NPL Report CISE 39/99

Report

to the National Measurement System Policy Unit, Department of Trade & Industry

INTERIM STATUS REPORT ON THE VALID ANALYTICAL MEASUREMENTS AREA FROM THE SOFTWARE SUPPORT FOR METROLOGY PROGRAMME

BY

W. A. HARDCASTLE,

S. L. R. ELLISON, LGC

AND

D. RAYNER, NPL

September 1999

Software Support for Metrology

Interim Status Report on the Valid Analytical Measurements Area

by W A Hardcastle, S L R Ellison (LGC) & D Rayner (NPL)

September 1999

Abstract

This report describes the status of the mathematics and software used to support the Valid Analytical Measurement (VAM) area of metrology, particularly as used within the NMS VAM programme. This is one of a set of status reports produced for the NMS Software Support for Metrology (SSfM) (1998-2001) programme mid-way through the current programme to inform the formulation of the next SSfM programme (2001-2004). The different aspects of mathematics and software are reviewed under the headings of the themes and project topics of the current SSfM programme. The SSfM programme is identifying best practice where it exists and disseminating guidance on that best practice to other metrology areas. The outputs of the SSfM programme will be generic, applicable to more than one metrology area. This report, therefore, not only identifies problems to be tackled and best practice to be disseminated by the SSfM programme, but also if appropriate possible future VAM programme projects applying SSfM outputs to specific problems in the VAM area.

ISSN 1361-407X

Authorised by Dr Dave Rayner,
Head of the Centre for Information Systems Engineering

National Physical Laboratory, Queens Road, Teddington, Middlesex, TW11 0LW

Contents

1. INTRODUCTION 1

2. SCOPE OF THE AREA COVERED 1

3. MODELLING TECHNIQUES 3

3.1 *Methods for Modelling Measurement data* 3
3.1.1 Types of Data 3
3.1.2 Modelling 5

3.2 Uncertainties and Statistical Modelling 9
3.2.1 Derivation of Statistical Models 9
3.2.2 Determination of the statistical distribution of measurement results 10
3.2.3 Development of Calibration Curves 10
3.2.4 Comparison of measurement results with limit values 11
3.2.5 Levels of confidence other than 95% 12
3.2.6 Supporting software 12
3.2.7 Supporting documentation 12
3.2.8 Training 13

3.3 Visual Modelling and Data Visualisation 14

3.4 Data Fusion 14
3.4.1 Applications 14
3.4.2 INTErSECT awareness 15

4. VALIDATION AND TESTING 16

4.1 Spreadsheets and Other Mathematical Software Packages 16
4.1.1 ‘Off the shelf software’ 16
4.1.2 Spreadsheets 17
4.1.3 Awareness of limitations 17
4.1.4 Testing and validation 17
4.1.5 Supplier advice 17
4.1.6 Traceability 18

4.2 Model Validation 18

4.3 Measurement System Validation 19
4.3.1 Embedded software 19
4.3.2 System validation 19
4.3.3 Reference data 19
4.3.4 Validating embedded software 19
4.3.5 Supplier guarantees 20
4.3.6 Supplier validation of embedded software 20
4.3.7 Access 20

5. METROLOGY SOFTWARE DEVELOPMENT TECHNIQUES 20

5.1 Software development methods 20
5.1.1 Software Quality 21

5.1.2 Management issues 21
5.1.3 Technical issues for software developers 22

5.2 Software reuse 23
5.2.1 Management issues 23
5.2.2 Technical issues for software developers 24

5.3 Virtual Instruments 24

6. SUPPORT FOR MEASUREMENT AND CALIBRATION PROCESSES 24

6.1 Automation of Measurement and Calibration Processes 24
6.1.1 LIMSs in use 25
6.1.2 Other processing methods 26
6.1.3 Duration of use of LIMS/other systems 26
6.1.4 Advantages and disadvantages 26
6.1.5 Finding a suitable LIMS 26
6.1.6 In-house development 27
6.1.7 Adherence to appropriate standards and procedures 27

6.2 Format Standards for Measurement Data 27
6.2.1 Need for standard data formats for measurement data 27
6.2.2 Need for sharing measurement data with other organisations 28
6.2.3 Need for a common format for calibration records 28
6.2.4 Need for a common format for exchange of measurement data 28
6.2.5 Different data formats 28

7. SUGGESTIONS FOR FUTURE ACTIVITIES 28

7.1 Expected Benefits of the SSfM Programme 28

7.2 Case Studies and Feasibility Studies 29

7.3 Future SSfM Topics 29

7.4 Future VAM Programme Topics 29

8. SUMMARY OF CHANGES 30

9. ACKNOWLEDGEMENTS 30

APPENDIX 1: GLOSSARY OF ABBREVIATIONS 31

APPENDIX 2: BIBLIOGRAPHY 33

APPENDIX 3: MODELLING TECHNIQUES - TABLES 34

1. Introduction

The purpose of this status report is to inform the National Measurement System (NMS) Software Support for Metrology (SSfM) (1198-2001) programme about the status of the mathematics and software used to support the **Valid Analytical Measurements** (VAM) area of metrology, concentrating on what mathematics and software are used within the NMS **VAM programme**. This and the companion status reports for the other metrology areas will inform the formulation of the next SSfM programme (2001-2004). It may also lead to appropriate linkage between the VAM programme and the SSfM programme.

The SSfM programme is an underpinning programme that provides generic support in the use of software and mathematics to the NMS Programmes for each metrology area. For details of the programme, its expected deliverables and the results already produced, see the SSfM web site: http://www.npl.co.uk/ssfm/.

The NMS Programmes for specific metrology areas provide metrological support to industry. The SSfM Programme has some direct impact upon industry, as evidenced by the SSfM Club membership. This relationship is depicted in Figure 1. It is because of this relationship that the Status Reports concentrate primarily on the use of software and mathematics in the other NMS Programmes.

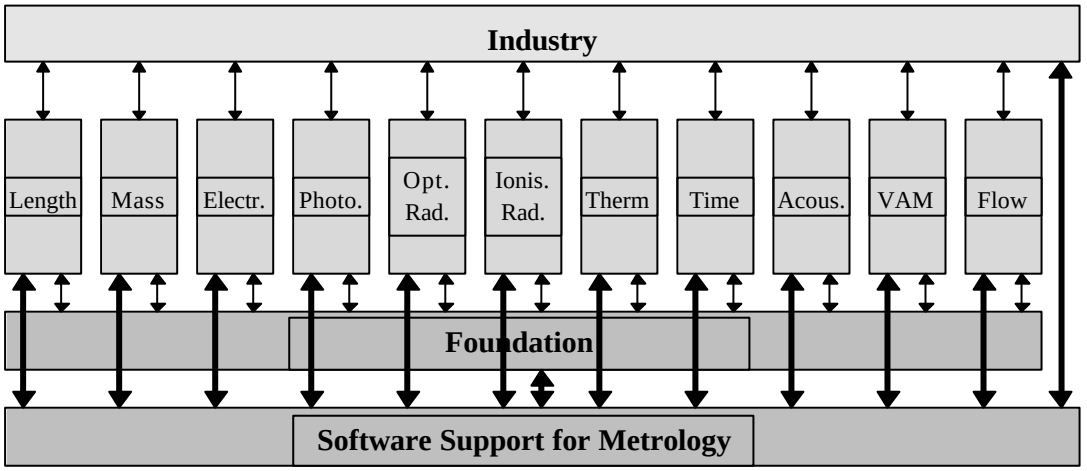

Figure 1. Relationship of SSfM to other NMS Programmes and Industry

In particular, this report addresses each of the themes within the SSfM programme and describes the status concerning the topics covered by each of the relevant projects. It also considers whether there are any important software or mathematics issues in the VAM area which are not addressed by the current SSfM programme.

This report is an update of an initial restricted status report produced in November 1998. That initial report was one of a set of restricted status reports which were synthesised into an overall status report for all metrology areas [13]. A summary of the differences from the initial VAM status report is provided at the end of this report.

2. Scope of the Area Covered

Valid Analytical Measurements - Scope

Analytical measurements are those made in the pursuit of analytical science; they are obtained using a broad range of techniques and are spread across several disciplines. The term ‘Valid

Analytical Measurements' was originally applied to the field of analytical chemistry and its sub-disciplines such as clinical chemistry. However, it should be noted that the boundaries of this discipline are diffuse and frequent use is made of techniques from other disciplines, for example, the biochemical and physical sciences. Although analytical chemistry has had the longest exposure to the concept of Valid Analytical Measurements, this report covers the whole scope of the current NMS VAM Programme, including both analytical chemistry and the physical analytical measurements.

Analytical chemistry is that branch of applied chemistry concerned with investigating the composition of materials. It is made up, traditionally, of two main strands: qualitative analysis and quantitative analysis. Qualitative analysis consists of activities designed to reveal the identity of a material or to show that a stated substance is a component of a given material. It often involves making quantitative measurements in order to establish whether the value of a given property is above or below a reference level. Quantitative analysis consists of activities designed to determine the amount of a particular component in a given material or the magnitude of a physical property.

In order to quantify the amount of a given substance it is necessary to measure an appropriate property of the substance and relate this, via a calibration, to a similar measurement on a known amount of the same substance. For the purposes of this report measurements are considered to be made on chemical, biochemical or physical properties of substances.

Physical analytical measurements within the NMS VAM programme include gas standards and gas analysis, surface analysis, measurements of particulates and aerosols, and electrical methods of measuring pH.

The physical analysis of gases also has quantitative and qualitative aspects. Quantitative analysis is based upon the use of primary gas standards, prepared using absolute gravimetric techniques so that concentrations are traceable to the primary standard of mass. In addition, there is quantitative analysis of particulates in gases. Qualitative analysis concerns the analysis of odours using novel techniques with the potential, eventually, to identify an odour measurement scale.

The physical analysis of surfaces of solids is a relatively new subject for which reference materials, as the basis for calibration of measurement instruments, are largely inappropriate. Instead, it is necessary to work from first principles in analysing the measurement process as a whole.

The measurement of particulates and aerosols includes the accurate measurement of ultrafine particles, sampling and measurement of bio-aerosols, and support for calibration of particulate measurement instrumentation. Apart from a calibration service for instrumentation measuring ultrafine particle concentration, the work mostly involves report and reference material production and hence does not require significant software or mathematics support, apart from that used in automation of the calibration process.

