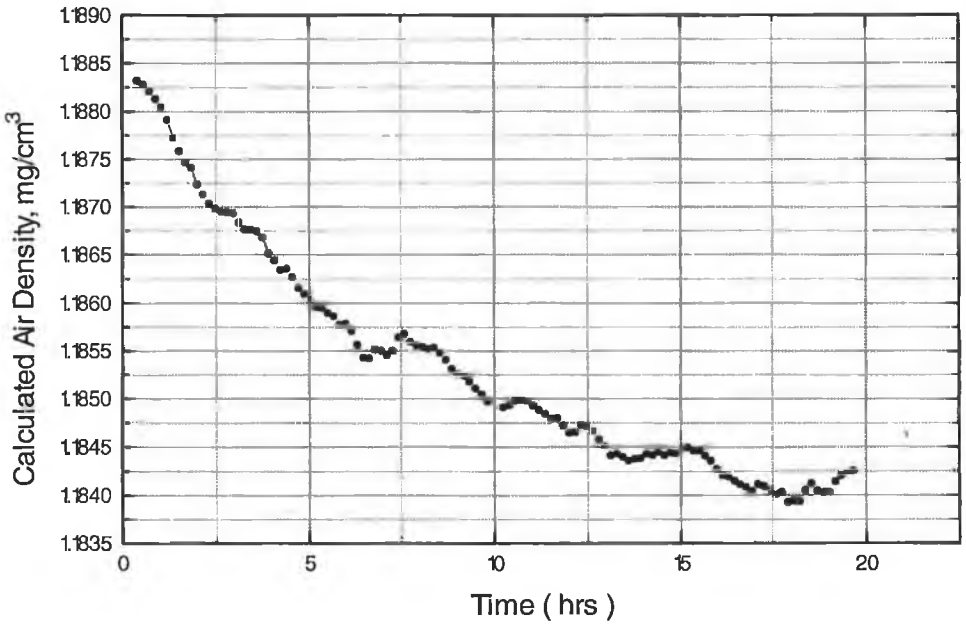


Fig. 11.8: Calculated Air Density For the data in Figs 11.5–11.7

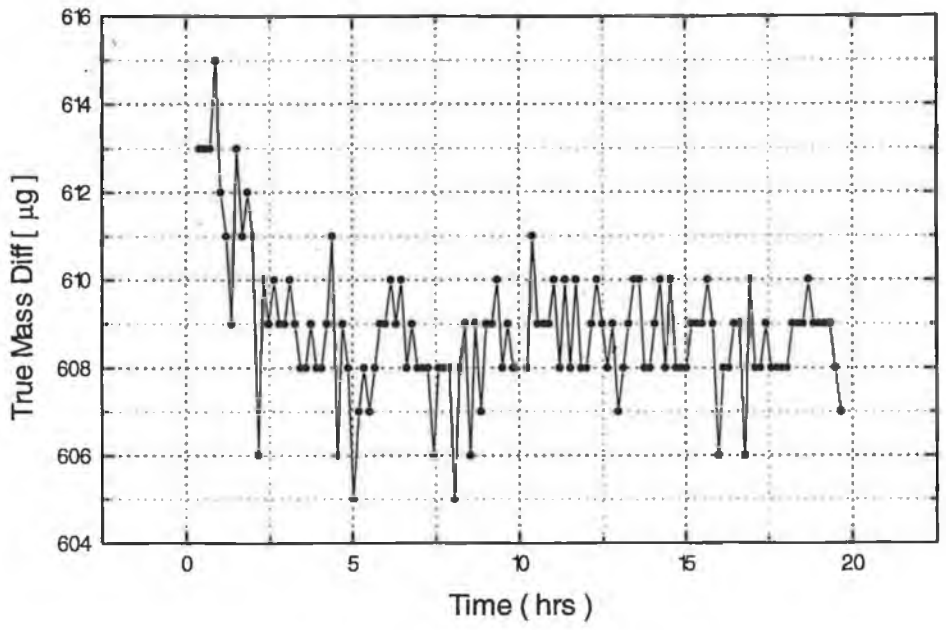


Fig. 11.9: Calculated True Mass Difference from Figs 11.4 & 11.8 [$\Delta V = -0.158 cm^3$]

An analysis of these graphs allows accurate information to be extracted about the comparison process and valid data to be available for the parameter estimation methods discussed earlier in this work. For example, in the graphs shown it is apparent that the system has settled after the first 10 hours and from that point on accuracy is limited by the resolution of the instrument. Indeed Figs (11.4) & (11.9) show clearly why we are justified in including a Uniform distribution in the uncertainty analysis of width equal to the display resolution of the instrument. Note also that there is a distinctive period of instability in the data during the first 4 to 5 hours which is not accounted for by the systematic buoyancy correction, since it is still there in Fig. (11.9) after the correction has been applied. That it is a systematic error is well known since it always appears on all data sets. It is quite likely to be related to a temperature effect since, from Fig. (11.5), we can see a temperature rise once the measurement process has begun. This is a systematic temperature rise always noted once the comparator commences operation and may be due to the motors used in the load alternator for example. Fig. (11.10) is a graph showing the temperature profile in the comparator chamber before, during and after a comparison experiment: the rise in temperature during the experiment and the subsequent fall-off afterwards is quite evident. The type of behaviour shown in Fig. (11.9) leads to the well-known idea of "exercising" the balance prior to obtaining data in order to ensure a steady state situation is reached. There is significant current research effort on-going to better understand the behaviour of flexure strips and torsion strips similar to those used in

mass comparators (e.g. Quinn *et al.* (1997a), (1997b), Kuroda (1995) to mention a few) and it is possible that there may be additional systematic effects concerning the balance mechanisms, such as anelasticity for example, which may also have a bearing on the comparator characteristics we observe. In any event we can conclude that there is a "stabilisation time" required once the comparator has been started before it is possible to arrive at the desired measurand. Of course if this preliminary effect could be adequately modelled, an appropriate additional correction could be applied and a corresponding standard uncertainty term included, but at the moment, since it is a transient effect the approach of allowing it to diminish is adequate for our purposes.

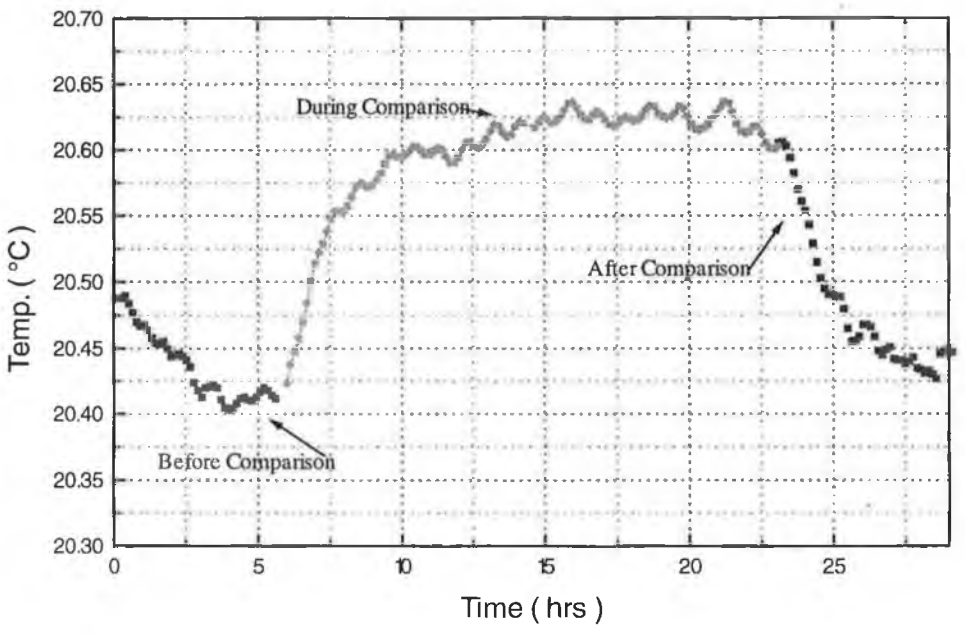


Fig. 11.10: Temperature plot for before, during and after a typical comparison experiment.

This type of analysis illustrates the importance of fully understanding the measurement process in order to correctly realise the desired measurand. For example, a practice of simply operating the comparator for 6, 8, or 10 cycles and taking mean values and standard deviations would be very inadequate here and would lead to systematic errors in the later analysis—which might well be evident if the MAP estimator were used, but their cause would not be identified without the type of practical analysis illustrated here. The practice in this work has been to extract 6 repeat measurements from an obviously stable region of the graph and obtain mean and variance information with these. Now this process is admittedly open to the criticism of being 'subjective' and such questions as 'when is the process obviously stable?' etc. In the present case we can counter with the criterion that when the

corrected measurement results are clearly only limited by the accuracy/resolution of the instrument, there are no further corrections needed, and the data can then be subjected to the desired statistical tools. However, as mentioned earlier, this is a place where Statistical Process Control methods could be implemented with profit: a criterion could be set for when the corrected results are deemed acceptable and the process repeated until it converges to this point. It would be important that only corrected results are used in such an analysis, since systematic errors like the buoyancy correction would introduce obvious drift. This would formalise the decision making process in a manner compatible with the general goal of performing measurements and uncertainty analysis in a uniform, coherent and well-defined manner.

Some systematic errors can be quite subtle and difficult to pin down. For example, in the course of some of this work, a problem of lack of reproducibility between experiments appeared. Two standards were compared over a period of a few weeks and conflicting results emerged. Figs. (11.11) to (11.18) show the data for 4 comparisons on the same pair of 1kg standards. In each case the weight-in-air difference, physical mass difference, and climate data is reproduced. It can be seen that in most cases the buoyancy correction removes the drift which is obvious on the weight-in-air plots, although in the case of comparison 2 (Fig. (11.13)) a significant drift–in the opposite direction–still remains afterwards. Fig. (11.19) shows the true mass values for all 4 comparisons on one plot and they are clearly not in agreement. This was troublesome and suggested some systematic effect was causing a problem. It was difficult to establish just what this was since the buoyancy correction for each individual comparison seemed satisfactory, but when several were compared together the inconsistencies surfaced. This would have caused problems in statistical fitting of the parameters and poor results would follow.