The work on electrical methods of measuring pH is based on measurements with a well-characterised hydrogen electrode system. Part of the internationally-agreed method requires a series of three or four values for the "acidity function" made at different added chloride concentrations to be extrapolated to calculate the effect for the hypothetical "zero added chloride" case. Mathematically, this procedure requires an extrapolation of a regression fit where the x and y variables (and their uncertainties) are strongly correlated. Statistically robust procedures that are capable of doing this and calculating the uncertainty have not yet been developed. Work on this problem is currently underway at another metrology institute (PTB) but no results are available at present. It is expected that this problem will require significant work in the future since it underpins international pH scales.

The software, and where appropriate the mathematics, employed in the VAM area falls into three main categories of use:

- data acquisition - including instrument control;
- data processing - including modelling, fitting, optimization, and evaluation of uncertainties;
- report generation.

Application areas covered by the VAM programme include, but are not limited to:

- forensic science;
- food & nutrition;
- pharmaceuticals;
- environmental monitoring;
- chemical industry;
- aircraft engine performance;
- water industry;
- clinical analysis and medical measurements;
- health & safety;
- patent protection and litigation;
- process control;
- R&D support.

This report does not cover R&D *per se* and hence excludes such topics as molecular modelling, computational (theoretical) chemistry and structure/activity relationship (SAR) modelling. These areas make intensive use of complex mathematical models and sophisticated software. The issues here are extremely diverse and are beyond the scope of the present report.

3. Modelling Techniques

3.1 *Methods for Modelling Measurement data*

3.1.1 Types of Data

There are three categories of measurement data sets:

1. discrete measurement data, which means:
   a) data obtained by sampling a discrete variable: e.g., a variable that may take only integer values such as in counting processes, and
   b) data obtained by sampling a continuous variable.
2. continuous measurement data, which means an analogue signal prior to any analogue to digital conversion.
3. hybrid measurement data composed of both discrete and continuous data sets.

In the VAM area, although some of the quantities for which data are collected are continuous variables (such as temperature and time), the data that are collected are nearly always discrete.

Sometimes the raw data are discrete. However, quite commonly spectral techniques are used, in which case the raw data are analogue and hence continuous, but this analogue data are fed into an analogue-to-digital (A/D) converter, resulting in large amounts of discrete data. In addition, occasionally, hybrid data may be needed for some applications.

Discrete data can usefully be categorised further into univariate (or *scalar*) data and multivariate, array or vector data. These categories are considered below.

Discrete data

Discrete data can be subdivided into numeric and non-numeric, the latter referring to the kind of data generated by qualitative analysis where the purpose is to classify an object. Classifications are generally bivalent and produce a 'yes/no' answer. Since, in the sense of this report, classifications do not feed into a model, this type of data will not be considered further here (although note that they may be included where chemometric methods of data processing are applied).

Quantitative analysis generates numeric data which principally characterises concentrations or amount of substance. A second significant type of numeric data is that describing the value of physical properties of a substance. These types of discrete data constitute the main class of intermediate and reported data in this area.

Examples of discrete data in the VAM area are:

- colony counts (microbiology);
- radioactivity counts (radiochemistry);
- particle counts (environmental chemistry).

In the VAM area, statistical treatment of such data uses appropriate discrete distributions. Modelling, for example of trend data or for calibration, is more likely to treat the data as approximately continuous. This is almost invariably justified by the large values involved; the intervals between data points are then very small compared to their values, the relevant distribution functions rapidly approximate normality, and the resulting uncertainties become negligible compared to variation from other causes.

Most quantitative data in the VAM area are related to variables that are essentially continuous, such as mass, concentration, time, temperature etc. However, such data are often collected as discrete data by digital sampling and are modelled using either discrete or continuous models as appropriate. Although, in principle, digital sampling of such parameters may result in numerical inaccuracy, in practice, digitisation and sampling intervals are very rarely a significant source of uncertainty.

Scalar and array data

Analytical measurements generally aim to provide one or more scalar quantities, usually concentrations. However, many instrumental methods of analysis generate data in which the value of a property varies continuously with some other property such as wavelength or time. The two most common classes of analytical methods producing this type of data are spectroscopy and chromatography. Specific examples of such data in analytical measurement include

- 'visible' spectra (~ 400 -700 nm);
- gas-liquid chromatograms;
- autotitrator pH/titrant volume curves.

Often, such data are treated on-line, for example by on-line integrators operating one data point at a time, and the data are not retained. Where the data are acquired and retained for subsequent processing, the output forms an array. In a large number of cases, whole-spectrum data are acquired as a function of time or some other property. Gas chromatography coupled with mass spectrometry is a well established example in which complete mass spectra are acquired rapidly over time, forming a large array (vector) of spectral values. In addition, array data such as spectra, or multivariate data consisting of a series of different property values associated with individual test items, may be subjected to modelling for quantitative calibration or for classification use. Acquisition, pre-treatment and processing of multivariate data in chemistry is generally referred to as *Chemometrics*; this topic is further discussed under 'data fusion' below.

Continuous data

In some cases, analogue data are collected and used without A/D conversion as continuous data.

Hybrid data

Hybrid data are not commonly employed in analytical measurements. Some interpretative modelling may use combinations of discrete and continuous data, but in most cases the continuous data will be converted to discrete (see above). In exceptional cases, classifications may be based on hybrid data; for example, in categorising the country of origin of drugs of abuse, the presence or absence of a particular 'cutting agent' may be combined with the concentration values of drug metabolites to classify samples.

3.1.2 Modelling

Types of model

There are three categories of model:

1. discrete models in which the outputs of the measurement system are related to measurements of its inputs by a system of algebraic equations;
2. continuous models in which the outputs of the measurement system are related to measurements of its inputs by a system of differential equations;
3. hybrid models in which the outputs of the measurement system are related to measurements of its inputs by a system composed of both algebraic and differential equations.

It is appropriate to distinguish here between the nature of the model and the approach to solving the model (which for a continuous model typically also involves solving algebraic equations).

Given infinite precision arithmetic, a discrete model can generally be solved exactly. This is not the case for a continuous model which typically requires approximations to be made, for example, in terms of domain discretisation.

Models for analytical measurements

Most analytical measurements take a number of single input values, such as masses, volumes, and intensities, and combine these to give the desired result. In general, this activity does not constitute 'modelling', though the resulting relationship may well constitute a partial model in the sense defined in the ISO Guide to the expression of uncertainty in measurement (GUM) [1]. Calculating a mass concentration from an observed mass divided by an observed volume is an example; though based on a very fundamental theoretical model which defines concentration as a linear function of mass per unit volume, it is not necessary to build and test a fitted model. However, the parameters may themselves be derived from fitted models, and these are considered further below.

It is useful to distinguish between models used to obtain values from raw data, models formed for calibration and prediction of measurement values, models used in interpretation, and measurement system models used for uncertainty estimation, and these different situations are considered separately.

Obtaining values from data

Quantitative intensity values are drawn from array data by several methods. Single point sampling, for example the estimation of peak height from a single data point within a spectrum or chromatogram, is simple, frequently used, and requires no mathematical modelling. Integrated intensities are most commonly obtained simply by summing and scaling a range of data values within a specified range. Again, mathematical modelling is unnecessary; the values are taken directly from the data and not from a fitted model. Of course, although some of these simple procedures apply no explicit model, elements of model building or model validation are nonetheless implicit or assumed. For example, selecting a local maximum from a spectrum as a measure of peak location or intensity implies a model based on peaks with defined maxima. Such simple implied models are typically validated by experiment, for example by adjusting sampling interval and observing the effect.

In a few cases, including spectroscopic integration in the presence of overlapping peaks, curve fitting may be used to obtain reliable integrated intensities. Analysts are more likely to avoid the necessity of curve fitting by choosing a system in which overlap is small. Where fitting to a model is done, it usually follows least squares fitting based on well established iterative numerical procedures. It is almost always possible to find an appropriate and sufficiently robust algorithm for a given application.

Further use of the data in calibration etc. of course corresponds more closely to mathematical modelling; this is dealt with below.

Calibration models

The majority of calibration models in the VAM area are linear relationships between scalar input and output quantities. Where possible, measurement systems are chosen or developed to give linear calibration models. There are exceptions, however. Immunoassay techniques, for example, produce sigmoidal response curves which are fitted by physical (e.g., kinematic) or empirical (e.g., log-logistic) models. Because these models are non-linear in their parameters, iterative methods (see below) are used to obtain the fit. Comparatively large statistical variability in the responses, together with exponential terms in the underlying model, can result in instability. Another class of calibration model is multivariate calibration, used increasingly in practical field measurements but not generally in standards calibration. Multivariate calibration establishes relationships between vector or multivariate observed data (usually spectra) and one or more properties of interest.

There is also interest in the development of calibration curves; this is addressed under Uncertainties and Statistical Modelling, section 3.2.

Interpretation

Modelling may be used to assist in interpretation of data, for example to establish trends or aid classification. Trend analysis is usually application driven and not undertaken for measurement purposes, so is not discussed in the present report. Classification is, however, common in regulatory and other applications of analysis, and may require construction of classification models based on multivariate data.

Measurement process modelling

In the VAM programme work on surface analysis, the complete measurement process is being modelled; these models are very much the subject of current research and development. In all other areas of the VAM programme, measurement processes are not generally modelled prior to routine development, except in that the basic relationship required to calculate the result from input parameters is written down. The approach to development is almost invariably practical and based on experience, not on prior modelling. There are well developed and applicable theories behind the physical methods used, but these are not generally useful in practical situations except to indicate limitations of a given technique. This is largely because the major influences on analytical measurement - effect of sample matrix, operator, interpretation of methods etc. - are not amenable to predictive mathematical modelling. This has a profound impact on studies of reliability and uncertainty. While good progress has been made in interpreting the ISO GUM [1] for chemistry via a EURACHEM document [2], measurement process models in the VAM area can usefully be based on measurable input quantities only in developing quite specific methods for individual test materials and determinands; typically, this applies only in reference measurements, and very rarely to routine measurement. The dominant approach to uncertainty and reliability study of both methods and reference materials is inter-laboratory comparison, supplemented by extensive in-house experimental studies ('method validation' for measurement methods). Studies of this kind model the process as a 'black box' influenced by factors such as Laboratory, Analyst, etc. This issue is discussed further below (see section 3.2, Uncertainties and Statistical Modelling). Recent work in the VAM programme area does, however, suggest a need to improve the application of modelling, particularly of substrate and related effects in reference materials characterisation; future work may accordingly result in increased modelling activity within the programme and eventually beyond.