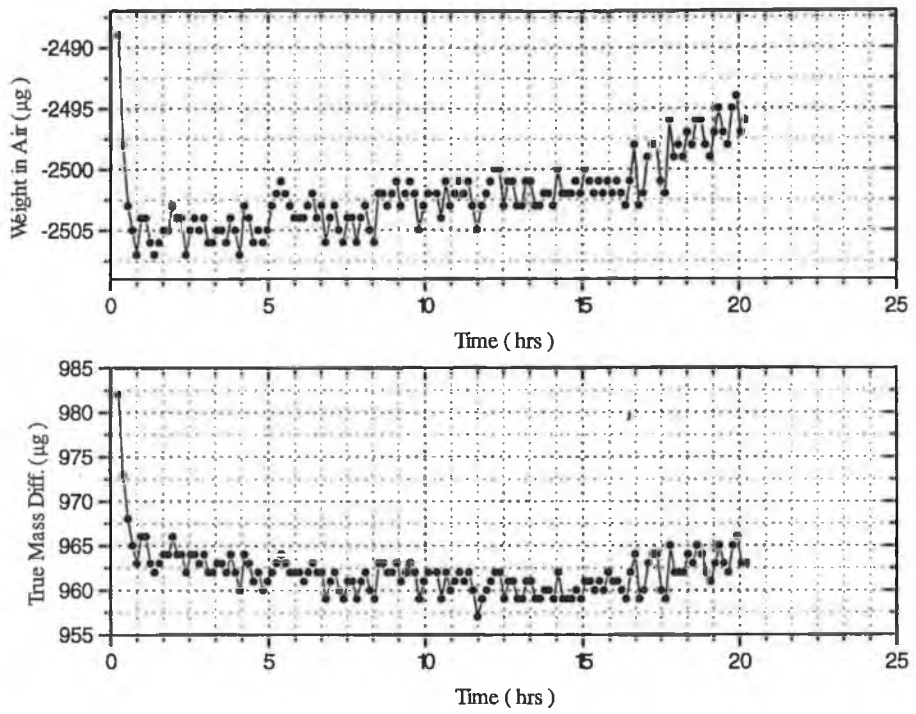


Fig. 11.11: Weight in Air & Physical Mass Differences for Comparison 1

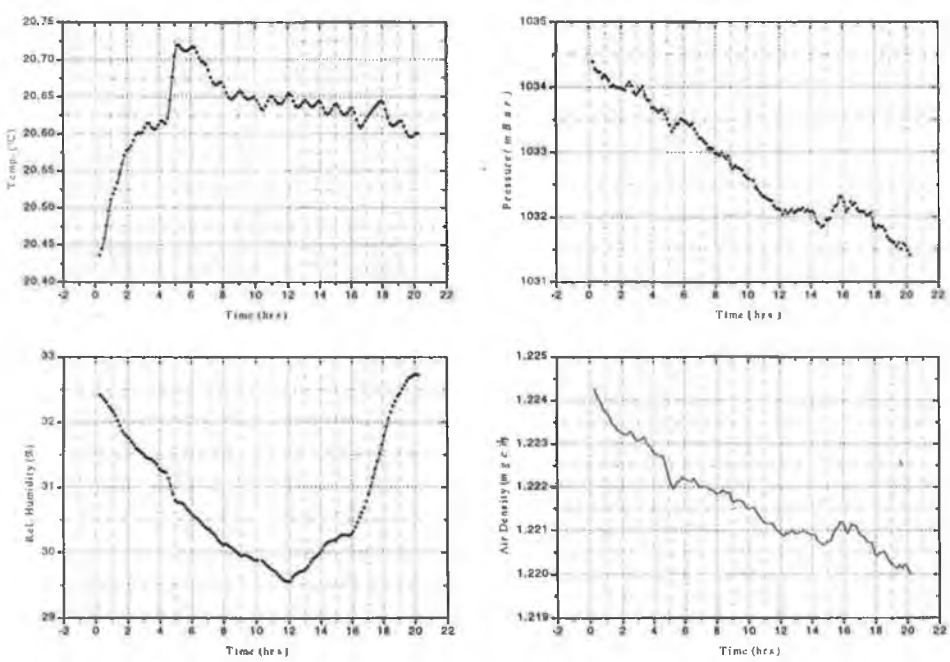


Fig. 11.12: Temp., Pressure, Rel. Hum. & Air Den. for Comp. 1

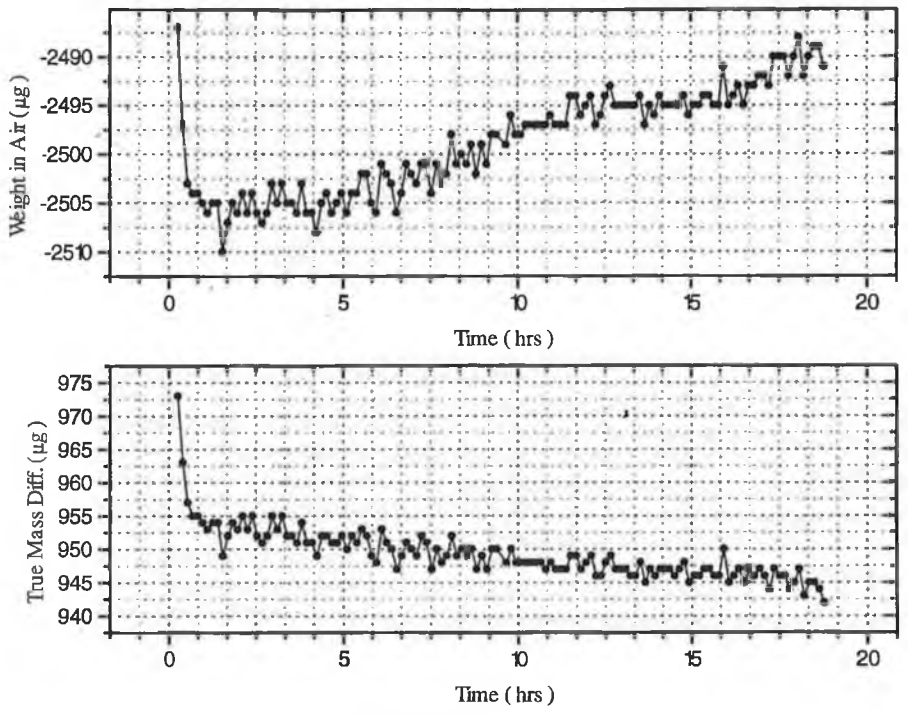


Fig. 11.13: Weight in Air & Physical Mass Difference for Comparison 2

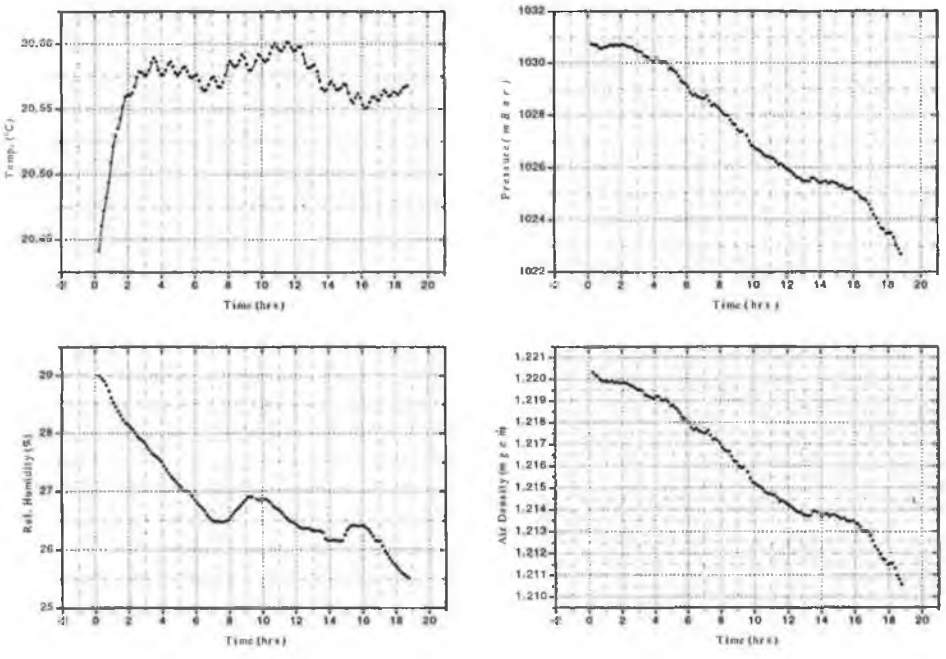


Fig. 11.14: Temp., Pressure, Rel. Hum. & Air Den. for Comp. 2

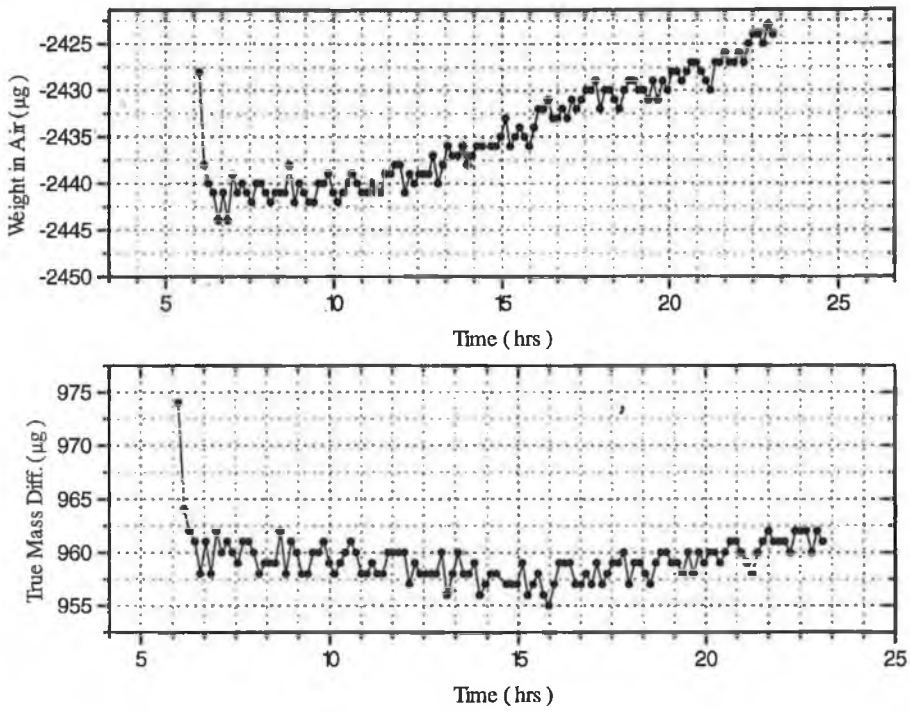


Fig. 11.15: Weight in Air & Physical Mass Difference for Comparison 3

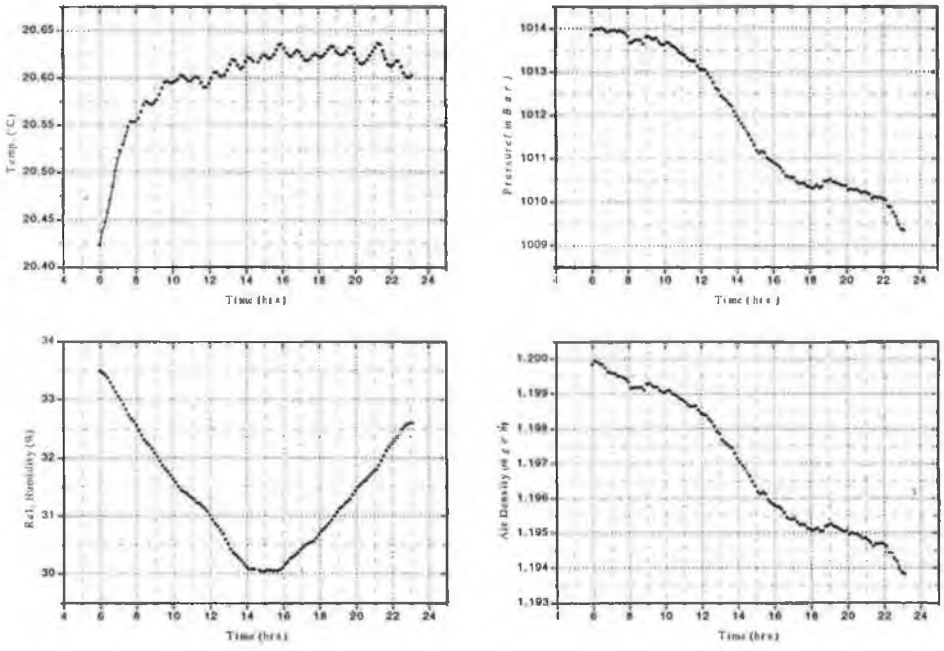


Fig. 11.16: Temp., Pressure, Rel. Hum. & Air Den. for Comp. 3

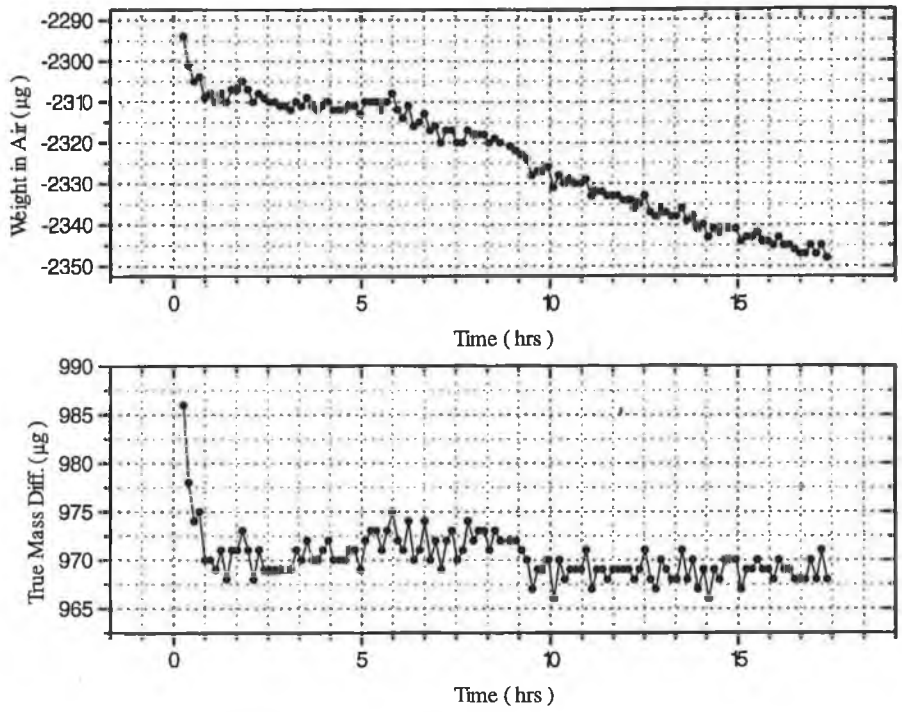


Fig. 11.17: Weight in Air & Physical Mass Difference for Comparison 4

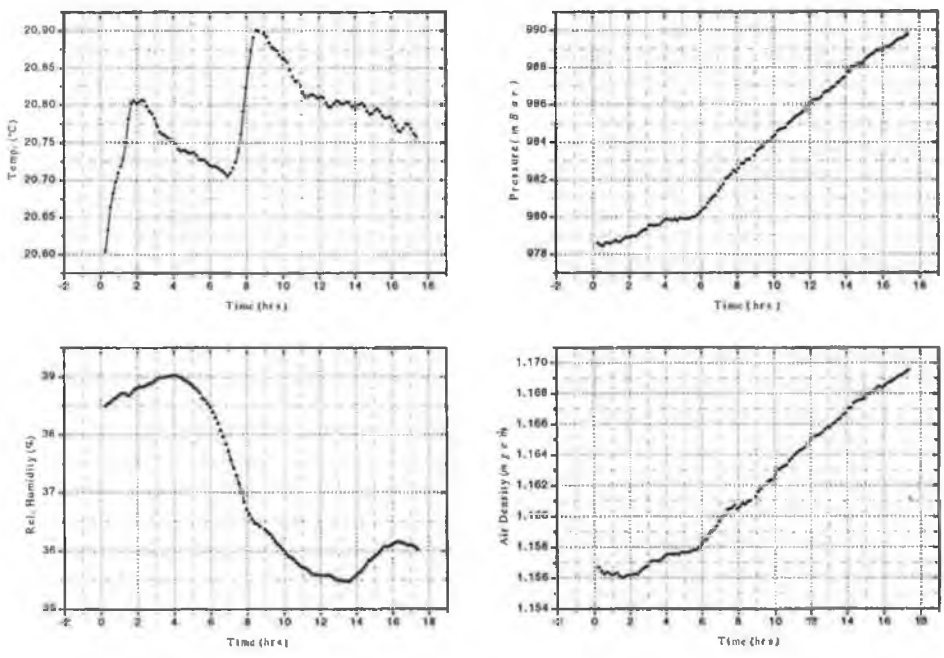


Fig. 11.18: Temp., Pressure, Rel. Hum. & Air Den. for Comp. 4

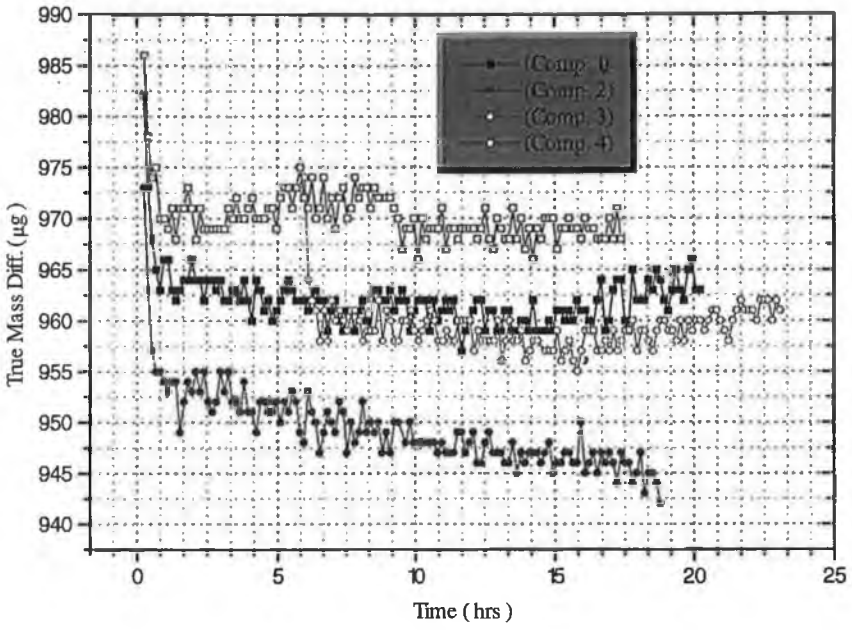


Fig. 11.19: Physical mass Difference plots from Comps 1 - 4 showing the inconsistencies referred to in the text.

However, when the true mass values are plotted as a function of humidity–see Fig. (11.20) below–a large non-linear effect immediately appears, being greatest for lower values of relative humidity. Thus there is a 'hidden' systematic effect, namely the comparators are being influenced by ambient humidity. This is most likely due to an electrostatic effect on the instruments. Now if this could be modelled properly a suitable correction could be applied to the data which would bring the divergent results back in line with one another and of course the functional relationship for the true mass difference would then be adjusted accordingly. In Fig. (11.20) a second degree polynomial is fitted through the data, but this is really for illustrative purposes to show the dependency and should not be taken as the correct fit to the data. It appears from the data that the effect diminishes at higher levels of humidity. This data was obtained during the startup phase of a new laboratory facility and once the humidity levels were stabilised at more suitable levels (~50%) the problem is not manifest.

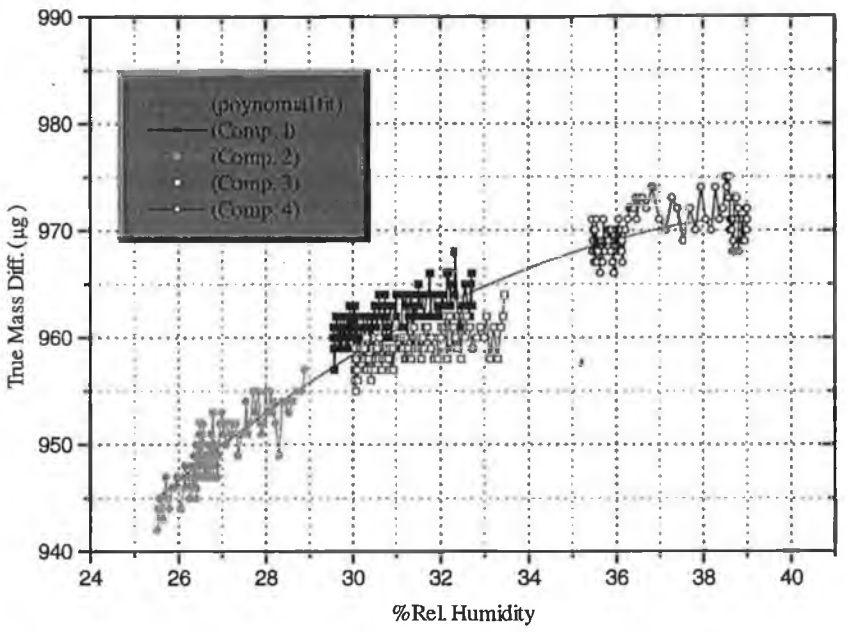


Fig. 11.20: Physical Mass Difference for Comps. 1 - 4 plotted as a function of % Rel. Humidity

Note as an aside that in Figs. (11.12), (11.14), (11.16) & (11.18), the temperature plots show the characteristic rise of ~ 0.2 K at the start of each experiment which has been mentioned before. Other examples of systematic effects due to convection and thermal gradients have been discussed by Glaeser (1990), Glaeser & Do (1993) and Macurdy (1964).

To conclude, we have shown in this section how the data used in the parameter estimation techniques was obtained and highlighted the principal procedures involved. It is very important that the experimental system is properly modelled and that steps are taken to eliminate or correct for any know systematic errors affecting the process. Computer control is useful here as large amounts of data can then easily be gathered from which trends in the data can be deduced. It is important that the functional relationship used to generate the measurand is supported by the observations from the data. Finally it is to this area one must return if the parameter estimation technique suggests disagreements/inconsistencies in the experimental information. It is always possible that an effect has been overlooked, or a possible physical correlation which should have been built into the covariance matrices was not.

12. Conclusion

A central theme of the work reported in this thesis is the establishment of a coherent methodology for analysing experimental data. This applies both to direct analysis of measurements in the area of uncertainty propagation and also to any subsequent processing or parameter estimation for which these initial measurements will form a basis. We have shown how some basic rules of consistent reasoning, underlying classical probability theory, can be applied to all aspects of the work, leading to improved results and a better understanding of the nature of the quantities under investigation.

A critical point relates to the understanding of probability theory in a broader sense than just an examination of relative frequencies in experimental trials and involves realising that probability measures give us a degree of plausibility or belief which we can assign to any estimate or measurement. This should be based on a complete analysis of all available information including any background, or prior information. We have pointed out that all probabilities are subjective inasmuch as they are based on the currently known information but that nevertheless it is possible to ensure that they are unbiased by adhering to the rules of consistency throughout the analysis.

The core of the uncertainty propagation method developed is the general law of error propagation as presented in the ISO Guide. The theory underlying this has been considered and shown to be in agreement with the Unified Approach developed in this thesis. We have highlighted how it leads to a coherent means of expressing measurement uncertainties in a way that can easily be incorporated into other work as required. Much effort has been expended in developing a measurement philosophy based on the view of probability theory outlined in the preceeding paragraph, in order to show the usefulness and accuracy of the ISO approach as a means to describe experimental measurements.

A useful and indeed necessary extension to the ISO procedure has been the consideration of the Maximum Entropy formalism which greatly aids in the implementation of an unbiased analysis. Once again the logical consistency considerations of probability theory are central requirements to this development.

This framework of the ISO uncertainty propagation, supported by considerations of logical consistency and maximum entropy, is then applied to mass determination, looking initially at the model parameterisation of the experimental procedure. The various influence quantities are considered in developing the Weighing Equation and

its uncertainty propagation, in particular including the components of uncertainty due to the systematic buoyancy correction. In this analysis the principles of consistent reasoning were applied throughout, in order to ensure that all the information was included in an appropriate manner. The multivariate form of the Weighing Equation was presented and its corresponding covariance matrix developed, care being taken to point out the nature of the variances and covariances as statements about our degree of belief in the measurand estimates. The system so described provides a very concise and easily manipulated description of the relevant information.

The second major section of the theoretical development of the thesis has been concerned with parameter estimation techniques and it is here that we have seen the full implications of the Unified Approach, the groundwork for which has been laid in the early chapters. The requirement has been that the same criteria and philosophy which guided model parameterisation and uncertainty analysis should also be present in this section and this unification of ideas led to the rejection of conventionally used Restrained Least Squares, both because of its internal inconsistencies and also because of its failure to analyse the data in a manner compatible with the basic criteria being implemented. The significance of this failure has been brought to light by examples showing the rigidity of the technique: the constraints restrict the range of possible values that can be selected and in the case of incorrect prior information being provided, the estimator is not able to either overcome the problems or even adequately highlight them.

In dealing with this situation, it was essential to employ a method which would take full account of all known data about the problem in the correct manner. The Augmented Design approach fulfilled these requirements while giving extra benefits in the form of easier calculations and, especially, smaller posterior variance / covariances than the alternative Restrained Least Squares technique. This was seen to be a particularly pleasing improvement, easily understood in the light of the stochastic nature of all the available information.

The explicitly Bayesian approach of Maximum a Posteriori estimation was presented in order to make a clearer illustration of the nature of the problem. We pointed out the importance of realising that prior information is simply that which is logically distinct from the current experiment and showed how all probabilities are in some way conditional on some background or prior information. The application of the basic criteria of probability theory led directly to Bayes' theorem and a powerful estimation technique which takes full advantage of all that is known about the problem. The crucial requirement is the consideration of the necessary constraint

information as prior data having its own dispersion characteristics and covariance matrix.

An interlude in this development discussed a generalised estimation technique, Generalised Gauss-Markov estimation, where we saw how the essential differences between the Bayesian and Least Squares methods lie in the difference of interpretation of the constraint information. GGM can thus handle either situation without prejudice to the other. Prior data having a null covariance matrix would be treated just like constraints in the Least Squares approach, but no decision has been taken in so doing that prevents other data with a non-zero covariance matrix from being analysed in full.

In considering the Bayesian estimator we found an important role for the degrees of belief assigned to the various parameters and the relative accuracies between data sets (prior & current) in obtaining the estimated parameter values. We saw how this is particularly important in dealing with incorrect data, this having greater or lesser influence depending upon the Degree of Belief assigned to it. We also looked at the ranges of possible values and variances that could be assigned to the parameters and found worst case scenarios of the prior information remaining unchanged as well as lower limits to the improvements in accuracy that could be achieved in sequential estimation.

In discussing errors on the prior information, we raised the particularly important issue of drift on mass standards which could render the prior information invalid or irrelevant. Since the prior information can exert a significant influence in the Bayesian analysis this is an important issue. We pointed out how it is crucial in mass calibration to have some standards involved which are recently calibrated to the necessary accuracy and that a "within-group" calibration of a given set of standards, although possible with the MAP estimator, should not be consecutively repeated without including other external standards. Providing this is done, then implementing a regime of Consistent Analysis and including all available information should easily allow drift to be uncovered. We pointed out that the primary way to do this was by adjusting the prior variances of some of the prior information. Then if the evidence demanded it, the updated posterior values would be adjusted as required. We pointed out that suspected drift could be included prior to applying the estimator to the data but that doing this was less important than adjusting the prior degree of belief. However there can be a fundamental problem here since in mass determination there is no absolute independent external information currently available so the analyst must take care to supply physically relevant data and be aware of the limitations of the mathematical methods.

We conclude that the Unified Approach, incorporating a Consistent Analysis, and careful attention to the physical nature of the problem under consideration, applied throughout the data analysis in mass determination leads to much improved estimates as a result of implementing a better understanding of the true nature of the information being processed. Further, the inclusion of all available information at the time of analysis is crucial in ensuring the success of the endeavour.

What has been produced is a unique package for data analysis whereby a small set of matrix equations can completely model all the available data and provide improved estimates of the parameters. The technique is remarkable for its simplicity and reliability, the capability to highlight inconsistencies and errors in data being particularly attractive. Since the data is processed in a unified manner it can easily be incorporated into other analyses.

One important point that we have highlighted concerns the significance of the buoyancy variance-covariance information. It usually happens that the uncertainty contribution from this source is by far the largest element of the experimental covariance matrix. Thus it is very important not to neglect this information, and also motivation to look for higher accuracy ways to determine the volumes of the standards since as we have shown in our case studies, these are a large contribution to the covariance matrix.

In discussing how the Bayesian Estimator copes with the problem of drift affecting the prior information, we pointed out the ultimate need for truly independent information to introduce into the calibration hierarchy. There are currently various endeavours in progress that we have cited previously which have this goal in mind. Ultimately this data on a non-artefact kilogram realisation should be of higher accuracy than existing standards and the Bayesian Estimator will then easily allow it to exert a correspondingly greater influence on subsequent parameter estimations. Thus the Bayesian estimation technique promises to be a useful analytical tool in this work.

We have already considered, in the final chapter, the use of computerised data acquisition systems. These will grow in importance as automated and electronic comparators continue to grow in sophistication and become ever more widely used (see Glaeser *et al* (1992), Helms (1997), Kajastie *et al* (1997) for example.) We have pointed out how many systematic effects in the measurement process can be pinpointed by analysis of large quantities of data with the aid of computer power. The software used for data acquisition was specially produced as part of this research and therefore has been tailored to the exact needs of the laboratory. (A modified version of

the software is currently used by the National Metrology Laboratory, Dublin, and is discussed in one of the publications reproduced in Appendix 5). This is the ideal approach since the software is then perfectly traceable and accessible, a necessary feature to ensure high standards in calibration (For other examples of computerised calibration analysis systems see for example Kruh *et al* (1994) or Dikken (1997)).

It is clear that the concise package of experimental modelling and parameter estimation could easily be built into a software package which should be able to interface with the data acquisition software. In this way all the experimental data can be loaded into an analysis program which can then construct the necessary vectors and matrices and solve the parameter estimation equations.

Thus an apparently mature field like the calibration of mass standards shows much promise for interesting contributions to the field of uncertainty analysis in the future!

****

Appendix 1: The Partial Derivatives of the Air Density Equation

If we look at the functional form of the air density equation as given in Chapter 3 we see:

$$\rho_a = \frac{PM_a}{ZRT}\left(1 - x_v\left(1 - \frac{M_v}{M_a}\right)\right) = f(P, M_a, M_v, Z, R, T, x_v) \qquad \text{(A.1.1)}$$

However, from the equations in (3.1.18) & (3.1.19) we can see that in fact Z & x_v are both functions of T, P & h. So in fact:

$$\rho_a = f(T, P, h, R, M_a, M_v') \qquad \text{(A.1.2)}$$

To evaluate the partial derivatives, a *Mathematica* program was used in which the functional forms of all the equations needed for Eq. (A.1.1) were entered so that an explicit form of the function in terms of the six parameters in Eq. (A.1.2) was produced. *Mathematica* could then easily evaluate all the partial derivatives, themselves also functions of the six influence quantities. Thus if the known values of R, M_v & M_a were supplied along with measured values of T, P & h it would be easy to find the particular values of the partial derivatives at that point. Program listing A.1.1 at the end of this appendix shows a portion of the program used, including the function to calculate the partial derivative with respect to temperature. The complex expression shown is the result of the *Mathematica* evaluation, expressed in a form suitable for a C program. The various constants used are all defined in the header file and are obtained from the published information on the 'BIPM Air Density Formula', Giacomo (1981), Davis (1992a).