Models currently used in the VAM area typically have a simple relationship between the input and output variables which themselves are often few in number. While all of the variables may carry measurement error, and may have considerable scatter (e.g., because of inhomogeneity), the measurements of their values usually have only small errors. Surface analysis is, however, an exception, sometimes having more complex relationships between input and output variables, sometimes having a number of inter-related parameters, and sometimes having large measurement errors. Thus, surface analysis would score 2 to 4 for each of the rows of Table 1.

The Normal, or Gaussian, distribution is the most common, but is not universal. Poisson distributions are expected for low level counting studies and are common in surface analysis (indeed surface analysis would score 2 in the penultimate row of Table 1 if the question was modified to relate to either Gaussian or Poissonian distributions). Most other cases are considered to follow a Gaussian distribution, except where a non-Gaussian distribution is expected from the underlying physics or chemistry. Even where a Gaussian distribution is expected, outliers are common, and this must be recognised, where appropriate, leading to data treatment assuming that the distributions are really non-Gaussian. The Gaussian approximation typically holds reasonably well up to a level of confidence of approximately 95%, but breaks down sharply at higher levels. This can often be attributed to human or other factors causing large errors in a small percentage of measurements. Irrespective of the origin of outliers, their presence requires suitable treatment of the data. Another complicating factor is that many practical measurements are made near a boundary value; concentrations below zero or above 100% are physically impossible. While the error may still be thought to be Gaussian, symmetrical estimates of uncertainty in the measurand value are not appropriate when the uncertainty interval would span such a limit.

Input quantities to the calculated result of measurement are, in general, well understood and all major influence factors identified. In particular, the best standards measurements, these parameters include all significant effects. However, for more routine measurements, particularly where different sample matrices are tested using the same method, uncertainties in the input

parameters to the calculated result are very often insignificant compared to the observed inter-laboratory distribution of results. This follows from a general approach in analytical method development; where an influence factor has an effect and can be controlled, it is invariably controlled to reduce its effect to insignificance compared to the available measurement precision.

The above scores apply to all of the VAM area with the exception that the first score does not wholly apply to surface analysis; in that area the first score could range from 2 to 5, depending on the technique used.

As noted above, measurement models employed in the Valid Analytical Measurement area are of several types. Basic calculations rely on such well established principles that further modelling is not necessary. Simple linear calibration models are usually widely accepted and based upon well understood chemical and physical theory, though in some systems it is recognised that some slight curvature is inevitable for instrumental reasons. Here, the use of quadratic fits is accepted as a pragmatic means of extending working range, even though there is no underlying theoretical basis for that form; nevertheless, such fits can be validated by demonstrating, where appropriate, that the inclusion of higher (cubic or quartic) polynomial terms gives no significant improvement. Some other calibration models (such as sigmoid response curves or the kinetic model) are based on well established theory, whilst others (e.g., log-logistic) are in the main empirical. In some routine monitoring applications, Artificial Neural Networks (ANNs) are becoming important tools; these may be regarded as arbitrary models with no underlying physicochemical basis (i.e., they are purely empirical models). Models for interpretation and classification (see above) may be complex in terms of the number of variables, but are likely to be based on simple functions (linear, quadratic); here too, ANNs have been employed.

Measurement process models may be very pragmatic, using functions such as:

$$\text{Observation} = \text{True value} + \text{laboratory bias} + \text{random error}$$

or may include individual influence quantities. It is common to find that major influences are not covered by adequate theory; for example, there is no reliable theoretical basis for analyte extraction from solids, which accordingly remains a major unpredictable influence. There are, however, simple measures of the overall efficiency of that process which can be used if required.

In analytical chemistry, no allowance for uncertainties is made in basic calculations, as this is not generally necessary; these calculations do not generally arise from fitted models, and the uncertainty does not affect the calculated result. In calibration modelling, statistical variability is usually the sole consideration, though it should be noted that this is because matters are generally arranged to ensure that other uncertainties in calibration are negligible. In both cases, other uncertainties are taken into account when required to estimate measurement uncertainty.

Measurement process modelling is very important in surface analysis but otherwise is generally undertaken only for uncertainty estimation purposes, and is considered in section 3.2.

The methods of model solution are typically sufficiently comprehensive for the intended purpose of the result. The solution methods do not generally take into account overall measurement uncertainty components or correlations amongst the input quantities. Weighting is not common, even where heteroscedasticity is recognised. However, it should be noted that uncertainties from this cause are assumed to be very small compared to other factors, although this assumption is rarely addressed.

The methods of model solution generally perform as expected. This is not surprising given that the dominant method is that of linear regression and that care is taken to ensure that (analytical) methods are operated within their linear region. The results from such calculations can easily be checked using independent software applications. A small number of analytical methodologies

produce a non-linear response which has to be fitted to a particular generic model (for example, a four parameter model having the form of an exponential decay function is used in connection with urine test kits). It is not unknown for the commercial software used in the above example to fail to converge when presented with a standard set of iteration parameters. Convergence can always be achieved by adjusting the iteration parameters but care is needed to ensure a sensible (and correct!) result is obtained.

The software used to solve the models executes rapidly and perturbing the input data generally returns a nearby solution. Statistical uncertainties based on non-iterative calculations are consistent with the 'noise' levels.

In many areas of surface analysis, although the scores in Table 6 do not necessarily apply, development of a valid model is the objective of the work.

In analytical science the measurement model is frequently a regression model (most commonly linear least squares) which is solved by a software application. In general, analysts do not attempt to validate the model independently of the method. The way analysts work in this case is to validate an analytical method and this has the effect of indirectly validating the model.

The special case of multivariate modelling is considered further in section 4.2.

3.2 Uncertainties and Statistical Modelling

3.2.1 Derivation of Statistical Models

There is considerable knowledge and experience of estimating measurement uncertainties in the gas analysis field. In other fields, the production of measurement uncertainty information is a current area of concern for those developing new analytical methodologies. Apart from in the gas analysis field, measurement uncertainty is a relatively new area for analysts and so its use, as implied in the ISO GUM [1], is not yet widespread. Analytical laboratories typically operate many different methods (frequently hundreds) and determining uncertainty estimates for results from these methods, based on ISO principles, can be extremely time consuming. In addition, it is commonly found that the main influences on analytical measurements are not associated with traceable measured input quantities. In the VAM area, reliability has historically been established through experimental studies of method performance ('method validation') with no formal intent to estimate a single uncertainty figure. Indeed, the range of method performance parameters determined during method validation studies is substantially more informative to a trained analyst. This kind of study still represents the mainstay of reliability assessment in analytical work, and that is unlikely to change quickly. Nonetheless, the recent emergence of measurement uncertainty as a key requirement of accreditation bodies has generated substantial effort to apply the concept in the VAM area. The following discussion accordingly describes current practice in a developing area.

Measurement uncertainty estimation in the VAM area follows the principles of the ISO GUM [1] to the extent achievable with available information. Models based on partitions of total observable variability, plus allowances for systematic effects based on comparisons with reference standards measured against the whole method, allow 'simple' models to express overall uncertainties effectively without detailed knowledge of individual input parameter uncertainties. With that caveat, it may be said that all significant sources of uncertainty can be modelled, whether the source is fully understood or not. In higher calibration standards work, for example in gas standards and surface analysis work, or in reference measurements of individual molecular entities in well characterised materials, it is practical to establish input parameter based models incorporating all significant effects.

In uncertainty estimation, variances are considered and taken into account. Correlations are not normally a concern in analytical chemistry, but are an important concern in gas and surface analysis. Even in analytical chemistry, correlations do in fact exist, but it is usually possible to reformulate the model to remove correlation effects (for example, an input mass determined by difference can be constructed from two individual values with correlated uncertainty due to systematic offset, or can be treated as a single 'mass by difference' value for which the uncertainty can be established without other correlations). In some cases controllable effects are reduced to insignificance during method development. However, in gas and surface analysis, work is on-going to understand and take account of correlations more fully. When differences between masses are used to eliminate systematic effects in gas analysis, the result can be that there is still a correlation of random effects among the various differences.

3.2.2 Determination of the statistical distribution of measurement results

In the VAM area, many statistical distributions are non-Gaussian. Though most such problems in field measurements can be dealt with by routine outlier checking or related measures, there remains a significant number of problems, particularly at reference and calibration level, for which the underlying distribution is truncated or non-Gaussian and must be treated as such. If this is to be handled within the ISO *Guide* framework as required by BIPM and accreditation agencies, some emphasis needs to be given to developing guidance which is complementary to the ISO *Guide* but which deals with non-Gaussian distributions.

It is recognised, if not widely, that expanded uncertainties calculated from first order expressions are not necessarily exact. This is not considered a problem however, given the nature of chemical analysis and the approximations and assumptions that often need to be made in order to obtain the combined uncertainty. Further, it is recognised that input uncertainties based on statistical observation are likewise only estimates, with their own uncertainties; the sampling distribution of standard deviation estimates of moderate degrees of freedom (up to 30) is in most cases expected to be far greater than the approximations inherent in the model equations. It is noted elsewhere (section 3.1.2) that a simple expanded uncertainty is not very appropriate in some important circumstances (values close to fundamental limits in particular). Correcting the approximations in the ISO *Guide* cannot be regarded as a priority endeavour in the VAM area, though a practical means of expressing uncertainty intervals near fundamental limits would be valuable.

3.2.3 Development of Calibration Curves

The field of gas analysis involves extensive use of multi-component calibration methods typically using between 5 and 7 standards to calibrate analysers that operate over dynamic ranges of x100 to x10,000. This is a specific example of the requirement to develop a "calibration curve" which can be used to calculate the appropriate "output" value for any measured unknown. The established method for developing the calibration curve from the measured data is by carrying out a regression of the measured values on the calibration (standard) values using ordinary least squares (OLS).

The increasing requirement for reduced uncertainty in gas analysis, particularly for the measurement of natural gas which has substantial economic impact, has led to various NMI's reviewing the use of OLS for this application. Important developments include:

- the use of generalised least squares (GLS) methods to develop calibration curves that take appropriate account of the uncertainties in the standards as well as the measured values. This method has now been mandated by ISO 6143 (developed by ISO TC 158);

- the use of weighted OLS to take account of known variations in the uncertainties of the measured values;
- the treatment of the results of OLS in a way that enables the results to be presented within the framework of the ISO GUM [1]. This method is being developed for incorporation in ISO 11222 which is being prepared by ISO TC 146.

Since the motivation for the implementation of these more sophisticated techniques is to reduce the uncertainty in the final result, there is an important requirement that they should be subjected to appropriate scrutiny before they are recommended for widespread implementation. Particular technical issues that are overlooked in some implementations are the possible existence of correlation in the values of the standards and the rationale for using higher-order polynomials to improve the quality of the fit.

The issue of calculating "calibration curves" is not specific to the examples given above, but is general across all of metrology. The prevalence of OLS in this type of application is largely a result of the availability of suitable software (e.g. Excel) for its implementation. If the possible benefits of these other methods are to be realised in practical situations, accurate and usable software is required to implement them. It is therefore an area that merits underpinning support from the SSfM Programme.

3.2.4 Comparison of measurement results with limit values

A strong driver for the measurement of gases is the requirement to determine whether the concentration of ambient gases within the environment, or gases emitted into the environment meet standards defined by the UK Government and the European Union. Tests as to whether a measured value exceeds a stated limit value introduce the requirement for statistically valid use of the measurement data together with information about its statistical distribution. This may be summarised in the form of an estimate of the uncertainty or be derived from repeated measurements of the difference between a particular measurement method and a "reference" method. The test must then be formulated as an appropriate "statistical hypothesis" that can be tested rigorously and unambiguously.

Specific examples include:

- the comparison of an estimate of the uncertainty in measurements made by an instrument with a stated requirement for its accuracy. This approach is being developed in a draft version ISO 14956;
- the test of the suitability of an instrument for an application by comparing a series of parallel measurements with a "reference" method. This method has been prepared by CEN TC 264 (Working Group 9).

These are specific examples of the difficulty of setting up valid statistical tests based on real experimental data. In many cases it is also difficult to interpret the legislative requirement unambiguously in terms of a suitable statistical test.

The issue of comparing measurement results with limit values is not specific to gas analysis, but is general across all of the VAM Programme and most areas of metrology. A useful response from the SSfM Programme to this important requirement would be in the form of a guidance note recommending best practice based on some case studies.

Although this discussion is specific to the comparison of measurement results with limit values, it is an example of a wider issue relating to the application of statistical methods that were not developed for metrological applications to the field of metrology. For example, ISO Technical Committee 69 develops standards relating to statistical methods that may be applicable to the

work of any other ISO Technical Committee. As such, there is a general requirement for the work of TC 69 to be followed by the SSfM Programme in order to ensure that the work is consistent with the requirements and applications of statistical methods within metrology.

It should be noted here that some TC 69 activities are closely followed by LGC and that, for example, the issue of comparison with limits is not always seen as requiring more rigour so much as better measurement capability in the first place!

3.2.5 Levels of confidence other than 95%

There is a general feeling in the industries affected by the VAM programme, that the 95% confidence level is about adequate for most applications in analytical measurement. For some purposes e.g., the presentation of certain types of forensic evidence or the formulation of certain types of pharmaceutical preparations, a higher level of confidence may be required, and in at least one forensic application, a nominal 3σ allowance is made in assessing breach of a statutory limit, corresponding to approximately 99.7% confidence assuming normality. In critical cases, however, it is substantially more useful to obtain independent verification of a value than to express a wider confidence interval. This follows from the typical rate of control failures; expressing a 99% or 99.7% confidence interval based on a normal assumption is futile when typical measurement performance leads to gross outliers in up to 3% of cases.

3.2.6 Supporting software

A large amount of quantitative analysis is performed with the aid of instrumental techniques. Many of these instruments are supplied with dedicated computers - usually PCs - which not only control the instrument but also provide a data processing function, including calculation of the measurement result. There is no known software supplied with such instruments which is capable also of providing an overall uncertainty figure; this is not surprising, given the multi-stage nature of analytical measurements.

By far the commonest type of general-purpose software package used for calculation is the spreadsheet, of which Excel is probably the most widely used example. A spreadsheet can be used both to calculate results and to determine an uncertainty estimate to accompany the result where required. Laboratories also often have statistics packages such as Statistica, Unistat, SAS, S+, etc. which may be used to perform statistical operations on data. In addition, in some areas (e.g., gas and surface analysis) considerable in-house software is used, for example written in FORTRAN or using the MatLab platform.

The packages employed for solving measurement models and determining associated uncertainty generally meet the requirements of the area adequately and are of sufficient quality. There is substantial anecdotal evidence, however, that general purpose business packages (spreadsheets) are not invariably up to the job, and that few users are fully aware of their limitations. User awareness is a general issue: few users fully appreciate the limitations of all software and the possibility of circumventing some of these limitations by appropriate pre-treatment of data.

3.2.7 Supporting documentation

The ISO Guide to the Expression of Uncertainty in Measurement, while being regarded as the base document in the area of uncertainty estimation, is viewed in the chemical analysis field as being too theoretical and remote from the practice of analytical measurement to be of much practical use. It has been interpreted for analytical chemistry by EURACHEM in the form of the document *Quantifying Uncertainty in Analytical Measurement* [2]. This document, although greatly improving the presentation of the topic and providing extensive worked examples, has

itself been criticised for being too detailed for routine application, though many chemical metrology laboratories have welcomed the interpretation. The EURACHEM document is currently nearing publication in substantially revised form, with an expected publication date early in 2000.

On the other hand, in the gas and surface analysis areas there is rather more experience of using the ISO Guide and, within its acknowledged limitations, there is no great problem with it. There is a real need in the gas analysis area for more relevant examples and companion guidance and examples on dealing with non-Gaussian distributions of data.

The UKAS document M 3003 is an interpretation of the ISO Guide for calibration laboratories working in the area of physical metrology. The main body of M 3003 provides a useful alternative presentation of the ISO principles, although the examples are not relevant to laboratories undertaking analytical measurements.

In the VAM area, all three documents fail substantially in one crucial respect. They appear to require a comprehensive model based entirely on measurable input quantities. Though practical and applicable for reference measurements, this is grossly unrealistic for most chemical testing. In the VAM area, most of routine testing is accordingly supported by inter-laboratory studies leading to statistical evaluations of whole method performance under stated conditions. The ISO Guide and its derivative interpretations give no guidance on use of this type of study, nor the data arising therefrom. It follows that the documents do not provide usable guidance for most of the routine chemical testing area.

It is, however, true that the fundamental principles of variance combination and equivalence of statistical and non-statistical uncertainty estimates can be applied to the types of data obtained in analytical method performance evaluation. It is considered possible to provide *'Guide*-compliant' guidance on uncertainty estimation using data from method evaluation studies. This issue is being addressed under the present NMS VAM programme, both through theoretical development and via input to EURACHEM and, more recently, ISO documents.

3.2.8 Training

Personnel in the VAM area are generally competent to develop and optimise methods by experiment, to set up basic calculations, and to apply routine calibration methods. Many are also competent to evaluate overall method performance using traditional statistical tools. Very few, apart from those in the gas analysis area, are familiar with the principles for characterising the error distribution of input quantities or for combining them. Modelling measurement processes is currently a major concern in the surface analysis area; in other areas it is a minority activity only.

Where laboratories have a need to estimate uncertainties (essentially accredited laboratories only at present), it is usual for one or two members of staff to be trained and competent to do so. Current demand for LGC training courses (currently the sole provider specific to the VAM area) is steady, but suggests a trained population limited to the few hundred accredited laboratories. In addition, for a general understanding of uncertainty estimation in accordance with UKAS M3003, there is the UKAS course. This is acceptable for dealing with Gaussian distributions, but a complementary training course is needed to deal with non-Gaussian results of measurement. The latter course, if developed, may initially, within the VAM field, be of limited interest; but it could be anticipated that within a few years there will be greater acceptance of statistical methods that deal with non-Gaussian distributions and, when the benefits of using those methods are seen, there will become a demand for training in those methods.

Analysts receive training in the use of appropriate software packages and systems for solving measurement models when necessary. The training ranges from on-the-job training conducted

by another experienced analyst to attendance on more formal training courses for particular systems.

3.3 Visual Modelling and Data Visualisation

Visualisation is used extensively in surface analysis, in general using commercial packages. There are no particular issues raised by this usage. The way existing software is developing is considered to be quite satisfactory. In-house plotting and imaging software mainly uses the MatLab platform.

In general, visual modelling and data visualisation are not used in the rest of the VAM field since the models encountered are invariably simple and pose few, if any, problems with their interpretation. Traditional graphical presentations of simple data such as regression plots, scatter diagrams, histograms and box and whisker plots seem to provide analysts with most of what they need. It is not clear what benefits would arise from use of more advanced visual modelling techniques.

Note, however, that applications such as surface modelling do appear in post-measurement interpretation. For example, the distribution of contaminants on development sites uses purpose-designed surface modelling software. It has been suggested that a visual indication of the uncertainties in such surfaces would be a valuable extension of such software.

3.4 Data Fusion

In analytical measurement science ‘data fusion’ is best identified with the discipline of *chemometrics*, the application of multivariate statistical methods to chemical data. This is now a well established discipline with a strong academic and industrial research base. Fresh applications of chemometrics techniques are reported weekly in the analytical literature and, in addition, there are regular meetings of special interest groups such as that of the Royal Society of Chemistry (RSC). Instrument manufacturers routinely incorporate chemometric techniques in their data processing software, either as explicit functions called by the user or as hidden routines of which the user may be unaware.

Some of the more common chemometric techniques employed in this area are Principal Components Analysis, Factor Analysis, Linear Discriminant Analysis, Cluster Analysis and, in particular for calibration, Principal Components Regression and Partial Least Squares. Artificial Neural Networks (ANNs) are often employed for classification purposes although, where comparisons have been carried out, it is not clear that the results produced are significantly better than those obtained using the more traditional multivariate statistical techniques. ANNs are also being used for quantitative measurements and an application of an ANN for the quantification of proton NMR spectra has recently been described [3].

The following comments accordingly focus on applications of chemometrics. It should be noted, however, that ‘data fusion’ in principle covers a broader field; there is, though, no evidence at present of applications other than the chemometrics applications described here.

3.4.1 Applications

1. *Processing and reconciling measurements from an array of sensors*

   Chemometric techniques are employed in industrial products for, among other things, just this type of application. An often quoted example is the ‘Warwick Nose’ (developed at Warwick University) which consists of an array of sensors designed to characterise the chemical components of gases and vapours. This system has been commercialised and one of its uses

is to detect the presence of off-flavours in beers and other beverages. The topic of electronic noses is also being considered in the work on odours within the gas analysis part of the VAM programme.

2. *Patching together measurements corresponding to different sets of parameters*

   Combining different sets of parameters in order to generate new information is also a well developed application area. In the field of environmental analysis, for example, unknown crude oils (from spillages, for instance) can be classified using chemometric techniques.