The data shown in Table A.1.1 below was obtained using this program with the values for t, P & h shown in the first column. For each set of three, the six partial derivatives are shown. The ranges of the parameters t, P & h taken are of typical order of magnitude for a Standards Laboratory and indeed cover most of the possible ranges that would likely be encountered. From the Table it can be seen that the data quoted in Eq. (3.2.3) is indeed representative of the likely values of the partial derivatives that would occur with standard laboratory data.

Table A.1.1 : Partial derivatives of the Air Density Equation for various values of t, P & h

t (°C) P (Pa) h (%RH)	$\frac{\partial \rho_a}{\partial t}$ $kg\ m^{-3} K^{-1}$	$\frac{\partial \rho_a}{\partial P}$ $kg\ m^{-3}\ Pa^{-1}$	$\frac{\partial \rho_a}{\partial h}$ $kg\ m^{-3}$	$\frac{\partial \rho_a}{\partial R}$ $kg\ m^{-3}\ J^{-1}\ mol\ K$	$\frac{\partial \rho_a}{\partial M_v}$ $mol\ m^{-3}$	$\frac{\partial \rho_a}{\partial M_a}$ $mol\ m^{-3}$
20 101325 50	-4.4×10^{-3}	1.189×10^{-5}	-10.5×10^{-3}	-0.144	0.4819	41.05
20 101325 30	-4.3×10^{-3}	1.189×10^{-5}	-10.5×10^{-3}	-0.144	0.289	41.297
20 101325 60	-4.5×10^{-3}	1.189×10^{-5}	-10.5×10^{-3}	-0.144	0.578	41.0
20 98000 50	-4.3×10^{-3}	1.189×10^{-5}	-10.5×10^{-3}	-0.139	0.4819	39.7
20 103000 50	-4.5×10^{-3}	1.189×10^{-5}	-10.5×10^{-3}	-0.147	0.482	41.79
19 101325 50	-4.4×10^{-3}	1.193×10^{-5}	-9.9×10^{-3}	-0.145	0.454	41.275
21 101325 50	-4.4×10^{-3}	1.185×10^{-5}	-11.0×10^{-3}	-0.144	0.511	40.93

Listing A.1.1

<HEADER FILE AIR_DEN.H>
==========

```
#ifndef __AIRDENH__
#define __AIRDENH__
#define A    1.2378847e-5
#define B   -1.9121316e-2
#define C    33.93711047
#define D   -6.3431645e+3
#define alpha    1.00062
#define beta     3.14e-8
#define gamma    5.6e-7
#define a0    1.58123e-6
#define a1   -2.9331e-8
#define a2    1.1043e-10
#define b0    5.707e-6
#define b1   -2.051e-8
#define c0    1.9898e-4
#define c1   -2.376e-6
#define d     1.83e-11
#define e    -0.765e-8
#define R     8.31451
#define Ma    28.9635
#define Mv    18.015
#define tAbs  273.15

extern double  enhance_fact(double, double);
extern double  vapour_press(double);
extern double  compress_fact(double,double,double);

extern double  air_den(double,double,double,double td = 0,double co2 = 0.0004);
#endif
```

****

<HEADER FILE MDEF.H>
==========

```
#define E   2.718281828

extern double Power(double x, double y);
```

****

<PROGRAM MFUNC.CPP>
==========

```
#include <math.h>
#include "header\mdef.h"

extern double Power(double x, double y)
{
  return pow(x,y);
}
```

****

<PROGRAM DERIV.CPP>
==========

/ Takes 3 arguments : Temp.(°C), Pressure (Pa) & Humidity (%) and evaluates the partial derivatives w.r.t the six parameters discussed in Chapter 3 & Appendix 1*
Only one of the functions (to calculate dp/dt) is reproduced here/*

```
#include <stdio.h>
#include <conio.h>
#include <stdlib.h>
#include <air_den.h>
#include <mdef.h>

double dpdt(double, double ,double);
double dpdP(double, double ,double);
double dpdh(double, double ,double);
double dpdR(double, double ,double);
double dpdMa(double,double,double);
double dpdMv(double,double,double);

void main(int argc , char *argv[])
{
 double P,t,h;
 char *end;

 t = strtod(argv[1], &end);
 P = strtod(argv[2], &end);
 h = strtod(argv[3], &end);
 h /= 100.0;

 printf("\n\nPartial Derivatives of the Air-Density Equation :>");
 printf("\n==========================================================");

 printf("\n\n%s%.8lf","Density wrt Temp.          (dpdT) =  :>  ",dpdt(t,P,h));
 printf("\n\n%s%.8lf","Density wrt Pressure       (dp/dP) =  :>  ",dpdP(t,P,h));
 printf("\n\n%s%.8lf","Density wrt Humidity       (dp/dh) =  :>  ",dpdh(t,P,h));
 printf("\n\n%s%.8lf","Density wrt Gas Const.     (dp/dR) =  :>  ",dpdR(t,P,h));
 printf("\n\n%s%.8lf","Density wrt Dry molar Mass  (dp/dMa) = :>  ",dpdMa(t,P,h));
 printf("\n\n%s%.8lf","Density wrt Moist Molar Mass(dp/dMv) = :>  ",dpdMv(t,P,h));
 getche();
}
```

```
double dpdt(double t, double P, double h)
{
 double deriv, ma, mv, T0, a, b, g;

 ma = Ma *1e-3;
 mv = Mv *1e-3;
 T0 = tAbs;  a = alpha; b = beta; g = gamma;

 deriv = -(ma*P*(1 - Power(E,C + D/(t + T0) + B*(t + T0) + A*Power(t + T0,2))*h*
        (1 - mv/ma)*(a + b*P + g*Power(t,2))/P)/(R*Power(t + T0,2)*(1 + Power(P,2)*
        (d + e*Power(E,2*C + 2*D/(t + T0) + 2*B*(t + T0) +2*A*Power(t + T0,2))*Power(h,2)*
        Power(a + b*P + g*Power(t,2),2)/Power(P,2))/Power(t + T0,2)-
        P*(a0 + a1*t + a2*Power(t,2) +(b0 + b1)*Power(E,C + D/(t + T0) + B*(t + T0) + A
        *Power(t + T0,2))*h*(a + b*P + g*Power(t,2))/P +(c0 + c1)*
        Power(E,2*C + 2*D/(t + T0) + 2*B*(t + T0) + 2*A*Power(t + T0,2))*
        Power(h,2)*Power(a + b*P + g*Power(t,2),2)/Power(P,2))/(t + T0)))) +
        ma*P*(-2*Power(E,C + D/(t + T0) + B*(t + T0) + A*Power(t + T0,2))*g*h*
        (1 - mv/ma)*t/P - Power(E,C + D/(t + T0) + B*(t + T0) + A*Power(t + T0,2))*h*
        (1 - mv/ma)*(a + b*P + g*Power(t,2))*(B - D/Power(t + T0,2) + 2*A*(t + T0))/P)/
        (R*(t + T0)*(1 + Power(P,2)*(d + e*Power(E,2*C + 2*D/(t + T0) + 2*B*(t + T0) +
        2*A*Power(t + T0,2))*Power(h,2)*Power(a + b*P + g*Power(t,2),2)/Power(P,2))/
        Power(t + T0,2)-P*(a0 + a1*t + a2*Power(t,2) +(b0 + b1)*Power(E,C + D/(t + T0) + B*
        (t + T0) +A*Power(t + T0,2))*h*(a + b*P + g*Power(t,2))/P +(c0 + c1)*
        Power(E,2*C + 2*D/(t + T0) + 2*B*(t + T0) +2*A*Power(t + T0,2))*Power(h,2)*
        Power(a + b*P + g*Power(t,2),2)/Power(P,2))/(t + T0))) -ma*P*(1 - Power(E,C + D/(t + T0)
        + B*(t + T0) + A*Power(t + T0,2))*h* (1 - mv/ma)*(a + b*P + g*Power(t,2))/P)*
        (-2*Power(P,2)*(d + e*Power(E,2*C + 2*D/(t + T0) + 2*B*(t + T0) + 2*A*
        Power(t + T0,2))*Power(h,2)*Power(a + b*P + g*Power(t,2),2)/Power(P,2))/
        Power(t + T0,3) + P*(a0 + a1*t + a2*Power(t,2) +(b0 + b1)*Power(E,C + D/(t + T0) + B*
        (t + T0) +A*Power(t + T0,2))*h*(a + b*P + g*Power(t,2))/P +(c0 + c1)*
        Power(E,2*C + 2*D/(t + T0) + 2*B*(t + T0) +2*A*Power(t + T0,2))*Power(h,2)*
        Power(a + b*P + g*Power(t,2),2)/Power(P,2))/Power(t + T0,2) -P*(a1 + 2*a2*t + 2*
        (b0 + b1)*Power(E,C + D/(t + T0) + B*(t + T0) + A*Power(t + T0,2))*g*h*t/P+ 4*
        (c0 + c1)*Power(E,2*C + 2*D/(t + T0) + 2*B*(t + T0) + 2*A*Power(t + T0,2))*g*
        Power(h,2)*t*(a + b*P + g*Power(t,2))/Power(P,2) +(b0 + b1)*Power(E,C + D/(t + T0) +
        B*(t + T0) +A*Power(t + T0,2))*h*(a + b*P + g*Power(t,2))*(B - D/Power(t + T0,2) +
        2*A*(t + T0))/P +(c0 + c1)*Power(E,2*C + 2*D/(t + T0) + 2*B*(t + T0) +2*A*Power(t +
        T0,2))*Power(h,2)*Power(a + b*P + g*Power(t,2),2)*(2*B - 2*D/Power(t + T0,2) + 4*A*
        (t + T0))/Power(P,2))/(t + T0)+ Power(P,2)*(4*e*Power(E,2*C + 2*D/(t + T0) + 2*B*
        (t + T0) + 2*A*Power(t + T0,2))*g*Power(h,2)*t*(a + b*P + g*Power(t,2))/Power(P,2) +
        e*Power(E,2*C + 2*D/(t + T0) + 2*B*(t + T0) +2*A*Power(t + T0,2))*Power(h,2)*
        Power(a + b*P + g*Power(t,2),2)*(2*B - 2*D/Power(t + T0,2) + 4*A*(t + T0))/
        Power(P,2))/Power(t + T0,2))/(R*(t + T0)*Power(1 + Power(P,2)*(d + e*Power(E,2*C +
        2*D/(t + T0) + 2*B*(t + T0) +2*A*Power(t + T0,2))*Power(h,2)*Power(a + b*P +
        g*Power(t,2),2)/Power(P,2))/Power(t + T0,2) -P*(a0 + a1*t + a2*Power(t,2) +(b0 +
        b1)*Power(E,C + D/(t + T0) + B*(t + T0) +A*Power(t + T0,2))*h*(a + b*P +
        g*Power(t,2))/P +(c0 + c1)*Power(E,2*C + 2*D/(t + T0) + 2*B*(t + T0) +2*A*
        Power(t + T0,2))*Power(h,2)*Power(a + b*P + g*Power(t,2),2)/Power(P,2))/(t + T0),2));
 return deriv;
}
```

Appendix 2. Data Processing Software

Here we illustrate the software used to perform the data analysis described in Chapter 11. Fig. (11.3), reproduced below as Fig. (A.2.1) described the main steps involved in the process. The data acquisition software of stage I has already been described in Chapter 11. Here we concentrate on the other stages.

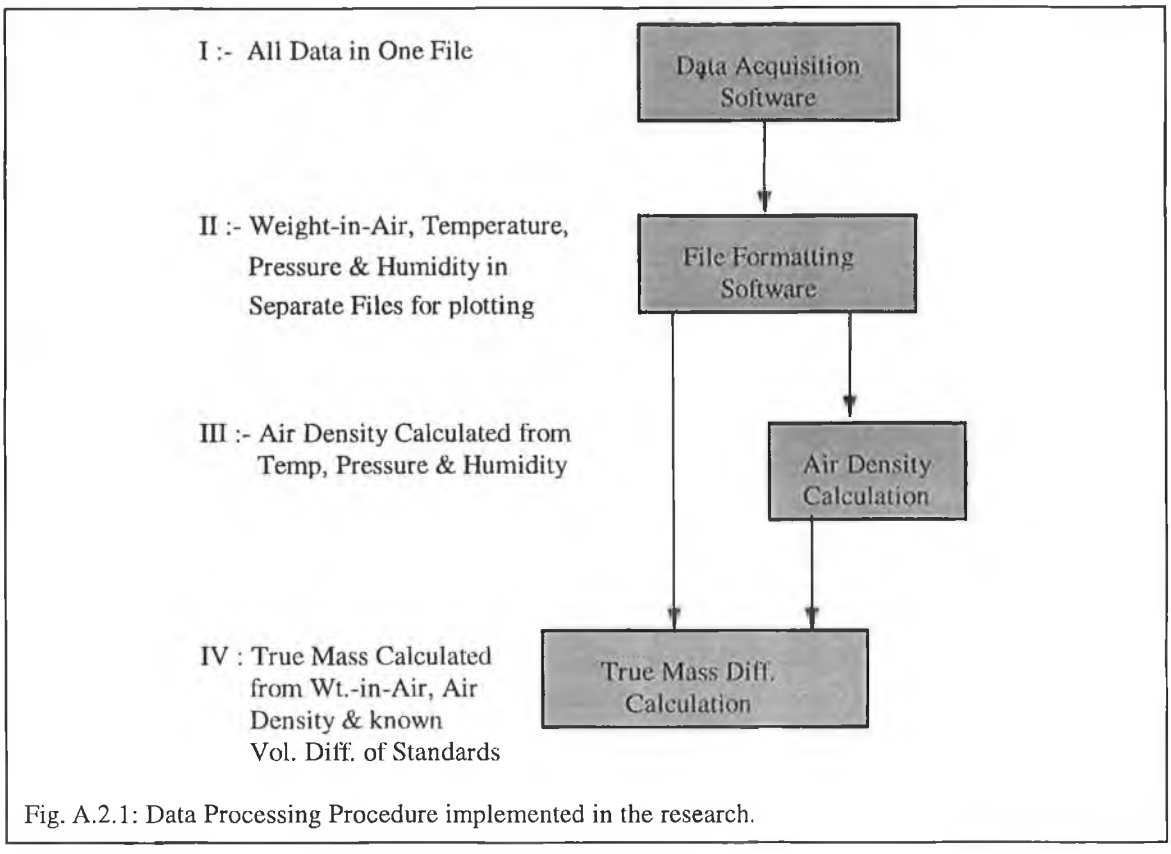


Fig. A.2.1: Data Processing Procedure implemented in the research.

Stage II: The data files produced by stage I, as shown in Fig. (11.2) must be processed before further analysis is possible. The 'raw data' files contain weight-in-air data, temperature, pressure and humidity data and a time sequence in seconds to allow graphing. Program *Explode* shown in Listing (A.2.1) takes such a data file and produces four output files: each containing the time data and one of the four influence quantities, temperature, pressure, humidity and weight-in-air data. From these the data can be plotted, Figs. (11.4) to (11.7) being typical examples of the graphs produced.

Stage III: It is then necessary to evaluate the air density, which is the input quantity required by the Weighing Equation (see Eq. (2.2.14) or the data-processing diagram, Fig. (3.0.1)). Program *density* shown in Listing (A.2.2) reads the data from three environmental data files produced in stage II above and writes out a data file of

air density values, calculated by implementing the "BIPM Equation" for air density. This too can be plotted, Fig (11.8) being an example.

Stage IV: Finally the desired measurand, the true mass difference for the comparison, can be computed. This is performed by the program *buoyancy* (Listing (A.2.3)) which reads the data from the corresponding weight-in-air and air density files and queries the user for the volume difference of the standards. Then a data file of true mass values is written out by applying the weighing equation to the data. Fig. (11.9) is an example graph produced by this means. From an analysis of the true mass graphs it is possible to see at what point the measurement process has stabilised and the data is then valid to use in the parameter estimation analysis. Program *buoyancy* can then be also run by specifying a selected number of data points from the graph to process, giving the mean values of weight-in-air, true mass and air density, as well as computing the variance of the weight-in-air data. This latter mode of operation is used to extract data to build up the various vectors and matrices for the analysis described in the main body of this work.

****

Listing A.2.1 - Program EXPLODE.CPP

```
#include <stdio.h>
#include <string.h>

void main(void)
{
 char fname[100] = "", wname[100] = "", tname[100] = "", pname[100] = "", hname[100] = "";

 // for i/p filename and 4 o/ps : weight, temp, pressure, humidity

 FILE  *fptr, *wptr, *tptr,*pptr,*hptr;
 double time,weight,tmp1,tmp2,tmp3, press1,press2,press3, hum1,hum2,hum3;

 do{
  printf("\n\nFileName :>  ");
  gets(fname);
 }while(((fptr = fopen(fname,"r")) == NULL) &&  printf("\n\nBad Filename !!"));
 printf("\n\nOutput Filename for Mass data :> ");
 gets(wname);
 wptr = fopen(wname,"w");
 printf("\n\nO/p Filename for Temperature Data :>  ");
 gets(tname);
 tptr = fopen(tname,"w");
 printf("\n\nO/P name for Pressure Data :>  ");
 gets(pname);
 pptr = fopen(pname,"w");
 printf("\n\nO/P filename for Humidity Data :> ");
 gets(hname);
 hptr = fopen(hname,"w");

 while(fscanf(fptr,"%lf%lf%lf%lf%lf%lf%lf%lf%lf%lf%lf",
                 &time,&weight,&tmp1,&tmp2,&tmp3,&press1,
                 &press2,&press3,&hum1,&hum2,&hum3) != EOF)
 {
  fprintf(wptr,"%lf   %lf\n",(time/3600.0),weight);
  fprintf(tptr,"%lf   %lf\n",(time/3600.0),(tmp1+tmp2+tmp3)/3.0);
  fprintf(pptr,"%lf   %lf\n",(time/3600.0),(press1+press2+press3)/3.0);
  fprintf(hptr,"%lf   %lf\n",(time/3600.0),(hum1+hum2+hum3)/3.0);
 }
 fcloseall();
 printf("\n\nFile Conversion Successful !");
}
```

Listing A.2.2 - Air Density Calculation Code
Header File AIR_DEN.H used by program AIR_DEN.CPP

```
#ifndef __AIRDENH__
#define __AIRDENH__
#define A    1.2378847e-5
#define B   -1.9121316e-2
#define C    33.93711047
#define D   -6.3431645e+3
#define alpha    1.00062
#define beta     3.14e-8
#define gamma    5.6e-7
#define a0    1.58123e-6
#define a1   -2.9331e-8
#define a2    1.1043e-10
#define b0    5.707e-6
#define b1   -2.051e-8
#define c0    1.9898e-4
#define c1   -2.376e-6
#define d     1.83e-11
#define e    -0.765e-8
#define R     8.31451
#define Ma    28.9635
#define Mv    18.015
#define tAbs  273.15

extern double  enhance_fact(double, double);
extern double  vapour_press(double);
extern double  compress_fact(double,double,double);

extern double  air_den(double,double,double,double td = 0,double co2 = 0.0004);
#endif
```

Program DENSITY.CPP

```
#include <stdio.h>
#include <conio.h>
#include <string.h>
#include <air_den.h>

void main(int argc, char *argv[])
{
 char   pname[100] = "", hname[100] = "", tname[100] = "", dname[100] = "";
 FILE   *pptr, *hptr, *tptr, *dptr;

 double  density, temp,press,hum,time,sto;
 strcpy(tname,argv[1]);
 strcpy(pname,argv[2]);
 strcpy(hname,argv[3]);
 printf("\n\n\nDensity Calculating Program : Enter o/p FileName :> ");
 gets(dname);

 tptr = fopen(tname,"r");
 pptr = fopen(pname,"r");
 hptr = fopen(hname,"r");
 dptr = fopen(dname,"w");

 while(fscanf(pptr,"%lf  %lf\n",&time,&press) != EOF)
 {
  fscanf(tptr,"%lf  %lf\n",&sto,&temp);
  fscanf(hptr,"%lf  %lf\n",&sto,&hum);
  density = air_den(temp,(press * 100), (hum / 100));
  fprintf(dptr,"%lf  %lf\n",time,density);
 }
 fcloseall();
 printf("\n\n\nDONE!!!!!");
}
```