3. *Fault and error detection/diagnosis by multivariate statistical process control*

   Fault and error detection/diagnosis using chemometrics is not a mainstream activity for analysts in this area. An emerging field, however, is multivariate statistical process control, which combines data from many different measurements to provide a rapid indication of process failure.

4. *Combining results from several very different measurement processes*

   The use of chemometric techniques to generate information from proxy analytes has been investigated. The typical motivation is where the target measurement is difficult, and hence expensive, to make but a number of cheaper measurements with which it can be shown to be correlated can be made more easily. For example, the total amount of pollutants entering a river from an industrial site has been estimated from a subset of all the parameters measured, and, as another example, the number of different measurements needed to characterise certain classes of hydrocarbon fuels has been reduced by up to 60% by identifying correlations between specific parameters.

5. *Dealing with predictive data from models, or historical or archive data, in tandem with current measurement data*

   There are no known examples of this type of application in the VAM area.

6. *Handling numerical, image, and/or qualitative data within a single framework*

   Combining numeric and non-numeric data within a single framework is most likely to be handled via expert-systems approaches, though it is possible to use ‘qualitative’ data in chemometrics models.

3.4.2 INTErSECT awareness

Awareness of the INTErSECT Faraday Partnership in this area is beginning to grow. One of the four INTErSECT flagship projects impinges upon this area. It is the “Intelligent data analysis & data fusion techniques in Pharmaceuticals, Bioprocessing and Process Control” project, which involves University College London (UCL), Unilever, SmithKline Beecham, GlaxoWellcome, Sira and Integral Solutions Ltd (ISL). The aim of this project is to show how hybrid intelligent systems techniques recently employed for processing large data sets and data mining in retail and finance business can be developed for, and transferred to, pharmaceuticals, bioprocessing and process control applications. The objectives are to develop and apply techniques derived from previous work at UCL together with recent theoretical developments in neural networks and computational learning to the analysis and management of large data sets in these application areas. Specific problems to be addressed include the following: understanding of structure-activity relationships in drug design (GlaxoWellcome); optimisation of high throughput screening processes (SmithKline Beecham); understanding the relationships between production parameters, sensory assessments, product appearance and consumer preferences (Unilever). The ultimate objective is to produce an automated system that selects the most appropriate hybrid techniques for a particular problem and can use data fusion to combine results in the most

intelligent way. ISL, a developer and supplier of data mining and decision support software (such as Clementine) aims to incorporate the results in future products. There does not appear to be any obvious requirement for additional INTErSECT sponsored projects in the VAM area, perhaps with the exception of the development of novel sensors and related data processing techniques for use in gas analysis.

4. Validation and Testing

4.1 Spreadsheets and Other Mathematical Software Packages

A wide variety of software packages is in use in the VAM area, many of which are supplied by instrument manufacturers to process data from associated instruments. These packages perform a variety of functions ranging from instrument control through data collection to report generation. Users of these packages are not generally aware of the limitations of the mathematical content of the software. When problems arise they are often attributed to 'bugs' in the program rather than to shortcomings in the underlying algorithms. Since users hardly ever have access to commercial computer code, they are in no position to assess the suitability of the algorithms employed.

Where general purpose packages are used, these tend to be either spreadsheets or statistics applications. Excel is by far the commonest Windows spreadsheet in use, but some analysts are still using older DOS based spreadsheets such as SuperCalc. In the gas and surface analysis areas, MatLab is used extensively. Several specialist packages are also used in the gas analysis area. Statistics packages are available; those in use in any one organisation often depend on the general business area concerned. BMDP for instance is orientated towards the medical sciences with SPSS towards the social sciences; STATISTICA has a large user base in the chemical and physical sciences and SAS is heavily used in the pharmaceutical sector. (Note: These are not the only packages used in each area and mention of specific packages is not an indication of any requirement).

When using a mathematical function within a package for the first time users will very often test the routines by feeding in data for which a result has already been obtained manually or with another system. This does not of course constitute a full validation but it does provide some degree of assurance that the function is at least capable of performing correctly on the test data.

It is not common for users of off-the-shelf software to seek assurances from software suppliers that mathematical functions employ algorithms which are valid, accurate, and stable over the likely range of values of the user's input data. It tends to be assumed that the manufacturer of a piece of software will ensure that this is so. If, however, users could be persuaded to press for assurances from suppliers, it might push suppliers to seek objective validation.

In the case of bespoke software, the software developer is generally required to demonstrate that the application performs satisfactorily on test data sets supplied by the customer. The test data sets are selected to cover the range of values likely to be encountered and may even be more severe. Despite such testing, users can often recount catastrophic failures arising from unusual data sets, for example, where the standard deviation of a data set forms the divisor in an expression and happens (correctly) to equate to zero. Such pathological cases are, fortunately, rare but illustrate the difficulty of trying to think of all possibilities.

4.1.1 'Off the shelf software'

The VAM area is highly automated and large amounts of numerical data are processed via proprietary programs supplied by instrument manufacturers. Where off-the-shelf software

packages are employed these tend to be mainly Excel or STATISTICA, with a few single user packages, such as MathCad.

4.1.2 Spreadsheets

Extensive use is made of spreadsheets in this area with computations at several levels of complexity being performed. As well as being used to evaluate algebraic measurement models of the type $y = f(x_1, x_2, \ldots)$ and regression models, they are also used for statistical calculations (means, standard deviations, t and F tests, ANOVA, etc.), processing measurement uncertainty data and experimental design. In-built functions (of Excel) that are used include: average, stdev, sumsq, sum, var, covar, max, min, sqrt, power, log, ln, exp, abs, and, if, ceiling, floor, rand, normdist, ... In addition, use is made of spreadsheets for the graphical presentation of results and as a common format for data exchange.

Most uses involve applying the in-built functions, but some use is also made of these functions within user- or professionally developed 'macros'; in addition, some users write their own data analysis routines. Some users apply features such as Direct Data Exchange (DDE) or Object Linking and Embedding (OLE) to help automate data acquisition, by driving instrument software from the spreadsheet application.

4.1.3 Awareness of limitations

Users tend to become aware of limitations in the software packages used when an obvious inconsistency arises, for example, when the results obtained with one package differ from those obtained with another package for the same input data. Limitations are also indicated when a package produces garbage output; this can be as a result of a 'bug' (coding error) or, not infrequently, through inappropriate use of a function, for example, attempting to fit a polynomial of too high a degree, such as the highest degree available, to a data set.

Within the gas and surface analysis areas, there is an awareness of the potential limitations in spreadsheets and other software packages. In the surface analysis area, specific tests have been developed to aid other analysts to validate software.

4.1.4 Testing and validation

Some testing and validation are generally performed on software packages, but the amount depends on the type of package and its scale of application. At one extreme (typical within gas and surface analysis areas) users will test mean and standard deviation results, say, from a spreadsheet by repeating the calculations using a calculator or a different software application having the same functionality. At the other extreme, in the case of a Laboratory Information Management System (LIMS) and commercial surface analysis data reduction software, extensive and highly structured testing and validation are performed. This testing includes the use of data sets for which results are known analytically.

4.1.5 Supplier advice

For low cost (< £1000) shrink wrapped software, users seldom seek advice from the suppliers about the mathematical performance characteristics of the package before purchase. They usually tend to rely on prior experience with earlier versions of the package or evaluation of demonstration versions coupled with expert reviews in the literature. Of course, should problems arise then the supplier will be contacted for advice. For low cost products advice from suppliers is variable. The general impression is that big name suppliers

in particular are often reluctant to do anything about reported faults other than to thank the customer and say that the information will be passed to the development team. Such customer reports are certainly not always acted upon; LGC has experience of reporting a fault on a beta version update of a well known spreadsheet only to find that the same fault was still present in the public release of the version! Smaller or niche market suppliers are usually more inclined to respond positively to user concerns. In the surface analysis area, one smaller supplier may respond with a disk by return whereas another may ignore any advice.

In the case of software/hardware systems representing large capital investments all aspects of the software will be scrutinised by the user. Where mathematical performance is an issue this will be tested to a level deemed appropriate by the user. There will normally be a contractual obligation on the supplier to co-operate in such testing and the supplier will normally be able to offer much useful advice and information about an established product.

In surface analysis, no problems at all have occurred in the use of software to promulgate the VAM programme, but faults have been observed, as expected, in commercial software tested in the programme.

4.1.6 Traceability

Where software is used in the production of a measurement result from raw data, it forms an integral part of the traceability chain. Demonstrable validity, including the use of appropriate configuration control and access control, is consequently important in maintaining traceability. It is, however, not currently considered necessary to establish the precise nature of algorithms employed; it is generally thought sufficient to verify output performance. This is not currently considered an issue in the gas and surface analysis areas.

Software may also support auditing and record keeping activities. Analytical laboratories operating under a recognised quality system are aware of the need for accurate documentation. Data files from spreadsheets and other mathematical software packages provide a convenient and compact representation of data and can be used to trace the chain of operations from raw data to reported results. The ability to demonstrate such traceability is mandatory in order to compete for certain types of contract, for example, where the customer requires work to be carried out under the Good Laboratory Practice (GLP) protocol [4,5].

4.2 Model Validation

Validation of a measurement model occurs during analytical method development as a matter of course. When methods are put to use they usually involve a calibration stage in which the response of the analytical system, for a set of known standards, is recorded. Analysts strive, and are frequently able, to arrange experimental conditions so that a straight line calibration is obtained. The calibration model is traditionally solved using unweighted linear regression. For some methods, the analytical technique employed leads to non-linear calibrations and here polynomials of degree 2 or occasionally 3 are used.

In the special case of multivariate modelling (including ANNs), several techniques for model validation are employed, particularly different approaches to cross-validation, including ‘leave-one-out’ recalculation, and, increasingly, techniques which subdivide the data set at random into ‘training’ and ‘validation’ sets. Models are built on the ‘training set’ and comparison of predicted and calculated values on the ‘validation’ set gives an effective indication of under- or over-fitting.

Experimental designs usually employed are standard designs which have already been validated such as factorial, central composite, Taguchi, etc. These are provided by a statistician working in collaboration with the user and so they are assumed to be appropriate.

In the area of surface analysis, model validation is an important area for further work. There are few calibration stages, but rather the measurement process is analysed from first principles. It is not possible to compare results with other methods of calculation, since often no other method exists.

4.3 Measurement System Validation

4.3.1 Embedded software

Most modern instrumentation contains embedded software but, since this cannot generally be accessed by the user, such instrumentation tends to be validated as complete systems.

4.3.2 System validation

System validation is treated to varying extents in different sectors. Accredited test laboratories are not generally required to validate measurement apparatus, but are required to check correct operation using appropriate test materials. This may involve as little as running a single test sample through prior to use, though method validation requirements would also have generated sufficient data to validate the precision and accuracy and to show that the instrument was performing adequately in context.

In the pharmaceutical sector, validation is a formal (legal) requirement under Good Manufacturing Practice (GMP) [6], and the pharmaceutical sector has accordingly developed a formal approach to instrument validation, termed ‘equipment qualification’ [7]. {The guidance parallels prior publications on ‘Good Automated Manufacturing Practice’ (GAMP) [8,9]}. This approach requires formal testing against specifications set out at the purchase stage, and will typically verify instrument control parameters and check overall performance using test materials. Where software forms part of the function, appropriate tests are used to verify the software performance; it is considered best practice to examine extreme cases where possible.

In the gas analysis field, each measurement system is validated as a whole.

4.3.3 Reference data

Reference data sets are not widely used to validate measurement systems. The primary reason is probably a low perceived need; system validation via reference samples is generally considered adequate for verification. In the gas analysis area there is some regression testing performed using previously captured data used as a reference data set. In the surface analysis area, reference data are being generated to validate commercial software for data reduction.

4.3.4 Validating embedded software

It is most unusual for instrument users in this area to be given the facility by manufacturers to validate embedded software. In the gas analysis area, control software is tested for extreme conditions. The general lack of validation of embedded software means that there is no detection of any discontinuities in the outputs that might be caused by software errors. In the surface analysis area, the development of standard import and export facilities allows

reference data sets to be used for validating aspects of data reduction software, but not the instrument control software.

4.3.5 Supplier guarantees

Users, as a rule, have to rely on the claims of manufacturers that embedded software has been adequately tested against supplier specifications. Users are nonetheless considered responsible for ensuring that purchased products meet specific needs. In the surface analysis area, the diagnosis system mentioned above has shown software bugs and the use of inadequate algorithms in commercial data reduction software.

Within NPL, software supplied by one part of NPL to another will normally be produced to a software quality standard forming part of the NPL quality system. This ensures that adequate testing is performed before the software is put into use.

4.3.6 Supplier validation of embedded software

Suppliers should validate embedded software as part of the design and implementation process. Current recommendations on software qualification require a 'design qualification' to verify that the design meets requirements; this is necessary but not considered sufficient. Most suppliers use a specification-design-code-test process through implementation; in conjunction with a design qualification, the specification/test pair constitutes validation. In addition, software needs to be tested *in situ* and under normal performance conditions.

In some sectors, particularly pharmaceutical, common guidelines exist. However, different sectors have different requirements, making general guidelines hard to provide. Development and testing in general are accordingly almost certainly best carried out within recognised software development systems, such as TickIT, which allow suppliers to choose their own processes to achieve fitness for purpose, accommodating sectoral guidance as required.

4.3.7 Access

Users should have sufficient access to embedded software to satisfy themselves that it meets their own requirement. There is no case for permitting access to software performing a purely control function for example but, where the software is operating on actual measurement data, such as in a calibration routine, it is important for the user to be able to test this independently of the manufacturer. Manufacturers of analytical instrumentation do not generally provide such a facility.

5. Metrology Software Development Techniques

5.1 Software development methods

Software development can be carried out either in-house or by external contractors. In-house development tends to occur where suitable expertise exists and can be utilised or where projects are relatively small (typically a few man-months).

The development methodologies employed have traditionally ranged from the quick and dirty construction, aimed at getting some working code in order to demonstrate feasibility, to the more formal approaches of a professional software house. The software development process used needs to reflect the risk associated with the software; so, for example, software used in prototyping is often written before the design documentation, whereas software used in the operation of a service needs to be developed using good software engineering practices.

External contractors invariably employ some type of formal development methodology and have quality assurance programs in place. These programs are usually based on the ISO 9001 model and, in the case of larger organisations in particular, very often accreditation to TickIT is held as well.

Most issues in software development are not directly relevant to the work of analysts in this area, although of course they make extensive use of software and depend heavily on it. In order to get a feel for how these issues relate to the commercial software employed in this area, a number of instrument manufacturers were approached and responses elicited from their software engineers.

5.1.1 Software Quality

1. Quality systems

All the instrument supply respondents contacted operate under an ISO 9001 Quality Management System. In addition, many software houses operating in the laboratory area are TickIT certified. NPL is certified to ISO 9001 plus TickIT, so any software NPL sells is produced in accordance with the requirements of TickIT.

2. Reliability records

Software is subjected to Software Quality Audit (SQA) before it is released and, on an ongoing basis, a system of recording and tracking faults is used. Fault tracking databases are reviewed regularly and attention drawn to reliability problems.

3. Reliability

The methods used provide a good measure of reliability, but it is recognised that the reliability can always be improved.

5.1.2 Management issues

1. Maintenance

Older software developed in-house can be expensive to maintain since much is poorly documented and the authors have frequently left the organisation. An exception is software developed in-house for research and prototyping purposes, which does not so much need maintenance as continual refinement as the research progresses; the development and refinement of such software can only be done effectively by the metrologist or by a programmer working very closely with that metrologist. For commercial software developers, however, the position is much better and the large investments made in reliability measures means that maintenance costs are generally minimal.

2. Development costs

In-house software development usually centres around spreadsheet macros or small applications using the Microsoft Office suite. Some small programs (usually add-ons to commercial programs) are developed using associated development tools such as Visual Basic. For these the development costs are seen as reasonable.

Commercial developers feel that, given the level of complexity of their products, the development costs are quite reasonable compared to the return on investment.

3. Problems

For software developed in-house the biggest problem is maintainability in the face of advancing technology. Some DOS based applications have had to be scrapped as they are no longer

compatible (on the hardware side) with modern PCs. With reductions in staffing there is also no time for staff with programming skills to learn a new language in order to recode such applications for Windows.

Commercial developers complain of projects not being sufficiently well defined in time for project start-up.

Users find configuration control difficult on multi-purpose systems; the prevalence of Windows systems is useful, but it is widely recognised that installation of apparently unrelated software can impact on the measurement software due to changes in versions of shared executables.

4. Quality attributes

The most important software quality attribute in metrology is probably reliability. Commercial developers cite usability first. Usability also impacts on reliability; well designed software interfaces reduce human error and alert the user to other problems. Conversely, ease of reconfiguration may lead to practical difficulties in configuration control.

5. Risks

Software error invokes a range of risks. Software failure in use results in down time, increased costs and potential for lost business for the laboratory. Data integrity failures may compromise large quantities of results, with considerable business implications. Access control failures may compromise confidentiality, leading to commercial losses for the laboratory or customer. The risk of software errors causing incorrect results to be reported is also serious; analytical measurements support both business and health decisions, and this in turn could have serious financial or safety repercussions for the users (i.e., the customers) of such results. In one case (notified privately) a pharmaceutical company lost several months' production revenue because a spreadsheet was giving incorrect results, delaying product approval. (Pharmaceutical production revenues may easily reach £1m a week.)

Commercial developers are also concerned about the loss of revenue that would arise as a result of the late introduction of a piece of software.

On the other hand, the risks associated with in-house software developed for research and prototyping purposes are quite low.

5.1.3 Technical issues for software developers

1. Development methods

Developers report using both the V model approach and prototype refining. Rapid prototyping followed by formal specification and re-engineering is also used by one major LIMS developer.

2. Software testing methods

The correctness of software is checked by one or more of the following methods:

a) checking the results by using another instrument;

b) checking the results by hand;

c) using a mathematical tool or spreadsheet, etc.

In addition, the user interface and functionality are subjected to test plans executed independently of the development team by SQA engineers. Algorithm checking is performed in the course of software validation which is carried out against published standards.

3. Independent testing

Instrument manufacturers report that parts of machine software can be tested independently of the hardware. It is highly probable though that a normal user would not be able to replicate such testing to the same degree.

4. Development problems

For instrument manufacturers one of the biggest problems is created by the very large number of different PCs, operating systems, and locales, which makes testing all combinations virtually impossible.

5. NPL Measurement Good Practice Guide 5, Software in Scientific Instruments

LGC has a copy of NPL Measurement Good Practice Guide 5, for information only. Guide 5 is widely available within NPL and has been extensively distributed outside through Competing Precisely events and through the SSfM Club. Feedback on its use has been received, but not yet from those active in the VAM field.

5.2 Software reuse

5.2.1 Management issues

1. Importance

Software reuse is seen as an important means of reducing software maintenance costs. Within the gas analysis area, re-engineering of software, rather than reuse, is the emphasis; existing software is re-engineered for use in a similar context or to solve a similar problem.

2. Mixed language programming

Mixed language programming tends not to be used for in-house projects in LGC but is employed in NPL and by commercial developers where, typically, the user interface will be programmed in Visual Basic and C/C++, FORTRAN or MatLab will be used for lower level functions.

3. Responsibility

The details of software reuse and mixed language programming tend to be left to the development teams, although general guidelines and policy may be provided by management.

4. Legacy software

Legacy software is generally not supported – in part because support is no longer available from the suppliers, but also because of 'technology pull' and a management desire to migrate all applications to modern systems. As noted elsewhere, some legacy software suffers from hardware incompatibilities and will not operate correctly on modern systems. In some cases legacy software continues to be operated on a 'no support' basis, while in other cases it is abandoned altogether when licences are not renewed.

In the gas analysis area, most of the legacy software is instrument control software, e.g., written in Quick Basic, which is used to control old instruments. There is very little need for maintenance of such software.

5. Year 2000

The Year 2000 problem has been recognised for some time. LGC and NPL have task groups working on this issue, testing every piece of appropriate equipment. It is not yet known what impact Y2K will have on the delivery of NMS milestones since, in part, it depends on how the relevant suppliers and subcontractors will be affected. Informal contact with suppliers indicates

that the issue has generated substantial administrative costs, simply to respond to the volume of requests for information.

5.2.2 Technical issues for software developers

1. Producing similar programs

There appears to be a preference for reusing existing compiler subroutines when producing a program very similar to an existing one. It is generally too expensive to recode from scratch and little value is seen in copying programs. The old program on which the new one is based is updated as necessary to minimise the software maintenance costs.

2. Software reuse strategies

This is not an issue which generally affects analysts and commercial developers have established software reuse strategies with which they are content.

3. Extent of mixed language programming

No particular problems are reported with mixed language programming. Developers, recognising the potential for problems, have design protocols for mixed language interfaces and these are adhered to. In addition regular reviews of performance issues around mixed language interfaces are held.