Program AIR_DEN.CPP

```
#include <stdio.h>
#include <math.h>
#include "header\air_den.h"

extern double air_den(double t, double p, double h, double td, double co2)
{
  double f, psv, xv, density, z, T = t+tAbs;
  double fd, psvd;
  double M = ( Ma + 12.011 * ( co2 - 0.0004 ) ) * 1e-3;

  f  = enhance_fact(p,t);
  psv = vapour_press(t);
  if(td == 0)
    xv  = (h * f * psv) / p;
  else{
    fd  = enhance_fact(p,td);
    psvd = vapour_press(td);
    xv  = (fd*psvd)/p;
  }
  z  = compress_fact(p,t,xv);
  density = p / ( z * T) * ( 1 - ( 0.378 * xv )) * ( M / R) ;
  return density;
}

//*****************************************************************************

extern double enhance_fact(double p, double t)
{
 double f;

 f = alpha + ( beta * p ) + ( gamma * t * t );
 return f;
}

//*****************************************************************************

extern double vapour_press(double t)
{
 double arg, vp, T = t+tAbs;

 arg = (A*T*T)+(B*T)+C+(D/T);
 vp  = exp(arg);
 return vp;
}

//*****************************************************************************

extern double compress_fact(double p, double t, double xv)
{
 double arg1, arg2, z, T = t+tAbs;

 arg1 = a0+(a1*t)+(a2*t*t)+((b0+b1*t)*xv)+((c0+c1*t)*xv*xv);
 arg2 = (p*p*(d+e*xv*xv))/(T*T);
 z = 1 - ((p/T)*arg1) + arg2;
 return z;
}
```

Listing A.2.3 - Program BUOYANCY.CPP

```
#include <stdio.h>
#include <stdlib.h>
#include "header\air_den.h"
#include <math.h>
#include <string.h>
#include <conio.h>

double round(double);
void data1(FILE *, FILE *,double);
void data2(int, int, double, FILE *, FILE *);

void main(int argc, char *argv[])
{
 char   mode;
 int    start, finish;
 double dvol;
 char   mname[100] = "", dname[100] = "";
 FILE   *mptr, *dptr;

 if(argc < 3 ){
  printf("\n\n Usage :> buoyancy file1 file2 "); exit(1);
 }
 strcpy(mname,argv[1]);   strcpy(dname,argv[2]);
 mptr = fopen(mname,"r");   dptr = fopen(dname,"r");
 printf("\n\nBuoyancy Correction Program : Enter Vol. Diff (cm^3) :> ");
 scanf("%lf",&dvol);
 printf("\n\nGraph Mode (1) or Data Mode (2) ??  :>  ");
 mode = getche();
 if(mode == '2')
 {
  printf("\n\nEnter start data point :>   ");
  scanf("%d",&start);
  printf("\n\nEnter finish data point :>   ");
  scanf("%d",&finish);
  data2(start,finish,dvol,mptr,dptr);
 }
 else
  data1(mptr,dptr,dvol);
 fcloseall();
 printf ("\n\n\nDONE !!");
}

void data1(FILE *mptr, FILE *dptr, double dvol)
{
   double time, mass, sto, meanDiff, density;
   FILE *opptr;
   char opname[100] = "";

   printf("\n\nEnter o/p file name :>   ");
   gets(opname);
   gets(opname);
   opptr = fopen(opname,"w");
   while(fscanf(mptr,"%lf  %lf\n",&time, &meanDiff) != EOF)
   {
     fscanf(dptr,"%lf  %lf\n",&sto,&density);
     mass = meanDiff + (density * dvol * 1000);
```

```
      mass = round(mass);
      printf("\n%s%lf%s%lf%s%lf","Density :> ",density," Weight :> ",meanDiff," Mass :>",mass);
      fprintf(opptr,"%lf  %lf\n",time,mass);
    }
  }

void data2(int start, int finish, double dvol, FILE *mptr, FILE *dptr)
{
  int pos, i;
  double meanDiff, density, mass,  time;
  double Sx_w, Sxx_w, xbar_w, var_w;
  double Sx_d, Sx_m, xbar_d, xbar_m;

  pos = 1; i = 0;
  Sx_w = Sxx_w  = xbar_w = var_w = Sx_d = Sx_m = xbar_m = xbar_d = 0;
  printf("\nWt-in-air    Air-Den    Mass");
  printf("\n=======================================");
  while(fscanf(mptr,"%lf  %lf\n",&time, &meanDiff) != EOF)
  {
    fscanf(dptr,"%lf  %lf\n",&time,&density);
    if(pos < start ) { ++pos; continue; }
    else{
      mass = meanDiff + (density * dvol * 1000);
    }
    ++pos; ++i;
    Sx_w += meanDiff;  Sx_d += density; Sx_m += mass;
    Sxx_w += pow(meanDiff,2);
    xbar_w = Sx_w/i; xbar_d = Sx_d/i; xbar_m  = Sx_m/i;
    if(i > 1)
      var_w = fabs(Sxx_w - (i * pow(xbar_w,2)))/(i-1);
    printf("\n%lf    %lf   %lf",meanDiff,density,mass);
    if(pos > finish) break;
  }
  printf("\n\n%s%lf","Mean, wt.in air :>    ",xbar_w);
  printf("\n%s%lf","Variance, Wt. in Air :>  ",var_w);
  printf("\n%s%lf","Mean, Air Den. :>  ",xbar_d);
  printf("\n%s%lf","Mean, True Mass :>   ",xbar_m);
  printf("\n%s%d","Number Data Points :>   ",i);
}

// This function rounds a number to the nearest integer

double round(double value)
{
  double result, mantissa;
  int decimal;

  decimal = int(value);
  mantissa = value - decimal;
  if(mantissa > 0.5) result = (double)decimal + 1.0;
  else if (mantissa < -0.5 ) result = (double) decimal - 1.0;
  else result = (double)decimal;
  return result;
}
```

Appendix 3: Glossary of Selected Terms

Apparent Mass: The result of an 'in-air' mass determination—by any measurement means—but prior to applying any corrections for buoyant forces, centre-of-gravity differences, etc.

Augmented Design: An estimation technique, based on the Least Squares criterion, or on the Gauss-Markov theorem in which both prior and *current information*[1] is treated as data having corresponding *covariance matrices*. These two sets of data are augmented into one extended set which can then be adjusted by either of the two criteria mentioned.

Bayesian Estimation: A probabilistic estimation technique using Bayes' Theorem—i.e. incorporating *conditional probabilities* and *prior information* with new *current data* in order to obtain updated posterior estimates.

BIPM Formula: An equation for the determination of the density of air, recommended by the Comité International des Poids et Mesures in 1981 and updated in 1991, used by the Bureau International des Poids et Mesures and most national laboratories.

BLUE: Best Linear Unbiased Estimator: An estimator which combines the characteristics of minimum variance, is unbiased, and a linear combination of the observations.

Buoyancy Correction: An *apparent mass* difference between two standards resulting from the different densities of their constituent materials and hence the standards experiencing different buoyant forces.

Combined Difference Standard Deviation: This is the root-sum-square of the respective standard deviations of the terms involved in calculating a *difference* or *residual vector*.

Combined Standard Uncertainty: A root-sum-square of *uncertainty contributions* due to the various *influence quantities* featuring in a *functional relationship*.

Comparison Calibrations: A calibration method in which residual differences between nominally equal quantities (mass, length etc.) are measured with resulting greater accuracy than could be achieved by measuring absolute values.

---

[1] Terms in *italics* are defined elsewhere in this Glossary.

Conditional Probability Distribution: A distribution describing a parameter or set of parameters that are subject to the existence of some other parameter, set of parameters, or hypothesis. The dependency represents logical rather than causal connections.

Constraint: *Prior information* which is treated as a deterministic quantity. A solution to a set of consistent equations for some parameter values to be found by either algebraic manipulation or statistical adjustment must be subject to the condition that the prior information remains unchanged.

Constraint Contribution: The true contribution to the *covariance matrix* of a parameter estimation due to the *constraints* that were applied to the estimation process. This information is however excluded from the estimation analysis in any technique that uses constraints (e.g. RLS) and must be added afterwards.

Conventional Mass: A term not widely used in the thesis but in common currency in mass metrology as a convenient approximation. The conventional mass of a standard is defined as being the mass of another standard which would exactly counter-balance it in air of density $1.2\ kg/m^3$. This other standard must further have a density of $8000\ kg/m^3$.

Corrected Realised Quantity: Value generated by the *functional relationship*: this is the *measurand* estimate.

Covariance Matrix: An important quantity in multivariate statistical analysis. It is a matrix in which diagonal terms give the variances of the elements of a corresponding data vector and the off-diagonal terms give the covariance between pairs of data elements.

Criteria of Consistency: Important tenets of Classical Probability Theory which indicate that all solutions to a problem should lead to the same result, that all available evidence relevant to a problem should be brought to bear on establishing its solution and that equivalent states of knowledge should be represented by equivalent probabilities / plausibilities. These criteria underlie much of the Unified Approach to data analysis.

Current Information / Data: New data obtained by the comparison experiment, to be used along with the *prior information* in obtaining parameter estimates.

Degrees of Belief: A statistical / probabilistic statement about the accuracy of a parameter estimate which has been constructed based on all known and available information.

Design Matrix: An *n x p* matrix indicating the form of the *n* comparisons to be carried out among the *p* parameters. The elements of this matrix are either *1, 0* or *-1* depending upon the role of each parameter in the particular comparison.

Deterministic Parameter: A parameter 'estimate' considered to be a constant and thus to have a zero variance and no covariance with any other parameters.

Difference Vector: The difference between the prior and posterior parameter estimates.

Dispersion Characteristic: A model for the uncertainty estimate of a parameter, incorporating a probability distribution function and a variance estimate.

Experimental Errors: Unknown contingencies / influences outside the scope of the model parameterisation but nevertheless effecting the observed outcome of an experiment (in an unknown way). In the *Unified Approach* these unknowns cannot be modelled but instead probability distributions are assigned to the *measurand* estimates to parameterise the resulting plausibility of the determined value.

Extended Model: A system model used with *GGM*, like that used in the *Augmented Design*, in which *prior information* is included in the analysis. The prior information takes the form of restraints which can be either *stochastic* or *deterministic*. The GGM method can deal with either situation.

Expectation Value: A "mean value", or value of highest probability in a distributed set of data, described by some probability density function.

Functional Relationship: The mathematical model for realising an estimate of the *measurand* from the known *influence quantities* and direct measurement data.

GGM: Generalised Gauss-Markov Method: A parameter estimation technique based on finding a minimum variance estimate without any restrictions on the quantities involved. It implements the method of generalised inverses and allows the use of an *extended model* in describing the data.

GM Theorem: Gauss-Markov Theorem: Proves a *BLUE* estimator exists. It is more general then Least Squares, among its features being that it does not require all data to be of equal variance as does Least Squares.

Gaussian Procedure: Usually based on a Taylor series expansion of the *functional relationship*. From this the contribution of the variances/covariances of the influence quantities to the final *measurand* estimate can be calculated.

General Law of Error Propagation: The term usually used in the ISO Guide to describe the procedure for evaluating uncertainties in a unified and coherent manner. Synonymous with the *Gaussian Procedure* above.

Gravitational Correction: An *apparent mass* difference resulting from a difference in heights of the centres of mass of two standards. It is one of the corrections to the *weight-in-air* difference that must be applied, via the *weighing equation*, to obtain the mass difference.

Inductive Logic: The process of inferring among several possible causes the most likely for a given set of observed effects.

Influence Quantities: Secondary or 'systematic' parameters whose influence must be included in the parameterisation needed to obtain the *measurand* estimate.

Jacobian: Mathematically, the Jacobian is the determinant of a matrix of partial derivatives, formed from a set of functions $f_1,..., f_n$, each of which contains *influence quantities* $u_1,...., u_m$. It is fully called the "Jacobian of $f_1,..., f_n$ with respect to $u_1,...., u_m$.

Likelihood Function: A *conditional* distribution describing the function which would generate the parameters needed to give an already observed distribution of data.

Linear Unbiased Estimator: An estimator which is a linear combination of the corrected observations and is an unbiased estimate of the unknown 'true' values of the parameters.

MAP: Maximum a Posteriori Estimation: This estimator combines the *MLE* criterion with Bayes's theorem to produce a *posterior estimate* from a set of *current data* and any available *prior data*. It allows the possibility of recursive and *sequential estimation* of parameters.

Mass Difference: The corrected result of a *comparison calibration* between a pair, or ensemble, of mass standards.

Mass Value (Physical Mass): The absolute value of a mass standard, in high accuracy work to be determined by statistical adjustment of a set of mass differences by some parameter estimation technique.

Maximum Entropy: (MaxEnt): The Entropy of a Probability Distribution is a numerical measure of the 'Uncertainty' it represents with respect to a parameter estimate. The Maximum Entropy approach involves maximising this uncertainty subject to the constraints of definitely known information. Such a technique ensures a 'maximally unbiased' estimate from a given set of *subjective* information.

Measurand: The specific physical quantity subject to measurement.

Minimum Variance Estimate: A parameter estimate, obtained from the measurement data and uncertainties, which has the lowest uncertainty among all possible estimates that could be produced by mathematical means.

MLE: Maximum Likelihood Estimation: An estimation method based on an analysis of probability distribution functions which seeks to maximise the *likelihood function* for the parameters, given the observation data which was obtained. Thus it is a technique which deals only with the observed data.

Over-Determined System: A design for a calibration experiment in which more comparisons are carried out than are needed to find a solution. The extra redundant information provides degrees of freedom for a statistical adjustment of the parameters.

Posterior Information: An estimate (with corresponding *covariance matrix* or probability distribution) formed by combining *prior information* with new experimental data on a parameter or set of parameters. The 'combining' takes the form of a *BLUE* adjustment.

Prior Information: Data about the parameters involved in the estimation which is available before the experiment is carried out. In the case of mass determination such information is necessary to obtain a particular solution.

Probability Density Function: A distribution function describing the range of likely values which would occur in attempting to estimate some parameter. In principle every measurement or data value (which is an estimate subject to unknown errors) can be described by some such distribution.

Realised Quantity: That which is directly obtained from the measurement and generally is not the measurand desired.

Relative Accuracy: In comparing two sets of information, or individual members of a set, the relative accuracy, or relative magnitude of their respective variances / covariances is important in predicting their influence in any parameter estimation.

Residual Vector: The difference between the experimental data vector and the estimated experimental data based on the *posterior parameter estimates*.

RLS: Restrained Least Squares: The standard LS criterion (possibly modified by *WLS*) but subject to a set of *constraints*, linearly independent of the observational equations. The method of Lagrange Multipliers is used to solve the LS criterion subject to these constraints.

Robustness: Applied to a description of any estimator in this thesis; a robust estimator is one whose *posterior* (or output) *estimate* is stable in the event of perturbations or inaccuracies in some of the *prior information* or *constraints*.

Sensitivity Coefficient: The partial derivative of a *measurand* estimate with respect to one of the *influence quantities*, at specified values of other influence quantities, if relevant.

Sequential Estimation: An extension of the *MAP* Estimator in which *the posterior estimates* form the *prior data* for a subsequent estimation with new (chronologically later) experimental data.

Standard Uncertainty: The positive square root of a variance. This quantity is to be used as a fundamental measure of the uncertainty associated with a parameter estimate.

Statistical Estimation Techniques: Methods for processing data from an over-determined measurement system in order to carry out an adjustment to get the 'best' (or minimum variance) estimate of the parameters.

Stochastic Parameter: A parameter estimate with an assigned *Degree of Belief*, i.e. given *variance* and covariance terms.

Subjective Estimate: Any estimate is subjective in the sense that the knowledge available at the time is finite. Later work may present new information, forcing an update to the current estimate. In other words, analogous to *Conditional Probabilities*, any parameter estimate is conditional on the background knowledge available at the time.

System Model: The relationship between the corrected observation data and the parameters to be determined, usually dictated by the *design matrix*.

Unbiased Estimate: The result of an analysis (an estimate) is unbiased if all the known information relevant to the problem has been utilised in a manner consistent with the basic criteria of the *Unified Approach*.

Uncertainty Contribution: The influence of the variance of an individual influence quantity on the final variance estimate of a *measurand*. This is dictated by its role in the *functional relationship* and is defined as a product of its *sensitivity coefficient* and *standard uncertainty*.

Unified Approach: A complete formalism for data analysis in which all quantities are treated equally, all available information is used and unavailable data is not considered or assumed. This approach is modelled on the essential criteria of probability theory and adherence to these *Criteria of Consistency* is considered a basic test for any analytical method in the Unified Approach.

Variance: A characteristic quantity of a probability distribution, used in describing its "width" or "spread". Hence it is an integral part of quantifying the uncertainty of an estimate.

WLS: Weighted Least Squares: A method for avoiding the problem of non-equal variances which prevents Least Squares from producing a minimum variance estimator. This method produces the same estimate as does the *GM Theorem*.

Weight-in-Air: The uncorrected result of a *comparison calibration* carried out in air, without corrections for *buoyancy*, centre of mass differences, volume expansion coefficients etc., having been applied.

Weighing Equation: The *functional relationship* among the complete set *of influence quantities* needed to generate the corrected mass difference of a pair of standards. It can be expressed in scalar or multivariate notation.

Within-Group Comparisons: The determination of the *mass values* of a set of standards by intercomparison among themselves, usually in an *over-determined* manner; the value of at least one of them being known from previous calibration.

****

Appendix 4: Bibliography & Citation List

Allan, D. W. (1987) *Should the Classical Variance be used as a Basic Measure in Standards Metrology?*. IEEE Trans. Inst. Meas. **IM-36**, No. 2 pp 646 - 654.

Allisy, A. *Some Statistical Methods used in Metrology.* (Publication details not to hand)

Almer, H. E. & Swift, H. F. (1975) *Gravitational Configuration Effect upon Precision Mass Measurements.* Rev. Sci. Inst. **46**, pp 1174 - 1176.

Arri, E. (1996) *Performing measurements through 'correct metrological procedures,* in: Advanced Mathematical Tools in Metrology II, P. Ciarlini et al, Eds, World Scientific, Singapore.

Baierlein, R. (1994) *Entropy and the second law: a pedagogical alternative"* Am. J. Phys. **62**, pp 15 - 26.

Ballantine, J. (1996), *The Elusive Properties of Magnetism in Mass Standards,* CalLab, July-August 1996, pp 42 - 45.

Bayes, T. (1763) *An Essay toward solving a Problem in the Doctrine of Chances*, Phil. Trans. Roy. Soc. **53**, p 370.

Beck, J. V. & Arnold, K. J. (1977) *Parameter Estimation in Science & Engineering.* Wiley, New York.

Becker, P. et al *The Silicon Path to the Kilogram* PTB Mitteilungen **106**, pp 331 - 329.

Bego, V. et al, (1995) *Voltage Balance for Replacing the Kilogram*, IEEE Trans. Inst. Meas. **44**, pp 579 - 582

Bernieri, A. (1996) *Statistical Problems in Calibration Design,* in: Advanced Mathematical Tools in Metrology II, P. Ciarlini et al, Eds, World Scientific, Singapore.

Bich, W. (1989a) *European Mass Intercomparison (Euromet Project No. A88/143).* IMGC Rapporto P152.

Bich, W. (1989b) : *IMGC Report P162*, a Supplement to Report P152.

Bich, W. (1990) *Variances, Covariances and Restraints in Mass Metrology.* Metrologia **27**, pp 111 - 116.

Bich, W. (1992) *Bias and Optimal Linear Estimation in Comparison Calibrations.* Metrologia **29**, pp 15 - 22.

Bich, W. et al. (1993a): *Euromet Project 231 on Uncertainty in Mass Metrology. The Assignment of Uncertainties to the NPLW43 set of Reference Standards : A Comparative Study of the different procedures.* IMGC Report P199.

Bich, W. (1993b) *Bayesian and Fisherian Estimation in the Recalibration of a Set of Mass Standards*, in Proceedings of the 13th IMEKO Force / Mass Measurement Conference, Helsinki, Finland, pp 175 - 185.

Bich, W. et al (1993/94) *Uncertainty Modelling in Mass Comparisons*. Metrologia **30**, pp 495 - 502.

Bich, W. (1996) *Simple Formula for the Propagation of Variances & Covariances.* Metrologia **33**, pp 181 - 183.

Bich, W. (1997) *The ISO Guide on Uncertainty Measurement: A Bridge between Statistics and Metrology* in: *Advanced Mathematical Tools in Metrology III*, World Scientific, Singapore.