4. Year 2000 problems

As far as can be ascertained, most major companies have made a substantial investment in identifying and solving Y2K problems.

5.3 Virtual Instruments

Apart from their use in gas analysis work for controlling new instruments, virtual instruments are not used much in the VAM area. Some instrument manufacturers employ them for developing and demonstrating instrumentation where it is impractical to have a complete instrument present.

LabVIEW is not widely used by instrument manufacturers for developing virtual instruments as proprietary development tools are available for this purpose. On the other hand, the use of LabVIEW at NPL is increasing; within the VAM programme, LabVIEW is being used for most new instrument control applications in the gas analysis area.

Where virtual instruments are in use, no particular problems were reported with design documentation or testing.

6. Support for Measurement and Calibration Processes

6.1 Automation of Measurement and Calibration Processes

There is no experience of using Laboratory Management Information Systems (LIMSs) in the gas and surface analysis areas. In the gas analysis area, databases of historical data are used to aid in running the measurement and calibration services, but the intention is to bring in a simple LIMS quite soon. Currently, measurement and calibration services are only just starting in the surface analysis area. Hence, the rest of this section will report on the chemical analysis area in which there is much more experience of using LIMSs.

A LIMS is in essence a database and associated management system designed to store and process laboratory data. This apparently simple definition belies the power and potential of a LIMS which can bring many benefits to a laboratory. As with any major IT system, however, it

can prove to be an expensive failure if not planned carefully or if it does not receive the full commitment and support of senior management.

Some of the features a LIMS can provide are:

- Acceptance of sample data entered manually at a terminal;
- Acquisition of data automatically from laboratory instrumentation;
- Storage of data;
- Processing of data (performing calculations on it);
- Controlling access to data;
- Report generation;
- Assistance with work scheduling;
- Provision of information on the status of an analysis (not started, in progress, complete, etc.);
- Tracking of samples through different locations within a laboratory;
- Maintenance of instrument service and calibration records;
- Utilisation of information from, and in return supporting, a corporate Management Information System (MIS).

The above list is not intended to be exhaustive but reflects the main uses to which LIMSs are put.

LIMSs started appearing in the mid 1980s. These early systems were mostly based on mini-computers but their appearance at that time was due to the increasing cheapness and availability of PCs which fuelled a growth in laboratory automation and a resulting data explosion. The main idea behind the concept of a LIMS was the desire to integrate the data storage and processing functions being performed, in a typical laboratory, by many disparate systems. For example, the process of checking in samples was often accomplished with manual registers or card boxes; sample reporting was starting to be done using word processors and price information may have been generated manually or via a separate computerised finance system. By bringing such functions together under one computerised information management system tailored to the needs of the laboratory, the expectation was (and still is) that information flows would be speeded up and operational efficiency thereby greatly improved.

Over the years commercial LIMSs have steadily evolved and today a wide range is available to suit almost every need. Networked PC based systems are popular for the small laboratory whilst, for larger systems, there is a move to more fully integrate the traditional LIMS functions with corporate computing functions to produce an enterprise-wide ‘solution’ to a company’s data management needs.

6.1.1 LIMSs in use

The main laboratory-wide LIMS in use at LGC is a Beckman CALS (current version 8.4) client server. There are, in addition, two smaller systems, a William Woodard dedicated to urine testing work and one developed in-house using dBase for tobacco work.

6.1.2 Other processing methods

Most of the work carried out in the VAM area is concerned with test data, although some calibration data necessarily accompanies this. Data may be processed in a number of ways other than through a LIMS, the commonest of which are:

1. Using a hand calculator;
2. Using a spreadsheet (possibly with the aid of user written macros);
   a) with manual data entry;
   b) with automatic data entry;
3. Using a proprietary data processing computer program;
   a) with manual data entry;
   b) with automatic data entry.

Methods 2b and 3b are often found in association with instrumentation supplied to implement a specific analytical technique. A variant of 1 involves the use of software calculators such as those provided by Windows or some Operating Systems such as Novell Netware.

6.1.3 Duration of use of LIMS/other systems

The current Beckman system has been in use for about three years. This system replaced an older Hewlett-Packard LIMS which had been in use for about six years.

6.1.4 Advantages and disadvantages

In addition to the general advantages provided by the features outlined above there are other specific benefits. For example, there is a real benefit in areas like DNA testing where there are large numbers of samples and it is necessary to keep track of the chain of custody. In the case of the DNA work, the system has been designed to make sure that the steps of the analysis are carried out in the correct order. It is necessary to ensure, e.g., that a sample is of the correct type, in the correct location and is available for testing. There is a lot of automation to reduce transcription errors, and triggered events to check the validity of the data entered. The database employed is the widely used Oracle system and so there is good potential for linking to other compatible systems.

The main disadvantage is the lack of speed of the system and in some cases data entry cannot be made flexible enough without making it too complex for the user. The system can accept input via terminals or instrument interfaces, but all input is handled through a single event queue and this can lead to bottlenecks occurring. Other disadvantages include the high cost of upgrades and the fact that the software is tightly coupled to Oracle versions. Though effective, the security features provided could also be improved. The system permits 16 levels of access, but these are not really adaptable enough for LGC's purposes.

6.1.5 Finding a suitable LIMS

There are many ways of obtaining information on LIMSs and their potential suitability for a particular application. The most fruitful would be:

- Consult the literature - *e.g.* journals such as *Scientific Computing and Automation*, *LIMS/Letter*; trade journals such as *Chemistry in Britain* are also useful sources of information, as are advertising publications like *Labmate* and *International Labmate*.

- Approach other organisations known to be using a LIMS and seek their opinions/comments. This option is best reserved for when a contact is already available in an appropriate organisation and the organisation is not a direct competitor.
- Attend seminars and conferences such as *Pittcon* or the *International LIMS Conference*; the latter is held annually and alternates between different locations in Europe and the USA.
- Search the WWW; all LIMS manufacturers have a presence on the Web and much useful information is available (see for example the *LIMS/Letter* site at http:/www.LIMSource.com - contains links to over 80 LIMS vendors).
- Employ a consultant. This will almost certainly be a necessity in order to confirm the suitability of a potential system, particularly if an organisation has no previous experience of commissioning and installing a LIMS.

6.1.6 In-house development

LGC's main LIMS is a commercial product but was partially tailored in-house using SSADM for template development. Given the ready availability of commercial systems, few organisations would now contemplate building their own one in-house. The small LIMS used for LGC's tobaccos work has been developed over many years to meet a particular need not catered for at the time by commercial systems. It is built around dBase, but no particular methodology was used in its development.

6.1.7 Adherence to appropriate standards and procedures

This question really concerns the issue of validation and this is something that needs to be given careful thought when the specifications requirements document is being drawn up. Commercial suppliers of LIMSs are well aware of the demands made by regulatory agencies such as the US Food and Drugs Administration (FDA) and the UK Department of Health (DOH). These demands generally involve demonstrating compliance with an international protocol such as GLP, GMP, or GALP [10], GAMP.

The software for the Beckman system was designed under an ISO-9001 regimen. Before accepting the system, LGC ran a large number of checks to ensure that the software manipulated data in accordance with its design, i.e., the correct tables were updated, calculations were consistent and reports pulled data from the correct fields. Every screen, entry and customised code module was exercised during testing.

LIMS suppliers are often able to provide a basic validation service but the onus is still on the user to ensure that a validation meets their full requirements. Issues such as the provision of an adequate audit trail facility and security arrangements appropriate to their needs will have to be addressed.

6.2 Format Standards for Measurement Data

6.2.1 Need for standard data formats for measurement data

Standard data formats for measurement data are only needed in a few specific areas within the VAM field.

JCAMP-DX is a human-readable data transfer format developed and frequently used in interchange of mid-infra-red and other chemical spectroscopic and chromatographic data.

In the surface analysis area, a standard format has been devised for storing and exporting data from the manufacturers' instruments. This standard format, developed by NPL, has now been standardized as ISO 14976 [11].

6.2.2 Need for sharing measurement data with other organisations

There is generally no need in the VAM field to share extensive amounts of raw measurement data with other organisations, except in the area of surface analysis. For some purposes, such as in Proficiency Testing schemes and for inter-company collaborations, there is a requirement to share processed data or small amounts of raw data. Where this is the case, data are usually stored and transmitted in the form of ASCII files or as spreadsheet files.

6.2.3 Need for a common format for calibration records

There is no need for a common format for calibration records specifically to enable international comparison of calibrations of a single instrument.

6.2.4 Need for a common format for exchange of measurement data

There is a perceived need in some quarters for a common format for electronic exchange of measurement data. Much work has been done, particularly in the USA, on the development of a common exchange format for chromatographic data. The Analytical Instruments Association (AIA), a US trade association, are promoting the Analytical Data Interchange (ANDI) protocols to "...facilitate the integration and use of data from multiple vendor's instruments". At present protocols exist for chromatographic (issued 1992) and mass spectral (issued 1994) data. These protocols implement, for their respective types of data, the network Common Data Format (netCDF) devised by the Unidata Corporation in the late eighties and further developed since then. These protocols, which were specifically designed for storing large amounts of array data., are beginning to appear in the latest generation of chromatography and mass spectral data stations although they are not as yet in common use.

In the surface analysis area, the common format [11] for exporting and importing measurement data has facilitated software development by third party manufacturers.

6.2.5 Different data formats

In most areas it is felt that the different proprietary data formats found in different measuring instruments is not a factor which would inhibit their use. A LIMS would circumvent this apparent problem through the use of data translators provided by the LIMS supplier. In addition, many instruments offer users a choice of alternative formats such as ASCII or CSV specifically to enhance portability of data. However, the standardisation of data formats for surface analysis has found support among most manufacturers and has shown many benefits of having software which works with data from many different instruments.

7. Suggestions for Future Activities

7.1 Expected Benefits of the SSfM Programme

If the current SSfM programme leads to improved testability of commercial software supplied with laboratory instrumentation – that is, if the user can inject his/her own data sets to test the

underlying algorithms – then this will be a significant benefit. This has already had significant results in surface analysis [12].

While many analysts would wish to check the reasoning and arithmetic behind results produced manually by others they tend to believe that results produced by a computer program are correct. The famous old adage "garbage in, garbage out" is widely appreciated but the possibility of "quality in, garbage out" is not. If the current SSfM programme leads to a wider appreciation among analysts of the limitations of the algorithms employed in commercial data processing software then this too will be of benefit.