Bohm, D. (1984) *Causality & Chance in Modern Physics*, Routledge & Kegan Paul, London.

Bölöni, B. (1997) *Increase of Information in the Course of Measurement* in *Advanced Mathematical Tools in Metrology III*, World Scientific, Singapore.

Bose, R. C. & Cameron, J. M. (1967) *Calibration Designs Based on Solutions to the Tournament Problem*, J. Res. NBS. **71B**, No. 4, pp 149 - 160.

Bowman, H. A. et al (1967) *Procedure for High Precision Density Determinations by Hydrostatic Weighing,* J. Res. NBS. **71C**, pp 179 - 198.

Bowman, H. A. et al (1974a) *The Utilization of Solid Objects as Reference Standards in Density Measurements,* Metrologia **10**, 117 - 121.

Bowman, H. A. et al (1974b) *A Density Scale Based on Solid Objects*, J. Res. NBS, **78A**, pp 13 - 40.

Bretthorst, G. L. (1989) *An Introduction to Parameter Estimation using Bayesian Probability Theory*, pp 53 - 79 in: *Maximum Entropy & Bayesian Methods*, Dartmouth, Kluwer Academic Publishers, Netherlands.

Buck, B. & Macaulay, V. A. (Eds) (1991) *Maximum Entropy in Action*, Clarendon Press, Oxford.

Cameron, J. M. et al (1977) *Designs for the Calibration of Standards of Mass.* NBS Technical Note 952.

Cohen, E. R. & Taylor, B. N. (1987) *The 1986 Adjustment of the Fundamental Physical Constants.* J. Res. NBS **92**, pp 85 - 95.

Cohen, E. R. & Tuninsky, V. S. (1995) *The Extended Least-Squares Treatment of Correlated Data*, IEEE Trans. Inst. Meas **44**, No. 2, pp 475 - 479.

Colclough, A. R. (1987) *Two Theories of Experimental Error.* J. Res. NBS **92**, pp 167 - 185.

Cox, M. G. (1996) *Constructing and Solving Mathematical Models of Measurement.* in: Advanced Mathematical Tools in Metrology II, P. Ciarlini et al, Eds, World Scientific, Singapore.

Cox, R. T. (1946), *Probability, Frequency and Reasonable Expectation,* Am. J. Physics **14,** pp 1 - 13.

Croarkin, C. (1989) *An extended error model for Comparison Calibration* Metrologia **26**, pp 107 - 113.

Cumpson, P. J. & Seah, M. P. (1994) *Stability of Reference Masses I: Evidence for Possible Variations in the Mass of Reference Kilograms Arising from Mercury Contamination.* Metrologia **31**, pp 21 - 26.

Cumpson, P. J. & Seah, M. P. (1994/95) *Stability of Reference Masses III: Mechanism and Longterm Effects of Mercury Contamination on Platinum-Iridium Mass Standards*, Metrologia **31**, pp 375 - 388.

Cumpson, P. J. & Seah, M. P. (1996) *Stability of Reference Masses IV: Growth of Carbonaceous Contamination on Platinum-Iridium alloy surfaces, and Cleaning by UV/Ozone treatment.* Metrologia **33**, pp 507 - 532.

David, F. N. (1962), *Games, Gods and Gambling,* Charles Griffin, London.

Davis, R. S. (1982) *Determination of Silicon Density to High Precision Using a Submersible, Servo-Controlled Balance*, Metrologia **18**, pp 193 - 201.

Davis, R. S. (1985) *Recalibration of the U.S. National Prototype Kilogram.* J. Res. NBS **90,** No. 4, pp 263 - 283.

Davis, R. S. (1987) *Note on the Choice of a Sensitivity Weight in Precision Weighing.* J. Res. NBS. **92**, No. 3, pp 151 - 172.

Davis, R. S. (1989) *The Stability of the SI Unit of Mass as Determined from Electrical Measurements,* Metrologia **26**, pp 75 - 76.

Davis, R. S. (1990) *New Assignment of Mass Values and Uncertainties to NIST Working Standards,* J. Res. Natl. Inst. Stand. Tech. **95,** pp 79 - 92.

Davis, R. S. (1992a) *Equation for the Determination of the Density of Moist Air (1981 / 91).* Metrologia **29**, pp 67 - 69.

Davis, R. S. (1992b) *Using Small, Rare-Earth Magnets to study the Susceptibility of Feebly Magnetic Materials.* Am. J. Phys. **60**, pp 365 - 370.

Davis, R. S. (1993) *New Method to Measure Magnetic Susceptibility* Meas. Sci. Tech. **4**, pp 141 - 147.

Davis, R. S. (1995a): *Determining the Magnetic Properties of 1kg Mass Standards.* J. Res. NIST. **100**, No. 3, pp 209 - 225.

Davis, R. S. (1995b): *Device to locate the Centre of Mass of a Test Object to within a precision of Micrometres.* Meas. Sci. Tech. **6**, pp 227 - 229

Debler, E. (1994) *A theoretical approach to measurement: Uncertainty of mass, Conventional mass & Force.* OIML Bulletin **XXXV**, No. 4, Oct. 1994, pp 9 - 14.

Dieck, R. H. (1997) *Measurement Uncertainty Models,* ISA Trans. **36,** No. 1, pp 29 - 35.

Dikken, D. A. (1997) *An Approach to Automating Mass Metrology Systems*, pp 47 - 54 in: *Weighing and Measurement in the Year 2000*, Int. Conf. 16 - 17 Sept. 1997, South Yorkshire Trading Standards Unit, Sheffield, England.

Eadie, W. T. et al (1971) *Statistical Methods in Experimental Physics*. North Holland

EAL (1997) *Expression of Uncertainty of Measurement in Calibration*, EAL Document R2.

Emery, A. F. & Fadale, T. D. *Design of Experiments Using Uncertainty Information* Trans. ASME **118**, pp 532 - 538.

Everhart, J. (1995) *Components of Mass Measurement Uncertainties and the Methods of Uncertainty Determination,* Measurement Science Conference, Jan. 26 - 27 1995, Anaheim, Ca.

Forbes, A. (1996) *Model Parameterisation,* in: Advanced Mathematical Tools in Metrology II, P. Ciarlini et al, Eds, World Scientific, Singapore.

Frantsuz, E. T. (1996) *A proposed Superconducting Magnetic Levitation System intended to monitor stability of the unit of mass.* in: Conference on Precision Electromagnetic Measurements Digest, Braunschweig, Germany, 17 -21 June 1996, pp 189 - 196.

Fritz, M. (1995) *Applying the ISO Guide to the Espression of Uncertainty in Measurement in a Mass Calibration Laboratory.* Measurement Science Conference, Jan. 26 - 27 1995, Anaheim, Ca.

Fröhner, F. H. (1997), *Assigning Uncertainties to Scientific Data,* Nuclear Science and Engineering, **126,** pp 1 - 18.

Gelman, A. *et al* (1995) *Bayesian Data Analysis,* Chapman & Hall, London.

Giacomo, P. (1981) *Equation for the Determination of the Density of Moist Air* Metrologia **18**, pp 33 - 40.

Girard, G. (1990) *The Washing & Cleaning of Kilogram Prototypes at the BIPM*, BIPM, Paris.

Girard, G. (1994) *The third Periodic Verification of National Prototypes of the Kilogram (1988 - 1992).* Metrologia **31**, pp 317 - 336.

Glaeser, M (1990) *Response of Apparent Mass to Thermal Gradients,* Metrologia 1990, **27**,pp 95 - 100.

Glaeser, M. et al (1992): *Automation of High Accuracy Weighing at some Western European National Measurement Laboratories,* PTB-Mitteilungen **102,** pp 163 - 171.

Glaeser, M. & Do, J. Y. (1993): *Effect of Free Convection on the Apparent Mass of 1kg Mass Standards,* Metrologia **30**, pp 67 - 73.

Godec, Z. (1997), *Standard Uncertainty in each Measurement Result Explicit or Implicit,* Measurement **20,** No. 2, pp 97 - 101.

Goldman, A. J. & Zelen, M. (1964) *Weak Generalised Inverse and Minimum Variance Linear Unbiased Estimation*, J. Res. NBS. **68B**, No. 4, pp 151 - 172.

Grabe, M. (1978) *Note on the Application of the Method of Least Squares.* Metrologia **14**, pp 143 - 146.

Green, J. R. & Margerison, D. (1978) *Statistical Treatment of Experimental Data*, Elsevier Scientific.

Greenspan, L. (1976), *Functional Equations for the Enhancement Factors for CO_2-Free Moist Air.* J. Res. NBS. **80A**, No. 1, pp 41 - 44.

Heierli, R. (1997) *High Accuracy Volume Determination for Weights up to 1kg*, pp 27 - 36 in Proceedings of the International Conference, *Weighing and Measurement in the Year 2000*, South Yorkshire Trading Standards Unit, Sheffield, England, Sept. 16 - 17, 1997.

Helms. A. (1997) *Automatic Control of 20 kg & 50 kg Weights in Large Numbers* Bulletin OIML **XXXVIII,** No. 2, pp 5 - 7.

Hughes, C. G. & Musk, H. A. (1972): *A Least Squares Method for Analysis of Pair Comparison Measurements*. Metrologia **8**, pp 109 - 113.

Hyland, R. W. (1975), *Correlation for the Second Interaction Virial Coefficients & Enhancement Factors of Moist Air,* J. Res. NBS. **79A**, No. 4, pp 551 - 560.

Ikeda, S. (1986a), *A Measure to Estimate the Effectiveness of Weight Series & Weighing Designs for the minimisation of propagated adsorbtion errors.* Metrologia **22**, pp 69 - 74.

Ikeda, S. & Kobayashi, Y. (1986b) *Decimal Weight Series for submultiples of mass standards to decrease Adsorbtion Error using "Simple Comparisons"* Metrologia **22**, pp 237 - 244.

ISO (1993) *Guide to the Expression of Uncertainty in Measurement.* Geneva: International Standards Organistion.

Jaeger, K. B. & Davis, R. S. (1984), *A Primer for Mass Metrology*, NBS Special Publication 700-1.

Jaynes, E. T. (1957) *Information Theory and Statistical Mechanics* Phys. Rev. **106**, pp 620 - 630.

Jaynes, E. T. (1982) *On the Rationale of Maximum Entropy Methods* , Proc. IEEE **70**, No. 9, pp 939 - 952.

Jaynes, E. T. (1983) *Papers on Probability, Statistics and Statistical Physics.* R. D. Rosenkrantz (Ed.), D. Reidel, Dordrecht, Boston.

Jaynes, E. T. (1985) *Where do we go from here?* in *Maximum Entropy & Bayesian Methods in Inverse Problems,* C. R. Smith & W. T. Grandy, eds, D. Reidel, Dordrecht, Boston.

Jaynes, E. T. (1996) *Probability Theory: The Logic of Science.* Fragmentary Edition available at ftp://bayes.wustl.edu/pub/Jaynes/book.probability.theory/

Jones, F. E. (1978): *The Air Density Equation and the Transfer of the Mass Unit.* J. Res. NBS. **83**, pp 419 - 428.

Jones, R. H. & Tyron, P. V.(1983) *Estimating Time from Atomic Clocks,* J. Res. NBS **88,** No. 1, pp 17 - 24.

Kegel, T. M. (1996) *Statistical Control of a Pressure Calibration Process,* ISA Transactions **35**, pp 69 - 77.

Kochsiek, M. (1982) *Measurement of Water Adsorption Layers on Metal Surfaces*, Metrologia **18**, pp 153 - 159.

Kochsiek, M. (1984) *The Mass Unit 'Kilogram', Precision Measurement of Mass, Attainable Uncertainties and Possibilities of a new Definition.* in: *Precision Measurements and Fundamental Constants II.*, B. N. Taylor & W. D. Phillips eds., NBS Special Publication 617, pp 427 - 435.

Kajastie, H. & Riski, K. (1997) *Automatic Weight Handlers for Mass Comparators*, pp 67 - 74 in : *Weighing and Measurement in the Year 2000*, Int. Conf. 16 - 17 Sept. 1997, South Yorkshire Trading Standards Unit, Sheffield, England

Kruh,D. et al (1994) *Computerised Calibration System - Makam,* Measurement **13,** pp 225 - 233.

Kuroda, K. (1995), *Does the Time-of-Swing Method Give a Correct Value of the Newtonian Gravitational Constant?,* Phys. Rev. Lett. **75,** pp 2796 - 2798.

Laplace, P. S. (1951) *A philosophical Essay on Probabilities* (Translation from the original French) Dover Publications, New York.

Levine, R. D. & Tribius, M. (Eds), (1979) *The Maximum Entropy Formalism*, MIT Press.

Lewis, S. L. et al (1990) *An Intercomparison of Standards of Mass and Mass Measurement Techniques at 50g and 10g, between four European National Standards Laboratories.* Metrologia **27**, pp 233 - 244.

Lieu, R. et al (1987) *Maximum Entropy in Data Analysis with Error-Carrying Constraints.* J. Phys. A : Math & Gen. **20**, pp 2379 - 2388.

Lira, I. H. & W. Wöger (1997) *The Evaluation of Standard Uncertainty in the Presence of Limited Resolution of Indicating Devices.* Meas. Sci. Tech. **8**, pp 441 - 443.

Luenberger, D. G. (1968) *Optimisation by Vector Space Methods.* Wiley, New York.

Macurdy, L. B.(1964) *Response of Highly Precise Balances to Thermal Gradients,* J. Res. NBS **68C,** No. 3, pp 135 - 140.

Mandel, J. (1964) *The Method of Least Squares :* Chapter 7 in *The Statistical Analysis of Experimental Data.*Wiley.

Mari, L. (1996) *The Meaning of 'Quantity' in Measurement.* Measurement, **17** No. 2, pp 127 - 138.

Matthews, R. (1997), *Faith Hope and Statistics,* New Scientist, 22 Nov. pp 36 - 39.

Mihailov, G. & Romanowski, M. (1990) *Calibration of the Multiples of the Unit of Mass.* Metrologia **27**, pp 17 - 18.

Morris, E. C. (1992): *Decade Designs for Weightings of Non-Uniform Variance.* Metrologia **29,** pp 373 - 377.

Mueller, J. W. (1984) *The Assignment of Uncertainties to the Results of Experimental Measurements;* in: *Precision Measurements & Fundamental Constants II.* Taylor, B. N. & Phillips, W. D., Eds., NBS Special Publication 617.

Myklebust, T. (1995) *Methods to determine the magnetic properties of weights and magnetic field and field gradients of weighing instruments,* National Measurement Service, Norway, Feb. 1995.

Nielson, L. (1997) *Least Squares Estimation Using Lagrange Multipliers*, in: *Advanced Mathematical Tools in Metrology III*, World Scientific, Singapore.

NPL (1990) Calibration Certificate, Ref. No. 08C021/8919, Weight Set Serial No. 109M, 26/4/'90.

OIML International Recomendation IR 33. *Conventional Value of the Result of Weighing in Air.*

OIML (1994) Internation Recomendation IR 111. *Weights of Classes E_1, E_2, F_1, F_2, M_1, M_2 and M_3.*

Olson, P. T. (1991): *Monitoring the Mass Standard via the Comparison of Mechanical to Electrical Power,* IEEE Trans. Inst. Meas. **40**, No. 2, pp 115 - 120.

Orford, G. R. (1996) *Why so Many Uncertainty Documents?* IEE Half-Day Colloquium on Uncertainties Made Easy, London. pp 4/1 - 4/3.

Petley, B. W. (1992) *The Role of the Fundamental Constants of Physics in Metrology* Metrologia **29**, pp 95 - 112.

Pinot, P. (1994/95) *Stability of Mass Standards made of XSH Alacrite: Gravimetric Study of the Influence of Cleaning,* Metrologia **31**, pp 357 - 365.

Pinot, P. (1997) *Stability of XSH Alacrite Mass Standards: Study of the effect of the BIPM Cleaning / Washing Method*, Metrologia **34**, pp 133 - 141.

Probst, R. & Kochsiek, M. (1984) *Investigation of a Hydrostatic weighing method for a 1kg mass comparator,* Metrologia **19**, pp 137 - 146.

Prowse, D. B. & Anderson, A. R. (1974) *Calibration of a Set of Masses in Terms of One Mass Standard.* Metrologia **10**, pp 123 - 128.

Prowse, D. B. (1982) *Calibration of Masses* in: *Asia/Pacific Metrology Program, Report on the Regional Workshop on Metrology for Developing Countries; 30 August - 10 September 1982, Sydney, Australia; Volume 2: Proceedings, Part 2: Mass Standards & Measurements*, pp 33 - 54

PTB (1994) Calibration Certificate Ref. No. PTB 1.11-11643/94, Weight Set 37 PTB '94.

Quinn, T. J. (1986), *New Techniques in the Manufacture of Platinum-Iridium Mass Standards,* Platinum Metals Review **30**, No. 2, pp 74 - 79.

Quinn, T. J., Speake, C. C., & Davis, R. S. (1986/87), *A 1kg Mass Comparator using Flexure Strip Suspensions: Preliminary Results*, Metrologia **23**, pp 87 - 100.

Quinn, T. J. (1991) *The Kilogram: The Present State of our Knowledge*, IEEE Trans. Inst. Meas. **40**, No. 2, pp 81 - 85.

Quinn, T. J., Davis, R. S., Speake, C. C. & Brown, L. M. (1997a), *The Restoring Torque and Damping in Wide Cu-Be Torsion Strips,* Physics Letters A **228,** pp 36 - 42

Quinn, T. J., Speake, C. C., Davis, R. S., (1997b), *Novel Torsion Balance for the Measurement of the Newtonian Gravitational Constant*, Metrologia **34,** pp 245 - 249.

Rao, C. R. (1973) *Linear Statistical Inference and its Applications.* Wiley.

Ridler, N. M. (1996) *Evaluating the Uncertainty in Measurements—General Principles.* IEE Half-Day Colloquium on Uncertainties Made Easy, London. pp 1/1 - 1/4.

Robinson, I. A. & Kibble, B. P. (1996) *Measuring Planck's Constant & Monitoring the Kilogram at NPL.* Conference on Precision Electromagnetic Measurements Digest, Braunschweig, Germany, 17 -21 June 1996, p 5.

Robinson, I. A. & Kibble, B. P. (1997), *The NPL Moving-Coil Apparatus for Measuring Planck's Constant and Monitoring the Kilogram,* IEEE Trans. Inst. Meas. **46,** No. 2, pp 596 - 600.

Ross, S. M. (1972), *Introduction to Probability Models,* Academic Press.

Schoonover, R. M. et al (1980) *A Practical Test of the Air Denisty Equation in Standards Laboratories at Differing Altitude,* J. Res. NBS **85**, No. 1, pp 27 - 38.

Schwartz, R. (1991) *Realisation of the PTB's Mass Scale from 1 mg to 10 kg*. Report PTB-MA-21e, PTB, Braunschweig.

Schwartz, R. (1994a) *Precision Determination of Adsorption Layers on Stainless Steel Mass Standards by Mass Comparison and Ellipsometry. Part I: Adsorption Isotherms in Air,* Metrologia **31**, pp 117 - 128.

Schwartz, R. (1994b) *Precision Determination of Adsorption Layers on Stainless Steel Mass Standards by Mass Comparison and Ellipsometry. Part II: Sorption Phenomena in Vacuum,* Metrologia **31,** pp 129 - 136.

Schwartz, R & Glaeser, M (1994c): *Procedures for Cleaning Stainless Steel Weights, investigated by Mass Comparison and Ellipsometry,* Meas. Sci. Tech. **5,** 1994, pp 1429 - 1435.

Schwartz, R. (1995) *Guide to Mass Determination with High Accuracy,* PTB Report PTB-MA-40, PTB, Braunschweig.

Seah. M. P. et al (1994) *Stability of Reference Masses II: The Effect of Environment and Cleaning Methods on the Surfaces of Stainless Steel and allied Materials.* Metrologia **31**, pp 93 - 108.

Shannon, C. E. (1948) *A Mathematical Theory of Communication,* Bell Systems Tech. Journal **27,** p 379.

Sivia, D. S. (1996) *Data Analysis: A Bayesian Tutorial* , Clarendon Press, Oxford.

Sobolev, V. I. (1996) *Conceptual Uncertainty of Measurement Errors*, Meas. Tech. **39**, No. 5 pp 466 - 469.

South Yorkshire Trading Standards Unit (1995) Calibration Certificate, Serial No. T15790, 19/6/1995, Weight Set ST1/810218

Spieweck, F. & Bettin, H. (1992), *Review: Solid & Liquid Density Determination,* Technisches Messen **59**, pp 237 - 244 & 285 - 292.

Steiner, R. (1996), *The NIST Watt Balance: Progress toward monitoring the Kilogram*, in Conference on Precision Electromagnetic Measurements Digest, Braunschweig, Germany, 17 -21 June 1996, pp 6 - 7.

Steiner, R. *et al* (1997), *The NIST Watt Balance: Progress towards monitoring the Kilogram,* IEEE Trans. Inst. Meas. **46,** No. 2, pp 601 - 604.

Stenbakken, G. N. , *et al* (1995): *Methods for Aligning the NIST Watt-Balance,* IEEE Inst. / Meas. Tech. Conf. Procs. , 24 - 26 April 1995, pp 247 - 251.

Sutton, C. M. & Clarkson, M. T. (1993/94) *A General Approach to Comparisons in the Presence of Drift,* Metrologia **30,** pp 487 - 493.

Taylor, B. N. (1991) *The Possible Role of the Fundamental Constants in Replacing the Kilogram* IEEE Trans. Inst. Meas. **40**, No. 2, pp 86 - 91.

Trebbia. P (1996) *Maxwell's Demon and Data Analysis*, Phil. Trans. Royal Soc. London, **452**, pp 2697 - 2711.

Tyron, P. V.& Jones, R. H.(1983) *Estimation of Parameters in Models for Cesium Beam Atomic Clocks.* J. Res. NBS. **88,** No. 1, pp 3 - 16.

Van den Bos, A. (1994) *Parametric Statistical Model-based measurement*, Measurement **12**, pp 237 - 249.

Weise, K. (1985) Preliminary English Translation of DIN 1319, Teil 4 (1985), *Grundbegriffe der Meβtechnik; Behandlung von Unsicherheiten bei der Auswertung von Messungen*, available from K. Weise, PTB, Braunschweig.

Weise, K. (1987) *Treatment of Uncertainties in Precision Measurements*. IEEE Trans. Inst. Meas. **IM-36**, No. 2 pp 642 - 645

Weise, K. & Woeger, W. (1992) *A Bayesian Theory of Measurement Uncertainty*, Meas. Sci. Tech. **3**, pp 1 - 11.

Weise, K. & Woeger, W. (1994) *Comparison of two Measurement Results using the Bayesian Theory of Measurement Uncertainty*. Meas. Sci. Tech. **5**, pp 879 - 882.

Wexler, A. & Wildhack, W. A. (eds.) *(1965) Proceedings of International Symposium on Humidity & Moisture, 1963, Washington; Volume 3: Fundamentals & Standards*. Reinhold, New York.

Wexler, A. (1976), *Vapour Pressure Formulation for Water in the Range 0 to 100°C: A Revision*, J. Res. NBS. **80A**, Nos. 5 & 6 pp 775 - 785.

Woeger, W. (1987) *Probability Assignment to Systematic Uncertainties by the Principle of Maximum Entropy.* IEEE Trans. Inst. Meas. , **36** pp 655 - 658.

Yoneda, K. (1996) *Optimal Number of Digits to Report* J. Operations Res. Soc. Japan, **39** No. 3, pp 428 - 434.

Zelen, M. (1962) *Linear Estimation and Related Topics* in *A survey of Numerical Analysis*, John Todd ed. McGraw Hill.

Zuker, M. et al (1980) *Systematic Search for Orthogonal Systems in Calibrations of Submultiples and Multiples of the Unit of Mass.* Metrologia **16**, pp 51 - 54.

****

Appendix 5. Publications

In : *Advanced Mathematical Tools in Metrology II.* pp 259 - 266, P. Ciarlini, M.G. Cox, F.Pavese & D. Richter, editors. World Scientific Publishing Company, 1996.

An Analysis of Statistical Estimation Methods used in High Accuracy Mass Determination

Thomas S. Leahy
School of Physical Sciences
Dublin City University
Dublin 9, Ireland
email :- tsl@physics.dcu.ie

Abstract

The application of Bayesian-based statistics to the solution of a set of over-determined equations resulting from comparison calibrations is considered. Following the recommendations of the ISO Guide on Measurement Uncertainty, a model parameterisation is developed which facilitates the inclusion of all known experimental information, and indeed prior information from previous calibrations, should this be available. A critical comparison, between this recursive approach and the classical solutions based on Lagrange Multipliers or the Gauss-Markov theorem, is made.

Some non-trivial differences between this approach and the conventionally used approaches in mass calibration have been found. An example from relevant experimental data, treated by both estimation methods, is included.

1. Introduction

The problem at hand is the data reduction of a set of n over-determined equations, to estimate p parameter values. This requires an adequate description of the experimental model and also a suitable Parameter Estimation method. The application of such a process to high accuracy mass determination is considered here. An implementation of the ISO Guide[1] for parameterising the experimental situation has been proposed [2] recently. This model is developed for a specific system and an example is given of its implementation. The feature of particular importance is the inclusion of uncertainty terms due to the systematic buoyancy correction in the overall model, allowing their inclusion in the estimation process.

With respect to the estimation process, two main approaches are highlighted: one based on the well known Least Squares (LS) method; and a Bayesian approach based on Maximum Likelihood Estimation (MLE). The former has been conventionally used in mass metrology (e.g. [3]) while the latter has been proposed [4, 5] as a means to circumvent some difficulties with the former. These arise as a result of the need to incorporate restraints in the estimation process in order to get a particular solution, since the calibration designs involve only differences between parameters which prevent a unique solution being obtained. The MLE approach is implemented here and its behaviour studied by looking at its performance in a specific example. It is shown

to be much more flexible than the LS approach and its treatment of the necessary constraints is highlighted as being preferable.

2. Model Parameterisation

At the outset of any experiment it is necessary to clearly describe the relationship between the parameters to be estimated and the data to be obtained for this purpose. The system of equations so generated may be referred to as a System Model, such as :

$$\mathbf{Y} = \mathbf{f}(\beta) \tag{1}$$

In the case of mass comparisons, the data can be described by

$$\mathbf{Y} = \Delta\mathbf{W}_{n\,x\,1} + \rho_{n\,x\,n}.\mathbf{X}_{n\,x\,p}.\mathbf{V}_{p\,x\,1} \tag{2}$$

while $\mathbf{f}(\beta) = \mathbf{X}.\beta$. Here, $\Delta\mathbf{W}$ is the vector of observed "weight-in-air" differences; ρ is a diagonal matrix whose elements are the relevant air densities for each comparison, $\mathbf{V}$ is a vector of volumes of the p parameters and $\mathbf{X}$ is a design scheme for the comparison experiment. Automated commercially available comparators are used to obtain $\Delta\mathbf{W}$. Sensitivity weights are not employed and other corrections such as those for center of gravity differences or volume expansion coefficients are not considered.

It is also necessary to establish a covariance matrix for $\mathbf{Y}$. This is done using the error propagation analysis of [1] with the matrix implementation described in [2]. If $\mathbf{Y} = \mathbf{f}(\mathbf{U})$, where $\mathbf{U}$ is the vector of influence quantities involved in generating $\mathbf{Y}$, the covariance matrix of $\mathbf{Y}$ can be calculated from:

$$\psi_Y = \mathbf{J}_u.\psi_u.\mathbf{J}_u^T \tag{3}$$

and

$$\mathbf{J}_u = \nabla_u.\mathbf{Y}^T \tag{4}$$

$$\mathbf{U} = \begin{vmatrix} \Delta\mathbf{W} \\ \rho \\ \mathbf{V} \end{vmatrix}, \text{ and } \psi_u = \begin{vmatrix} \psi_{\Delta W} & 0 & 0 \\ 0 & \psi_\rho & 0 \\ 0 & 0 & \psi_v \end{vmatrix}$$

where $\psi_\mathbf{u}$ is the variance / covariance matrix of $\mathbf{U}$. It is assumed that there are no covariances among the influence quantities as each results from separate measurements using separate instrumentation. Thus with $\mathbf{U}$ and $\psi_\mathbf{u}$ as opposite and using Eq.(2), Eq.(3) can be evaluated to give:

$$\psi_Y = \psi_{\Delta W} + diag\{\mathbf{x}.\mathbf{v}\}.\psi_\rho.diag\{\mathbf{x}.\mathbf{v}\} + \rho.\mathbf{x}.\psi_V.\mathbf{x}^T.\rho \tag{5}$$

$\psi_{\Delta W}$ is diagonal as each comparison is a separate measurement. $\psi_{\Delta W} = s^2.\mathbf{I}$ is not assumed since this is very rarely, if ever, true in practice. The second term is the contribution to $\psi_\mathbf{y}$ due to the air density term in Eq.(2). This is diagonal since $\rho = diag\{\rho_1,\rho_2,...\rho_n\}$ and not $\rho.\mathbf{I}$. Thus each ρ_i is an independent measurement. The third term gives the contribution due to the volume term in Eq.(2). While $\psi_\mathbf{v}$ is diagonal, the third term as a whole, is non-diagonal as a result of the form of $\mathbf{X}$; and introduces covariances into $\psi_\mathbf{y}$. Eqs.(2) & (5) provide all the necessary tools to give a full parameterisation of the calibration experiment.

3 Selecting an Estimation Method

The most commonly used estimator which can be derived from either a Weighted Least Squares (WLS) approach, or more generally via the Gauss-Markov theorem (GM), gives solutions of the form

$$\beta_{GM} = (\mathbf{X}^T.\psi_Y^{-1}.\mathbf{X})^{-1}.\mathbf{X}^T.\psi_Y^{-1}.\mathbf{Y} \tag{6}$$

$$\psi_{\beta_{GM}} = (\mathbf{X}^T.\psi_Y^{-1}.\mathbf{X})^{-1} \quad (7)$$

(see for example [6, 7, 8])

However, ($\mathbf{X}^T.\mathbf{X}$) is singular which is an inevitable consequence of a calibration design involving only differences in parameters so Eqs.(6) & (7) cannot be used directly. The conventional solution to the problem is to impose constraints on the Normal Equations; the constraint usually being the value of a selected parameter involved in the comparison. In general the constraint can be expressed as a linear combination of the parameters so that:

$$\mathbf{A}^T.\beta = \mathbf{C} \quad (8)$$

where **C** is the constraint vector. The implementation of this method has been outlined in [4] and results in estimators whose form can be summarised as:

$$\beta_{LM} = \mathbf{L}.\mathbf{Y} + \mathbf{M}.\mathbf{C} \quad (9)$$

$$\psi_\beta = \mathbf{L}.\psi_y.\mathbf{L}^T + \mathbf{M}.\psi_c\mathbf{M}^T \quad (10)$$

where **L** and **M** are linear, non-random matrices.The problems with this estimator have been well documented [9], chief among them being its inadequate treatment of the constraint information which is considered deterministically to get a solution and then stochastically to obtain the proper covariance matrix

It is better if all information necessary to obtain a complete solution can be included in the estimation process from the beginning. In this respect Maximum a Posteriori estimation, based on Bayesian statistics is superior. Here the constraints are viewed as prior information which is to be updated by the current information obtained in the experiment. The estimators in this case are:

$$\beta_{MAP} = \beta_0 + (\mathbf{X}^T.\psi_y^{-1}.\mathbf{X} + \psi_{\beta_0}^{-1})^{-1}.\mathbf{X}^T.\psi_y^{-1}.(\mathbf{Y} - \mathbf{X}.\beta_0) \quad (11)$$

$$\psi_{\beta_{MAP}} = (\mathbf{X}^T.\psi_y^{-1}.\mathbf{X} + \psi_{\beta_0}^{-1})^{-1} \quad (12)$$

(See for example [6] and also [10, 11] for some general comments on Bayesian Estimation and the treatment of measurement uncertainties)

With this estimator, there is no problem with ($\mathbf{X}^T.\mathbf{X}$) being singular since the only condition is the existence of Eq.(12).

4. Example

4.1 Input Data

The comparison experiment involves three 50g standards, labeled b_1,b_2,b_3. The design matrix is given in Fig.1. Table I gives the relevant experimental data while Table II gives the prior information. The β_0 data is given as deviations in μg from a nominal 50g.With the aid of Eq.(5) ψ_y is calculated (see Fig. 3) The three constituent parts are shown, illustrating that the volume term gives rise to the largest variance/covariance contribution.

$$X = \begin{vmatrix} 1 & -1 & 0 \\ 1 & 0 & -1 \\ 0 & 1 & -1 \\ -1 & 1 & 0 \\ -1 & 0 & 1 \\ 0 & -1 & 1 \end{vmatrix} \quad \& \quad \beta = \begin{vmatrix} b_1 \\ b_2 \\ b_3 \end{vmatrix}$$

Fig. 1 Design Matrix and Parameter Vector

4.2 Bayesian Estimation

The estimators in Eqs. (11) & (12) are used, giving the estimated values and covariance matrix as illustrated in Fig. 2. It is immediately obvious from a comparison with Table II that a reduction in the estimated variances of the parameters has been effected. This is a consequence of freeing the parameters from a deterministic situation and then applying a minimum variance estimator. It is clear from Fig 2 that a complete solution has been established and no further calculation is necessary.

Table I - Experimental Information

Data Point	ΔW (μg)	diag ($\psi_{\Delta W}$) $(\mu g)^2$	diag(ρ) ($mg.cm^{-3}$)	X.V (cm^3)	Y (μg)
y_1	66.0	0.107	1.199	-0.1419	-104.1
y_2	-109.0	0.227	1.202	-0.1266	-261.2
y_3	-173.0	0.190	1.202	0.0153	-154.6
y_4	- 65.6	0.190	1.200	0.1419	104.7
y_5	109.5	0.112	1.191	0.1266	260.3
y_6	172.4	0.140	1.207	-0.0153	153.9

Table II - Prior Information

Parameter	β_0 (μg)	diag(ψ_{β_0}) $(\mu g)^2$	V (cm)^3	diag(ψ_V)$^{1/2}$ (cm^3)
b_1	- 63.0	25.0	6.2202	0.0011
b_2	34.0	225.0	6.3621	0.0009
b_3	186.0	225.0	6.3468	0.0009

$$\begin{vmatrix} 0.107 & 0 & 0 & 0 & 0 & 0 \\ & 0.227 & 0 & 0 & 0 & 0 \\ & & 0.190 & 0 & 0 & 0 \\ & \text{symm} & & 0.190 & 0 & 0 \\ & & & & 0.112 & 0 \\ & & & & & 0.140 \end{vmatrix}$$

(a)

$$\begin{vmatrix} 3.02 & 0 & 0 & 0 & 0 & 0 \\ & 2.40 & 0 & 0 & 0 & 0 \\ & & 0.035 & 0 & 0 & 0 \\ & \text{symm} & & 3.02 & 0 & 0 \\ & & & & 2.40 & 0 \\ & & & & & 0.035 \end{vmatrix} * 10^{-3}$$

(b)

$$\begin{vmatrix} 2.9 & 1.74 & -1.16 & -2.9 & -1.72 & 1.17 \\ & 2.9 & 1.17 & -1.17 & -2.89 & -1.17 \\ & & 2.34 & 1.17 & -1.16 & -2.35 \\ & \text{symm} & & 2.90 & 1.73 & -1.17 \\ & & & & 2.86 & 1.16 \\ & & & & & 2.36 \end{vmatrix}$$

(c)

$$\begin{vmatrix} 3.01 & 1.74 & -1.67 & -2.90 & -1.72 & 1.17 \\ & 3.15 & 1.17 & -1.74 & -2.89 & -1.17 \\ & & 2.53 & 1.16 & -1.15 & -2.35 \\ & \text{symm} & & 3.10 & 1.72 & -1.17 \\ & & & & 2.98 & 1.16 \\ & & & & & 2.50 \end{vmatrix}$$

(d)

Fig 3. The components of ψ_y : (a) $\psi_{\Delta W}$ term ; (b) term due to air density variance ; (c) term due to volume variance ; & (d) the complete ψ_y

$$\beta = \begin{vmatrix} -64.7 \\ 40.0 \\ 195.1 \end{vmatrix} (\mu g) \; ;$$

$$\psi_\beta = \begin{vmatrix} 20.53 & 20.11 & 20.11 \\ 20.11 & 22.60 & 21.40 \\ 20.11 & 21.40 & 22.59 \end{vmatrix} (\mu g)^2$$

Fig. 2 Estimated Parameter Values and Covariance Matrix Using Eqs. (11) & (12)

4.2.1 Analysis

Table III gives the estimated parameter values and variances for several cases of prior information, ranging from the minimum necessary, to full prior information about all parameters. When just one parameter is included in the prior information, only one solution is possible in each case and so no reduction in variances can be effected. This is illustrated in Cases B,C & D in Table III, which show that under these conditions it is preferable to use the parameter which has lowest variance to get the best solution. It is also clear that the parameter of lowest variance exerts the greatest influence on the result. For example, in Case G, the variance of b_2 and b_3 is reduced to 113 μg^2. However, in Cases E & F ,as well as A & B, where the much lower variance of b_1 is included in the prior information, the estimated variance for b_2 and b_3 is now much reduced.

The relative accuracy of the prior information influences the estimated *values* as well as the estimated *variances*. When b_1 is not included in the prior information (i.e. it is given infinite prior variance and arbitrary value) it is adjusted much more significantly. In the limit of zero variance, on the other hand, the parameter would not be adjusted in any way. This is a useful feature for dealing with suspect prior information which may not be accurate. In Table IV parameter b_3 is given an in-error prior value which should adversely effect the estimated values. In the first column (case (i)) the estimated values can be seen to be quite different to the data in Table III; while in the second column, where $s^2(b_3) = 900\ \mu g^2$, the in-error prior value is adjusted significantly while the other two are adjusted much less. In this respect, the relative accuracy of the prior information and the experimental information is important: prior information which disagrees with the experimental evidence will result in large residuals (where residuals here are the differences between observation data and estimated observation data, i.e. $(\mathbf{Y} - \mathbf{X}.\beta)$. However, if the experimental information is significantly more accurate than the prior information, the influence of the latter will be reduced. To simulate this, the prior information was given a variance of 4 μg^2 ,which is similar in magnitude to that of the experimental information as given in Fig. 3. This is shown in Table IV, case (iii). Table V gives the residuals for the various cases of Table IV : the difference in residuals between cases (i) & (ii) is minimal, while the difference between cases (iii) & (iv) is significant.

Table III Estimated Values and Variances for Various Prior Information

Estimated Parameter Values (μg)							
Prior Data.	Case A	Case B	Case C	Case D	Case E	Case F	Case G
b_1 : -63	* -64.7	* -63.0	-71.5	-74.3	* -63.8	* -64.1	-72.9
b_2 : 34	* 40.6	42.5	* 34.0	31.1	* 41.5	41.3	* 32.6
b_3 : 186	* 195.4	197.4	188.8	* 186.0	196.5	* 196.1	* 187.4
Estimated Parameter Variance (μg)^2							
Prior Variance	Case A	Case B	Case C	Case D	Case E	Case F	Case G
$s^2(b_1)$: 25	* 20.53	* 25.0	227.99	227.99	* 22.53	* 22.53	114.8
$s^2(b_2)$: 225	* 22.63	27.99	* 225.0	227.42	* 24.89	25.16	* 113.1
$s^2(b_3)$: 225	* 22.60	26.76	227.42	* 225.0	25.12	* 24.86	* 113.1

Note :- " * " => corresponding prior information used in the estimation process

Table IV - Simulating an Error on one Parameter (using all prior information)

Variances (μg)^2	case (i)	case (ii)	case (iii)	case (iv)
$s^2(b_1)$	25	25	4	4
$s^2(b_2)$	25	25	4	4
$s^2(b_3)$	25	900	4	900
Prior Values (μg)		Estimated Values		
b_1 -63.0	-77.4	-67.4	-74.0	-66.1
b_2 34.0	27.4	37.6	28.4	37.0
b_3 160.0	180.9	192.5	176.6	192.7

Table V Comparison of Residuals (Y - X.β) for Table IV data

Table IV - case (i) residuals (μg)	Table IV case (ii)	Table IV case (iii)	Table IV case (iv)
0.82	1.02	-1.61	-0.81
-2.84	-1.18	-10.61	-2.27
-1.16	0.29	-6.50	1.04
-0.12	-0.32	2.31	1.51
1.94	0.28	9.71	1.37
0.36	-1.09	5.70	-1.84

4.2.2 Principle Conclusions

Complete evaluation of ψ_β with reduced error covariance matrix for all parameters is possible. Degrees of belief associated with the prior information plays an important role: if some information is considered more accurate than others, the estimates will be constrained more in that direction.

The estimator can be very robust in dealing with errors in either the prior or current information, but this depends on the relative accuracies of the data. If a higher degree

of belief is attached to some information which is in error, it does not seem possible to counteract this, and the estimator is then less robust. The problem will be highlighted however, in either large residuals or estimated values which are significantly different from their prior values, indicating that at least one set of information is in error.