Particularly in the gas analysis area, benefits are looked for in terms of more guidance on uncertainty estimation, particularly covering non-Gaussian distributions, correlations, using multivariate error analysis (generalised distance regression - i.e., handling errors in all variables). It is also hoped that SSfM will provide support to software development by way of guidance and tools to help the development process.

7.2 Case Studies and Feasibility Studies

Three possible areas for case study have been identified:

- the transfer of LIMS experience from the chemical analysis area to the gas analysis area and possibly other metrology fields;
- the transfer of format standards experience from the surface analysis area to other metrology fields.

There is also the possibility of a feasibility study for applying data fusion to electronic noses in the odour work within the gas analysis area.

7.3 Future SSfM Topics

It is becoming increasingly common for professional bodies to issue software with guidance documents but there is concern about how such software should be validated. For example, the RSC statistics subcommittee has recently issued a set of MINITAB macros for kernel density estimation but it is not clear who should validate them and how this should be done. A future SSfM project to address this need could be the production of a generic guidance document on the validation of software issued by professional institutions such as the RSC, IoP, BSI, ISO etc.

In the next SSfM programme, the work currently being done on uncertainties and statistical modelling could be extended to more fully research statistical methods used in fields outside metrology to determine the extent of their applicability to metrology. Furthermore, contributions could be made to the work of ISO TC69 and the work of that committee should continue to be tracked to identify metrology-relevant work items and, where relevant, to ensure that requirements and guidance used in metrology are consistent with the requirements and guidance issued by TC69.

Support should be provided for generalised least squares methods and for weighted ordinary least squares methods. This support should include ensuring that accurate and usable software is available to metrologists to enable them to apply these methods.

7.4 Future VAM Programme Topics

Recent work within the VAM programme area shows a need for better understanding and treatment of effects arising from the sample matrix. A sample matrix is the material in which an analyte of interest is embedded and may be a solid, liquid or gas. For example the Mercury (II)

ion experiences a different environment in fish meat than it does in sea water, and the difference in environment causes systematic effects on the measurement result. These effects often represent the major source of uncertainty in chemical measurement.

One possible solution to this problem is to characterise the sample matrix more fully (either in terms of the observed behaviour of the analyte or of the matrix composition), model the effects mathematically and hence quantify them. The result may be partial compensation or improved uncertainty estimates. This, together with increasing efforts to quantify effects in sample pre-treatment and other parts of the process, suggest a general likelihood of increased modelling activity within the VAM area, with consequent increasing dependence on effective mathematical treatment.

8. Summary of changes

In this **Interim Report** the main changes from the initial restricted status report on the VAM area are as follows:

- addressing the whole scope of the physical analytical measurements part of the VAM programme, including particulates, aerosols, and electrical methods of measuring pH;
- within the section on uncertainties and statistical modelling, adding subsections on development of calibration curves and comparison of measurement results with limit values;
- updating the text regarding INTErSECT awareness;
- updating the text on the need for a common format for exchange of measurement data;
- updating the suggestions for possible future SSfM activities to include a number of specific suggestions for additional work on uncertainties and statistical modelling;
- identifying, in the VAM chemical area, a nascent trend toward increased modelling (from the current very low level) in reference chemical measurements with potential future increase in software dependence.

9. Acknowledgements

The authors gratefully acknowledge the contributions of Maurice Cox, Peter Harris, Martin Milton, David Nettleton, Martin Seah and Peter Woods – all from NPL.

Appendix 1: Glossary of Abbreviations

Abbreviation	Expansion
A/D	Analogue to Digital
ANDI	Analytical Data Interchange
ANN	Artificial Neural Network
ANOVA	Analysis Of Variance
ASCII	American Standard Character for Information Interchange
BSI	British Standards Institute
GALP	Good Automated Laboratory Practices
GAMP	Good Automated Manufacturing Practice
GLP	Good Laboratory Practice
GMP	Good Manufacturing Practice
DDE	Direct Data Exchange
DNA	Deoxyribose Nucleic Acid
DOH	Department of Health
DOS	Disk Operating System
FDA	Federal Drugs Administration
GLS	Generalised least squares
GUM	ISO Guide to the expression of uncertainty in measurement
IoP	Institute of Physics
ISL	Integral Solutions Ltd.
IR	Infra-Red
ISO	International Organization for Standardization
JCAMP	JCAMP-DX is a human-readable data transfer format developed and frequently used in interchange of mid-IR and other chemical spectroscopic and chromatographic data.
LIMS	Laboratory Information Management System
MIS	Management Information System
NMR	Nuclear Magnetic Resonance
NMS	National Measurement System
OLE	Object Linking and Embedding
OLS	Ordinary least squares
PC	Personal Computer
RSC	Royal Society of Chemistry

SQA	Software Quality Audit
SSADM	Structured Software Analysis and Design Methodology
SSfM	Software Support For Metrology
UCL	University College London
UKAS	United Kingdom Accreditation Service
VAM	Valid Analytical Measurement
Y2K	Year 2000

Appendix 2: Bibliography

1. *Guide to the Expression of Uncertainty in Measurement,* International Organisation for Standardisation, 1993, ISBN 92-67-10188-9.
2. *Quantifying Uncertainty in Analytical Measurement,* EURACHEM, 1995, ISBN 0-948926-08-2.
3. Amendolia, S.R; Doppiu, A; Ganadu, M.L; Lubinu, G. *Anal. Chem.*, 1998, **70**, 1249–1254.
4. *OECD Series on Principles of Good Laboratory Practice and Compliance Monitoring, Number 1: OECD Principles of Good Laboratory Practice (as revised in 1997),* Environmental Directorate, OECD, Paris, 1998.
5. *The Good Laboratory Practice Regulations 1997*, Statutory Instrument 1997:654, HMSO, ISBN 0-11-065105-1.
6. *Rules and Guidance for Pharmaceutical Manufacturers and Distributors*, ISBN 0-11-321995-4.
7. P Bedson, M Sargent. *Accred. Qual. Assur.* 1996, **1,** 265-274.
8. *Good Automated Manufacturing Practice, GAMP3: For Validation of Automated Systems in Pharmaceutical Manufacture*. Available from: ISPE, 3816 W. Linebaugh Ave., Suite 412, Tampa, FL 33624, USA.
9. *Good Automated Manufacturing Practice, GAMP96: Supplier Guide for Validation of Automated Systems in Pharmaceutical Manufacture*. Available from: ISPE, 3816 W. Linebaugh Ave., Suite 412, Tampa, FL 33624, USA.
10. *EPA Directive 2185, Good Automated Laboratory Practices: Principles and Guidance to Regulations for Ensuring Data Integrity in Automated Laboratory Operations*, 1995. Available from: United States Environmental Protection Agency, Office of Information Resources Management, Research Triangle Park, NC 27711, USA.
11. Surface Chemical Analysis - Data transfer Format, BSI/ISO 14976: 1998.
12. M.P. Seah and M. Brown, J. Elec. Spectrosc., 1998, **95**, 71-93.
13. D. Rayner, *Initial Report on the Status of Software and Mathematics in each Metrology Area,* NPL Report CISE 18/99, February 1999.

Appendix 3: Modelling Techniques - Tables

Table 1. Complexity of models used in the VAM programme

Simple	Score*	Complicated
Simple relationship between input and output.	1	Pages of equations to describe the system.
Small number of parameters.	1	Large number of parameters, with complicated inter-relationships.
Measurement errors limited to one measured quantity.	3	Many variables subject to significant measurement error.
Measurement errors consistent with Gaussian distribution.	4	Non-Gaussian behaviour, outliers likely under a Gaussian assumption.
Input quantities well controlled, understood.	2	Inputs estimated, not all influence factors have been identified.

** 1=First statement applies strongly; 5=second statement applies strongly*

Table 2. Maturity of models used in the VAM programme

Mature or Accepted	Score*	Developing or Speculative
Well understood and accepted physical theory.	2	A number of competing models or model incomplete.
Comprehensive theory adequately taking into account all main features.	3	Approximations employed, terms omitted to simplify model. Some components modelled by empirical functions, e.g., polynomials, splines. Behaviour of some components not understood and/or ignored.

** 1=First statement applies strongly; 5=second statement applies strongly*

Table 3. Incorporation of uncertainties in the VAM programme

Thorough treatment	Score*	Simplistic treatment
Measurement uncertainty modelled for all input quantities.	2	Measurement uncertainty modelled for only one or two input quantities. Other quantities either assumed to be exact or uncertainty contributions are assigned arbitrarily.
Variances and correlation amongst input quantities fully taken into account.	4	Possible variances and correlation among input quantities ignored.

** 1=First statement applies strongly; 5=second statement applies strongly*

The score in the first row above does not apply to physical analysis, where a score of 3 would be more appropriate.

Table 4. Comprehensiveness of solution methods for models used in the VAM programme

Comprehensive	Score*	Approximate
Solution method solves the computational problem posed by the model.	2	Approximations are employed. Some terms simplified/linearised/ignored.
Solution method takes into account all the sources of measurement error, variances and correlation amongst the input quantities.	3 or 4	Easiest method of finding a solution employed. All data treated as exact, independent and uncorrelated.

** 1=First statement applies strongly; 5=second statement applies strongly*

Table 5. Effectiveness of models in the VAM programme

Effective	Score*	Ineffective
Solutions consistent with those predicted by theory, established by other methods.	1	Solutions at variance with other results.
Method always returns a solution.	2	Method produces unrealistic solutions, fails to provide a solution, crashes.
Rerunning the method on perturbed data/ input quantities/ produces a nearby solution.	2	Small changes in the input data/quantities produces large changes in the output.
Iterative method converges quickly.	2	Method can make slow progress, fail to converge.
Estimate of statistical uncertainties consistent with the expected measurement error, noise in the data.	1	Unrealistic uncertainties calculated.

** 1=First statement applies strongly; 5=second statement applies strongly*

Note: In the surface analysis area, many of the models are still under development, and hence the above scores and those in Table 6 do not necessarily apply.

Table 6. Adequacy of model validation in the VAM programme

Good validation	Score*	Poor validation
Fit of model to the measurement data consistent with the expected measurement error.	1	Residual errors much larger than expected, have a systematic pattern.
Results agree with other methods of calculation.	1	
Results agree in situations for which an analytic/reference solution exists.	1	Solutions at variance with other results.
Numerical results in accordance with predicted behaviour of the algorithm (rate of convergence).	1	Behaviour at variance with prediction.

** 1=First statement applies strongly; 5=second statement applies strongly*

Note: See note to Table 5 *re* surface analysis.