4.3 Restrained Least Squares Estimation

The conventional least squares solution is applied to the case of just one constraint. This gives the same result as the Bayesian estimator with one piece of prior information, as there is only one solution in this situation and all estimators will generate it. The solution is shown in Fig. 4. The incomplete covariance matrix is shown along with the "constraint contribution", ψ_c, which must be added to give a complete solution. This estimator performs very badly if there are two constraints in this case, since all the adjustment must be carried out on only one parameter. This results in very poor agreement with the experimental data, large residuals and an unsatisfactory fit. As there is no facility with this method to enter any information about the relative accuracy of the prior information, the inherent variability of the constraints cannot be used to advantage in the estimation process.

$$\beta = \begin{vmatrix} -63 \\ 42.5 \\ 197.4 \end{vmatrix} (\mu g) \quad \psi_\beta = \begin{vmatrix} 0 & 0 & 0 \\ 0 & 2.99 & 1.76 \\ 0 & 1.76 & 2.95 \end{vmatrix} (\mu g)^2; \quad \psi_C = \begin{vmatrix} 25 & 25 & 25 \\ 25 & 25 & 25 \\ 25 & 25 & 25 \end{vmatrix} (\mu g)^2$$

Fig 4 - Estimated parameter values and covariance matrix using Restrianed Least Squares Solution

6 Conclusion

The requirements for establishing a complete parameter estimation method for dealing with over-determined, singular design comparison calibrations have been outlined. This includes the necessity to fully parameterise all variance/covariance components associated with the input data. Off-diagonal terms in the covariance matrix of the input data should always be included as there are cases where their inclusion does effect the result. It is also necessary to find a method for suitably incorporating any constraints required to obtain a solution. It has been shown that Bayesian estimators are best for this as they take a more appropriate view of the prior information than does the traditionally used estimator. Also, some of the characteristics of the Bayesian Estimator have been examined, noting how it performs with various combinations and accuracies of prior/current data. It has been shown to be very robust under most conditions and to clearly highlight problems even where it cannot circumvent them. In contrast, the conventionally used solution is much more rigid and cannot react to varying information about the prior data. The overall approach described allows a uniform treatment of the data which is in line with current guidelines on uncertainty estimation.

Acknowledgments

The author is indebted to the Irish National Metrology Laboratory for providing laboratory access and equipment related to the experimental aspects of the research

and to Dr. B. Lawless of Dublin City University for many fruitful discussions and ideas.

References

1. I.S.O. 1993 *Guide to the Expression of Uncertainty in Measurement.*
2. W. Bich et.al., Metrologia 1993/94 ,**30**, 495 - 502, *Uncertainty Modeling in Mass Comparisons*
3. R. Schwartz, PTB-Bericht ma-21-e, Phys.-Tech. Bundesantalt, May 1991 *Realisation of the PTB Mass Scale from 1mg to 10kg.*
4. W. Bich, Metrologia 1992, **29**, 15 - 22, *Bias and Optimal Linear Estimation in Comparison Calibrations*
5. W. Bich, Proc. 13th IMEKO Int. Conf. on Force & Mass Measurement, Helsinki, May 1993 , 173 - 183, *Bayesian and Fisherian Estimation in the recalibration of a set of Mass Standards : a Comparison of Methods.*
6. J. V. Beck & K. J. Arnold, Wiley, 1977, *Parameter Estimation in Science and Engineering.*
7. D.G. Luenberger, Wiley, 1969, *Optimization by Vector Space Methods.*
8. C.R. Rao, Wiley, 1973 *Linear Statistical Estimation and its Applications.*
9. W. Bich, Metrologia 1990, **27**, 111 - 116, *Variances, Covariances and Restraints in Mass Metrology.*
10. K. Weise & W. Woger, Meas. Sci. Tech. 1992, **3**, 1 -11, *A Bayesian Theory of Measurement Uncertainty.*
11. K. Weise, IEEE Trans. Inst. Meas., June 1987, **IM-36 #2**, 642 - 645, *Treatment of Uncertainties in Precision Measurement.*

In : *Advanced Mathematical Tools in Metrology III.* pp 266 - 267, P. Ciarlini, M.G. Cox, F.Pavese & D. Richter, editors. World Scientific Publishing Company, 1997

MAXIMUM ENTROPY & BAYESIAN APPROACHES TO PARAMETER ESTIMATION IN MASS METROLOGY

THOMAS S. LEAHY
School of Physical Sciences
Dublin City University
Dublin 9, Ireland
Email : tsl@physics.dcu.ie

EXTENDED ABSTRACT

Data analysis models should be both mathematically sound and physically relevant. Consistency and homogeneity should also be characteristics of the analysis. The treatment of measurement uncertainties is critical to the accuracy of the method developed. The logic of Classical Probability theory provides the most appropriate approach to dealing with data of a non-deterministic nature as it is a method of reasoning in the absence of certainty[1,2]. A Bayesian view of probability is adopted, based on desiderata of consistency—namely that the same conclusion should be reached, irrespective of the intermediate routes of evaluation, that all known and available information must be incorporated into the analysis and that equivalent states of knowledge must be described by equivalent mathematical statements[1,3]. From this foundation a complete probability analysis can be constructed.

A thorough practical basis for implementing a measurement analysis strategy has been presented in the ISO Guide[4], now widely accepted. One of its key aspects is a uniform and consistent treatment of all influence quantities. Establishing variance measures for all elements is achieved by assigning Degrees of Belief which reflect the extent of available knowledge about each parameter. This will include all relevant information, but only that information which is available about the parameter[1,2,5].

The Principle of Maximum Entropy (MaxEnt) ensures that Degrees of Belief can be assigned in a consistent manner, free from possible bias. The entropy of a probability distribution gives a measure of the amount of "uncertainty" it represents. Maximising this subject to the constraints of whatever *is* known about the parameter generates the most honest probability assignment since it assumes the least knowledge about the parameter while reproducing any known features. For example, maximising the entropy tells us, as we would expect, that a uniform distribution is the best assignment when nothing is known except the range of values the parameter could adopt. Also when an estimated mean and variance are known, MaxEnt indicates a Gaussian Distribution to be the least subjective.

The analysis of mass calibration data involves the parameterising of an experimental situation and subsequent data reduction of a set of over-determined comparisons to evaluate the mass values of a set of standards[6,8]. A Unified Approach is desired to all aspects of this work, which will ensure optimal estimation and full use of all available information in a non-biased manner[7]. This is done using the principles of consistent reasoning of the ISO Guide, MaxEnt and Bayesian Parameter Estimation.

Statistical Adjustment by the Bayesian Method[6,7,8] ensures the process of consistent reasoning is maintained. It utilises all the available information while not assuming unknown data. It results in parameter estimates of improved accuracy, often providing error-detection and even error-correction, should there be inconsistencies or systematic errors in the data.

References

1. Jaynes, E. T. "*Probability Theory: The Logic of Science*"; Fragmentary Edition, March, 1996, available on the WWW.
2. Garrett, A. J. M. "*Macroirreversibility & microreversibility reconciled: the second law*" pp 139 - 170 in *Maximum Entropy in Action*, B. Buck & V. A. Macaulay, eds. Oxford, 1991.
3. Sivia, D. S. "*Data Analysis: A Bayesian Tutorial*", Oxford, 1996
4 I.S.O. "*Guide to the Expression of Uncertainty in Measurement*", 1993
5 Weise, K. & Wöger, W. "*A Bayesian theory of Measurement Uncertainty*"; Meas. Sci. Tech. **3**, 1992, pp 1 - 11.
6 Leahy, T. S. "*An analysis of Statistical Estimation Methods used in High Accuracy Mass Determination*", pp 259 - 266 in *Advanced Mathematical Tools in Metrology*, P. Ciarlini *et al*, eds. World Scientific, Singapore, 1996.
7 Leahy, T. S. "*An Analysis of Statistical Estimation Methods & Model Parameterisation in Mass Calibration*", Postgraduate Research Thesis, in preparation.
8 Bich, W. "*Bias & Optimal Linear Estimation in Comparison Calibrations*", Metrologia, **29**, 1992, pp 15 - 12.

In: *Weighing and Measurement in the Year 2000*, pp 115 - 126, Proceedings of the South Yorkshire Trading Standards Unit International Conference, Sept. 16 -17, 1997.

SOFTWARE APPLICATIONS IN MASS METROLOGY

Thomas S. Leahy

School of Physical Sciences
Dublin City University
Dublin 9, Ireland

National Metrology Laboratory
Forbairt
Dublin 9, Ireland

Email : tsl@physics.dcu.ie

ABSTRACT

Computer-controlled data acquisition and processing systems are becoming more widespread in the mass calibration laboratory in recent years. Comparator manufacturers often provide software accompanying their instruments, especially in respect of the automated comparators now widely used in high accuracy work.

This paper presents the motivation for a laboratory to develop its own software rather than relying on externally produced packages. The scope of software applications is considered in the light of metrological requirements. Problems resulting from the use of different software packages with different instruments are discussed in terms of data storage and presentation formats.

Access to source code is considered essential for the traceability and accuracy that is required in a metrological context. It must be possible to verify the algorithms used, particularly in relation to uncertainty calculation. The need for a coherent, unified approach in this regard is presented.

The advantages of in-house software development are highlighted by reference to software developed recently at the National Metrology Laboratory, Dublin. This software is able to access a range of different mass comparators and is also able to interface with additional instruments for measuring temperature, pressure and humidity within the laboratory. Some modifications to the software to deal with special requirements are also discussed.

Experimental data obtained and analysed with this software is presented showing systematic influences on the automated mass comparators which would not have been so easily highlighted without the use of this data acquisition software.

1. INTRODUCTION

The instrumentation employed in the mass calibration laboratory is becoming more and more sophisticated in recent years. This concerns both technical properties of the measurement instruments and also, of particular interest in this paper, software capabilities. Often, many functions are provided on software EPROM's in the instrument, allowing various menus to be accessed from the front panel controls and nearly always an interface port is provided, allowing these functions to be controlled remotely via a computer.

This situation opens up a whole range of opportunities for automating the measurement and analysis process, in particular if the measurement instruments are able to perform automated comparison calibrations. Manufacturers, having designed their instruments with these facilities have naturally utilised the possibility of computer control by providing software applications for data acquisition, manipulation and storage. These labour-saving devices are of great interest but must also be the subject of great scrutiny owing to the particular requirements of a calibration laboratory with respect to accuracy, traceability and documentation.

This paper considers some software applications in the mass laboratory, primarily concerned with data acquisition and processing for use in routine calibration work. Software developed at the National Metrology Laboratory in Dublin will be used as an example and some experimental observations on automated mass comparators obtained with the aid of this software will be discussed.

2. SOFTWARE APPLICATIONS AND LABORATORY REQUIREMENTS

The typical commercial calibration laboratory is involved in the calibration of standards from the lower OIML 'M' classes up to 'F1' and perhaps even some 'E2' standards. In this context usually a large number of calibrations are carried out, covering the spectrum from *mg* weights through to *20kg* and perhaps up to *50kg*.

Thus there will certainly be a range of mass comparators in the laboratory and it is more than likely that they will not all be from the same manufacturer. The instruments in use may include conventional manually-operated comparators and also perhaps some of the more sophisticated automated instruments equipped with weight exchange mechanisms. To focus on our specific interest in this paper, most modern instruments are equipped with serial interface ports as standard, and will have a proprietary set of commands allowing bi-directional communication with a peripheral device such as a computer. Comparators produced by a given manufacturer will probably have a similar instruction set which may be hierarchical—more sophisticated machines will utilize all of the commands of their simpler relatives' along with additional controls.

However, it is very unlikely that instruments from different manufacturers will incorporate similar instruction sets. And it is here that the problems begin to appear as each comparator may have its own supplied software package, resulting in the laboratory having possibly several different pieces of software, each performing essentially the same task. Apart from the additional workload for the operator in becoming familiar with several different packages, there are more fundamental problems: the data storage formats may be different, for example some may write data in plain ASCII text while others may use binary formats, only readable with the

proprietary software, leading to difficulties if the user wishes to further analyze the data, perhaps with other software. The data analysis may differ between packages for example, some may provide a Reference weight database, allowing the final value of the test weight to be determined and displayed, while others may just provide summary information about the comparison such as the mean value and standard deviation. Not all packages attempt to perform an uncertainty analysis and if such an analysis is implemented, there may be further problems in regard to traceability and documentation. The presentation and display of the data will be different for different packages, again leading to unhelpful differences within the laboratory.

For higher accuracy work, to 'E2' & 'E1' level for example, it is most likely that buoyancy corrections to the comparison data would be needed. This means that more instrumentation must be incorporated with the software. Typically a laboratory will have several different types of instrument for measuring air temperature, barometric pressure, relative humidity or dew point temperature and perhaps carbon dioxide content as well. Some mass calibration software provides the facility for the user to manually enter data from these instruments during the comparison, which is not an ideal solution since operator transcription errors are not impossible and in any event it defeats the purpose of automated data acquisition software, especially if automated comparators are in use. A particularly useful feature of these instruments is that they can be operated when the laboratory is empty and optimum conditions are realizable. Some manufacturers have produced climate-monitoring systems which can be connected to the PC along with the comparator. However, the specifications of these instruments are not always adequate for high accuracy requirements. The best approach is for the laboratory to provide itself with the necessary equipment first and then to consider a software implementation afterwards which will meet its own requirements.

3 GENERAL REQUIREMENTS FOR TRACEABILITY

So far the scope of software applications in mass metrology has been considered. Apart from matters of utility and convenience, there are also important issues concerning traceability and documentation of software and these provide the strongest motivation for a laboratory to take a keen interest in the software it uses, [1]. Modern metrology and calibration is founded upon a carefully controlled system of documentation and traceability so that every measurement is connected to an approved and agreed-upon standard. In the case of mass metrology there is a hierarchical chain of standards back to the International Prototype Kilogram held at BIPM and ancillary measurements needed in calibration, such as temperature data for example, are also traceable via national and international standards. However, if we insert a software

processing unit at some point in these proceedings, there is a great danger of it becoming a "black box" whose form and functions are unknown. Fig. 1 highlights the important position of the software in this respect. Much has been said and written about software testing and verification techniques which do help to deal with this difficulty, but our primary interest here is in knowing exactly what the software is doing with our data, rather than simply verifying it with standard test data for example, [2].

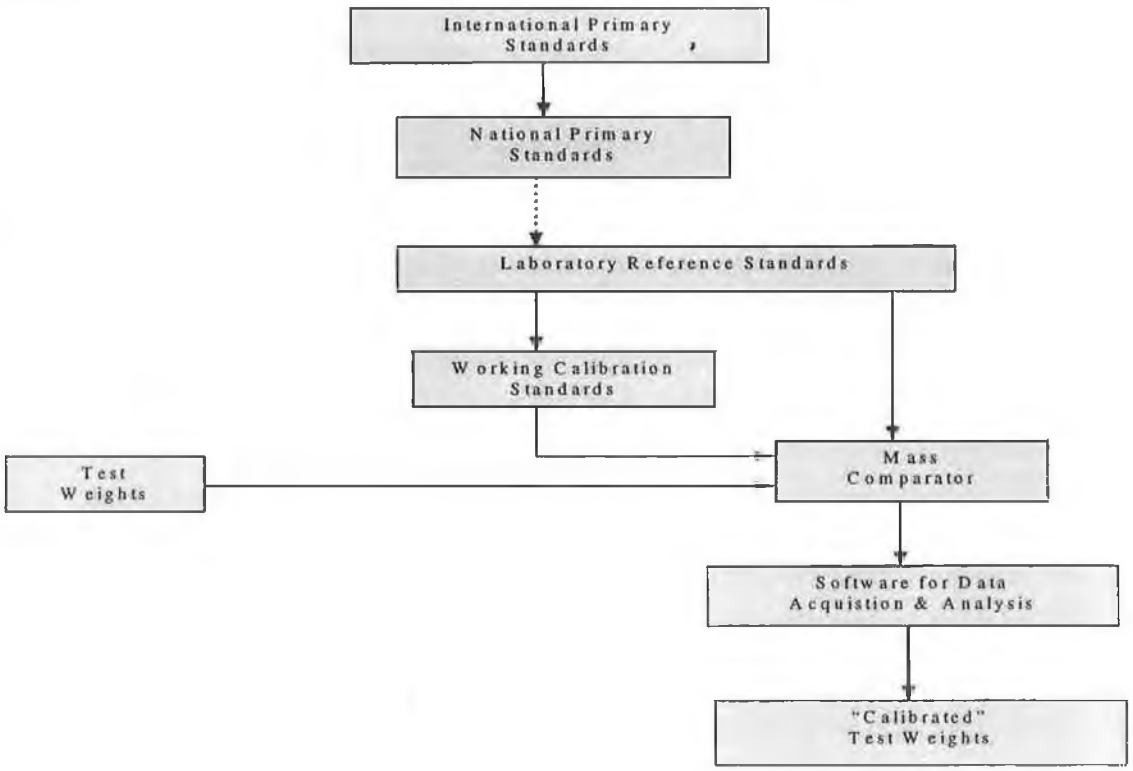


Fig. 1: The Position of Laboratory Data Acquisition & Analysis Software in the Calibration Hierarchy. Unless the software is well understood it can be an unknown quantity within the Traceability Chain.

The difficulty is that most software documentation provides information on how to use the software and what type of inputs are required, but does not often expound on what the software does and how it does it. This type of information is crucial for a metrologist. In order to have complete confidence in output data from a processing routine, how it is treating the data must be made clear. Perhaps a flow chart should be provided indicating the chain of events, the equations used should be specified along with the algorithms which implement them. There are of course difficulties here since manufacturers of software are not too likely to release source code which is ultimately what we require for verification purposes. Therefore, this aspect of software provides a strong motivation for in-house development which allows complete control over all aspects.

It is also important to be able to access the data in a useful form. Verifying that the data displayed by the instrument is that which is received by the computer program is relatively easy, and the instrument itself can be verified by calibration with high-accuracy known standards. However, it should be possible to access the stored data independently of the software for verification and perhaps further processing. It may be necessary to graph the data using some other software in order to compare comparator variations with temperature variations, for example. To do this the format of the data must be known so that it can be extracted and parsed as required. We may also want to independently check the data to ensure that we agree with the results of the processing software. We need to know with what precision the data is printed to the file so that we can be aware of any rounding / truncating which may occur.

For a calibration laboratory with a large volume of work, there is certainly scope for software applications to enhance efficiency. However, as we have attempted to point out in the preceding paragraphs, there is a great deal of information about the software which is vital in order to have complete confidence in it. A "discontinuity" between software development and metrological requirements can lead to as many difficulties as are solved by using software. This, coupled with the need to design software specifically for a laboratory's needs provides a strong motivation for in-house development.

4. AN EXAMPLE

At the National Metrology Laboratory in Dublin a range of equipment is used in the mass calibration laboratory, facilitating calibration over a wide spectrum of mass values and to a wide range of accuracies. Various software packages had been tried and while each did have individual advantages, no single application met all requirements, which was the primary motivation for developing something more suitable. The practical design constraints needed were to ensure that

1. Data acquisition from all existing mass comparators should be possible, as well as from the ancillary equipment for measuring temperature, pressure and humidity. Additionally, there should be a modular approach that would allow further instruments to be supported in the future without significant alteration.
2. We needed a program that would run on a 'simple' computer running the DOS operating system rather than a Windows-based application. This was to facilitate the use of older office PC's which were being replaced. They are nevertheless quite adequate for the slow rate of data acquisition and simple mathematical analysis needed in calibration experiments.

The approach implemented was to abstract the data analysis and presentation sections from the data acquisition parts of the program. In this way it is possible to use the same user interface for all the comparators. The operator must provide information for each comparison, including calculating information such as reference weight value and density (selectable from a database) and test weight density; as well as archiving information such as data file names and names/serial numbers for the test weights.

Since this type of information is needed for all comparisons, irrespective of which comparator is used, there is no need to have different programs to process it. By developing our own software we were able to produce a total software solution, useful in all parts of the mass laboratory. There are of course some comparator-specific options that the user must set, perhaps relating to configuring the instrument for example. These are presented as required, depending on which comparator is selected from the supported list. Our software allows automated climate data acquisition so an additional option is provided, allowing the user to enable this feature if a buoyancy correction is needed.

4.1 Data Presentation

The data is presented in a standard format, irrespective of which comparator is used to obtain it. The most useful form for regular work where a small number of measurements are made is a tabular form where individual comparator readings along with the result of each double or single substitution cycle are given. The overall mean value and standard deviation are also presented, as shown in the screen-shot in Fig. 2.

On completion it is possible to view the final values for the calibration where the essential summary information is presented. This includes the mean "weight-in-air" value and the final value of the test weight, expressed as both physical and conventional mass. The screen-shot in Fig. 3 illustrates a typical data set. If a buoyancy correction is not required only conventional mass for the test is given. This information is printed to a data file in plain text format. This avoids any complications for printing and also ensures the data is easily and independently accessible.

A crucial feature of any mass calibration is the accuracy to which it is performed. The standard deviation of the measurement is the best indicator of this and can be compared with accepted values for the comparator at the given range. In our case, the laboratory's quality control measures have stipulated the maximum uncertainty allowed for each denomination, depending on its OIML class. With known accuracies for the buoyancy correction (if applied) and the reference weights, it is then easy to state the maximum allowed standard deviations for the comparator. These are stored in a file which the program consults on completion of the measurement and analysis.

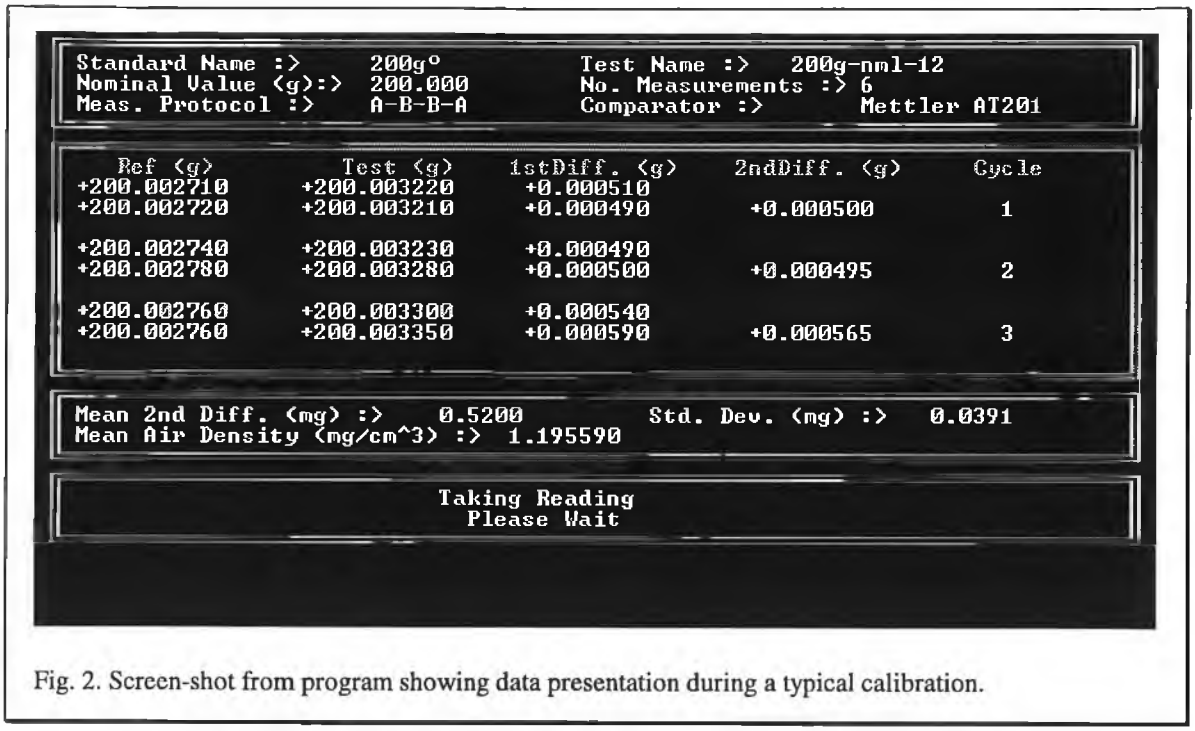


Fig. 2. Screen-shot from program showing data presentation during a typical calibration.

Fig. 3: Screen-Shot showing portion of the data file, where the essential summary information is presented.

A message is printed to the screen and to the data file either validating the result or warning that the standard deviation is too large. This is a convenient way for the operator to assess the acceptability of the data.

However, long term use of the automated comparators has shown that there can be drifts in the measurements over a period of several hours. Some examples of this

behaviour are shown in the following section. The standard deviation of just 6 or 10 measurements would not highlight this drift at all. Because of suspected systematic errors which arose in some high accuracy work, the program was modified, allowing the comparators to be operated for much longer periods and also allowing start and stop times to be specified. In this way it was possible to operate the instruments overnight when the laboratory would be empty and more stable. To deal with the larger amounts of data generated, the output sections were adjusted to give a graphical display. This was highly convenient as the user could glance at the output screen and immediately assess the stability and reproducibility of the comparator data. An example screen-shot from the program operating in this mode is shown in Fig. 4.

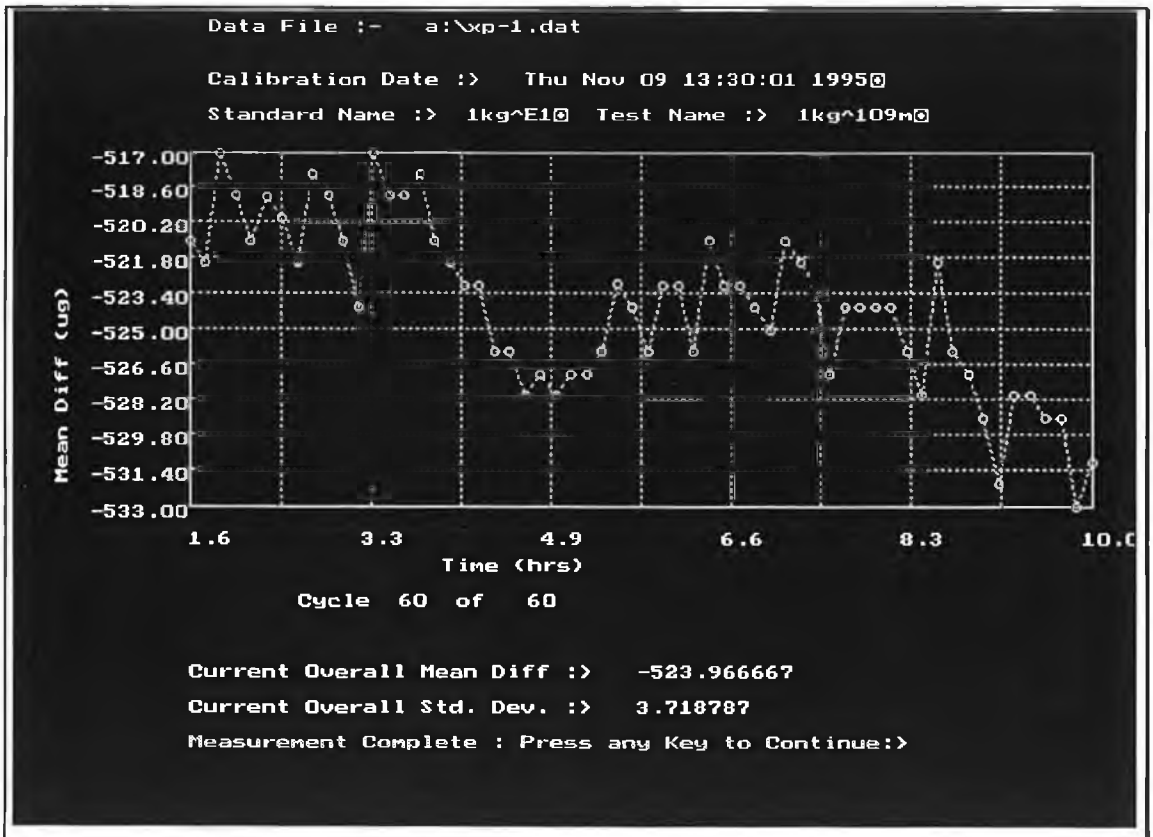


Fig. 4 Alternative data presentation in graphical format

The emphasis has been on developing a convenient, coherent code for the calibration laboratory which fulfilled specific requirements, while the facility to alter and adjust the program provides great flexibility for any special analysis that might be needed.

5. EXPERIMENTAL ANALYSIS

In the course of research work undertaken by the author it was necessary to perform "within-group" comparisons on a set of standards. In this method more comparisons

are performed than there are weights in the group, leading to an over-determined problem and the possibility of applying some suitable parameter estimation technique to obtain "best-fit" values for the standards in the light of the information obtained in the comparison experiment, [3]. [4]. In performing these calibrations, the automated mass comparators in the National Metrology Laboratory were used with the software described in the preceding section. Since the software allowed automated collection of temperature, pressure and humidity data, as well as the mass comparison data, it was possible to obtain large amounts of data with little effort. The computer could calculate air density from the resulting data, apply a buoyancy correction and hence produce a data file of true mass differences for the weights used in the comparison. It was usual to operate the comparators overnight when conditions would be more stable and as a result typically 8—12 hours of data would be available for each comparison.

Fig. 5 graphically presents the information available from a typical comparison. The 'weight-in-air' and buoyancy-corrected true mass differences are shown along with temperature, pressure, humidity and calculated air density data. The great advantage of the software used was that it was easy to access the raw data in order to process it or present it in any desired format. The graphical format makes it easy to analyse the data and investigate the behaviour and stability of the mass comparator.

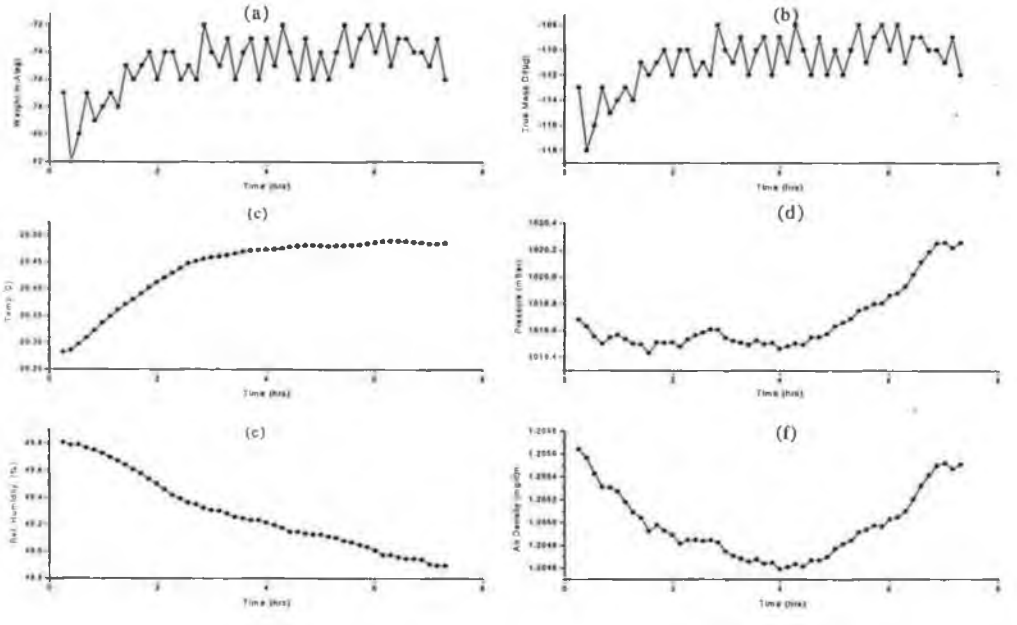


Fig. 5: Shown here is the relevant data for a typical comparison experiment. (a) shows the uncorrected "weight-in-air" difference from the comparison. (b) gives the buoyancy-corrected true mass difference between the standards. (c) is the temperature during the course of the comparisons, obtained with a probe within the weighing enclosure. (d) gives the corresponding barometric pressure while (e) gives the relative humidity, also within the weighing enclosure. The calculated air density is shown at (f). This comparison involved a 1kg standard of density 8050.3 kg/m^3 & two 500g stackable weights of density 8048.3 kg/m^3. All standards are of E1 classification.

A striking feature of the two mass-comparison graphs at the top of Fig. 5 is the drift during the first two hours of measurement (graphs (a) & (b)). The fact that this is not corrected in the true-mass difference graph (graph (b)) suggests that it is not buoyancy-related. It appears that the comparator needs to operate for about 13 double-substitution cycles in order to reach stability. After this time its stability is only limited by the resolution of the instrument (2 μg in this case). The temperature probe was within the weighing enclosure of the comparator during the measurements, and from the temperature graph of Fig. 5 (graph (c)) we can see a characteristic rise of ~ $0.2K$. We can correlate this with the mass comparison graphs and conclude that there is a systematic drift in the comparator before it reaches equilibrium. This phenomenon appears in spite of taking appropriate precautions such as ensuring the weights reached thermal equilibrium within the weighing enclosure before beginning comparisons. The source of the problem may perhaps be temperature gradients within the weighing enclosure due to the load alternator motors. For example, Fig. 6 shows the temperature within the weighing enclosure before, during and after a measurement period. The temperature rise during the comparison is clear from this. The practice adopted as a result of this behaviour was to take 6 or 10 measurements from a stable portion of the graph and to use these to obtain a mean mass difference and standard uncertainty for the comparison.

During the analysis of one group comparison exercise significant systematic errors among the input data for the parameter estimation method were noted. This was in spite of care having been taken to perform measurements for sufficient time to stabilise the systematic drift mentioned above. Therefore some of the measurements were performed again and surprisingly did not agree at all with the previous data, all corrections having been applied. Indeed several comparisons for a pair of 1kg weights produced different results as shown in Fig. 7. An obvious reason for this did not appear, but graphing the physical mass difference as a function of relative humidity revealed a clear correlation. This is illustrated in Fig. 8 where the root of the problem is shown to be that the humidity levels in the laboratory were too low at the time of the measurements, and the performance of the comparator was suffering as a result. A systematic error, possibly due to an electrostatic effect, was being introduced, and because not suspected or quantified, could not be corrected in the subsequent analysis.

For the purposes of this paper, the important conclusion we wish to draw is the great benefit of convenient software for data acquisition and analysis. Much of the information presented in this section would have been difficult or impossible to obtain without the help of computer power, properly harnessed for the particular experimental requirements. Some of the effects noticed would introduce significant systematic errors in the calibration results if corrective action were not taken.

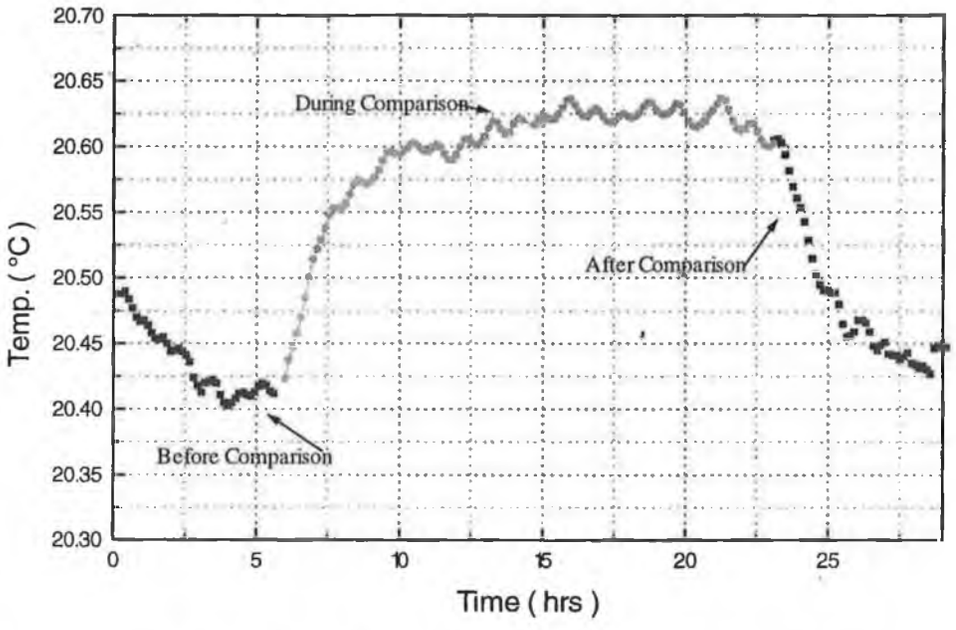


Fig. 6: Temperature plot for before, during and after a typical comparison experiment.

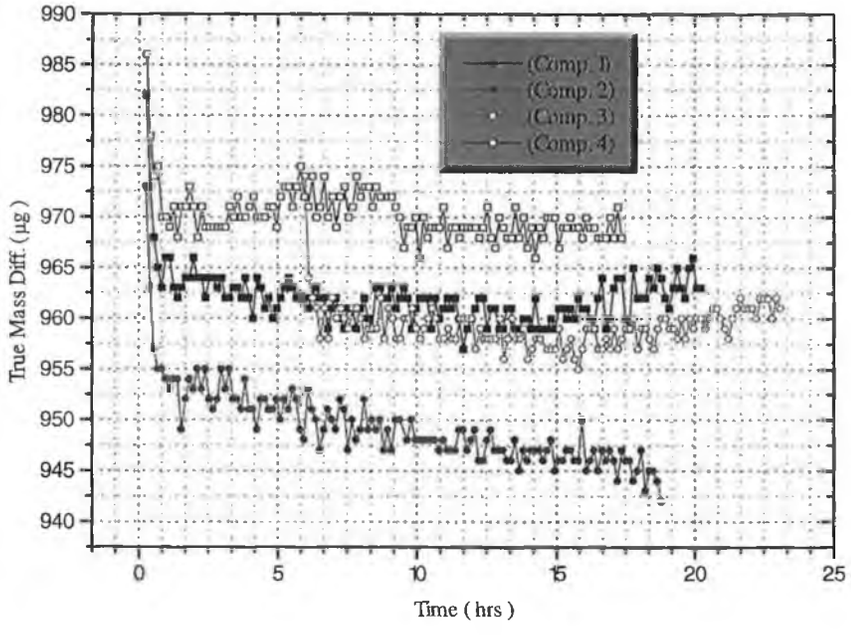


Fig. 7: True mass difference for four comparisons of the same pair of 1kg standards. In spite of applying buoyancy corrections to the data (which was all obtained at different times), significant inconsistencies remain unaccounted for.

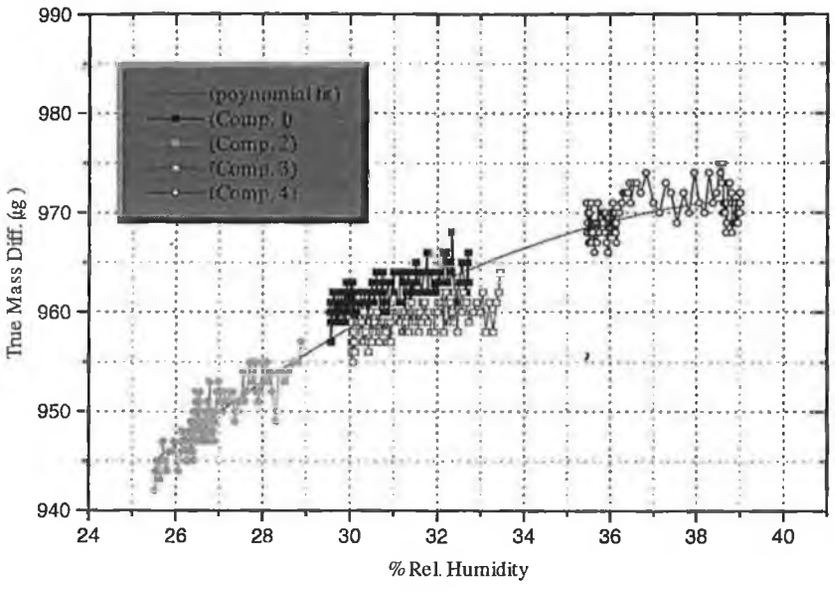


Fig. 8: Plotting the True Mass difference as a function of of % Rel. Humidity for each comparison highlights a clear correlation. This shows that the comparators are effected by the ambient humidity and that values of relative humidity much below 40% lead to problems with the data

6. CONCLUSION

In this paper we have briefly discussed some of the software requirements for the mass calibration laboratory. Because of the variety of equipment in use it is difficult to obtain a total software solution from an external source. We have shown by example how the internal development of a software package gives great flexibility in designing an application which will meet individual laboratory requirements. The added benefit of access to source code helps to ensure the transparency necessary for traceability within the metrological system. We have also highlighted some insights into the operation of the automated mass comparators in our own laboratory in which the use of suitable data acquisition software was instrumental.

7. REFERENCES

[1] Norbert Grief *et al,* "*Software problems in calibration services–a case study.*" "Advanced Mathematical Tools in Metrology III, Berlin, Sept. 1996".

[2] B. Beizer, "*Software testing techniques*", Van Nostrand, New York, 1990.

[3] T. S. Leahy, "*An Analysis of Statistical Estimation Methods used in High Accuracy Mass Determination*". In: P. Ciarlini, *et al*, Eds., "Advanced Mathematical Tools in Metrology II", World Scientific, Singapore, 1996

[4] T. S. Leahy "*A unified approach to statistical estimation and model parameterisation in mass calibration"* Post-graduate research thesis, in preparation.