Preprints.org

Modelling Metrological Traceability

[Blair D. Hall](#) *

Article

MODELLING METROLOGICAL TRACEABILITY

Blair D. Hall

Measurement Standards Laboratory of New Zealand, Lower Hutt, New Zealand; blair.hall@measurement.govt.nz

Abstract: Metrological traceability is essential for ensuring the accuracy of measurement results and enabling comparison of results to support critical decision-making in society. This paper explores a structured approach to modelling traceability chains, focusing on the role of residual measurement errors and their impact on measurement accuracy. This work emphasises a scientific description of these errors as physical quantities. By adopting a simple, static modelling framework grounded in physical principles, the paper offers a formal way to track the effects of errors through an entire traceability chain, from primary reference standards to end users. Real-world examples from microwave and optical metrology highlight the effectiveness of this rigorous modelling approach. Additionally, to further advance digital systems development in metrology, the paper advocates a formal semantic structure for modelling, based on principles of Model-Driven Architecture. This architectural approach will enhance the clarity of metrological practices and support ongoing efforts toward the digital transformation of international metrology infrastructure.

Keywords: Metrological traceability; measurement error; measurement uncertainty; calibration; digital transformation

1. Introduction

Traceability is a fundamental concept in metrology. Quantities must be expressed on a metrologically traceable measurement scale to enable meaningful comparison. For this reason, traceability is critical to the functioning of international quality infrastructures (QIs) and to the dissemination of measurement units within the global measurement system [1]. Also for this reason, traceability is often mandated when critical decisions depend on physical quantities.

A pair of international arrangements that rely on traceability have been established. These arrangements have fostered the development of national and international quality infrastructure, which provides access to necessary traceable measurement services in society. In 1999, the International Committee for Weights and Measures (CIPM) Mutual Recognition Arrangement (MRA) provided a framework for National Metrology Institutes (NMIs) to demonstrate the equivalence of their measurement capabilities to those of international peers [2]. Similarly, in 2001, the International Laboratory Accreditation Cooperation (ILAC) multi-lateral MRA provided a framework to support the provision of calibration and testing services across the developed world [3].

The concept of traceability emerged late in the 20th century [4]. Four possible definitions for *traceability* were considered in an early paper by Belanger [5]. In 1982, Nicholas and White, in their monograph *Traceable Temperatures*, proposed a variation on one of Belanger's suggestions [6]. Then, in 1984, the first edition of the *Vocabulary of Metrology* (VIM) included an entry for traceability, drawing on another of Belanger's suggestions ([7], 6.12). Later editions of the VIM have expanded and refined the definition of traceability. The second edition, in 1993, emphasised that traceability is a quantitative characteristic of measurement by including reference to *uncertainty* ([8] 6.10). This revision coincided with the release of the first edition of the *Guide to the Expression of Uncertainty in Measurement* (GUM) [9]. The current definition of traceability in the third edition of the VIM is ([10] 2.41):

> **metrological traceability**
>
> property of a measurement result whereby the result can be related to a reference through a

documented unbroken chain of calibrations, each contributing to the measurement uncertainty.

This incorporates terms with specific meanings—*measurement result*, *calibration*, and *measurement uncertainty*—which are also defined in the VIM. The definition is accompanied by eight explanatory notes, one of which elaborates on the meaning of *reference*.

It is notable that the definition of traceability includes a requirement for documentation. Metrology is a scientific discipline, and traceability is one of its fundamental principles. It is unusual to see such a practical requirement in a scientific definition. However, traceability plays the functional role in society of ensuring the reliability of measurements. The need for documentation stems from the real-world requirement to audit the reliability of stages along a traceability chain. Thus, the VIM definition integrates both the scientific concept of traceability and the practical necessity of generating records that can be reviewed.

The international metrology community has recently embarked on a digital transformation of its systems and processes, which may entail substantial change [11]. This initiative is motivated by a belief that significant gains in efficiency and reliability can be achieved. However, designing a digital infrastructure which simply automates the various tasks that people perform today is unlikely to foster innovation. Metrology is practised by skilled professionals throughout the world's quality infrastructures, whose work is shaped by their interpretations of authoritative documents like the VIM and GUM. The implementation details can vary—between organisations, between economies, and between regional metrology organisations. This diversity will complicate an overarching digitalisation of quality infrastructure activities, as neither the analysts eliciting business requirements nor the metrologists describing their work have explicit guidance about the balance between practical considerations and scientific requirements. For example, the definition of traceability does not explain the purpose for measurement uncertainty or why documentation is needed [12]. These important aspects are left to the discretion of metrologists.

This article examines traceability as a fundamental scientific principle and advocates for explicit measurement modelling, which provides direct and valuable insights into how various types of measurement achieve traceability. Starting with a few basic assumptions, we show that measurements can be modelled with mathematical expressions. These models can represent the influence quantities that ultimately determine the accuracy of measurement results at the end of a traceability chain. Traceability is established by accounting for these influences; indeed, the elements that need to be "traced" along a metrological traceability chain are the residual measurement errors arising from uncontrolled influences. Mathematical notation provides a concise, consistent, and logical structure, offering several advantages: it transcends linguistic differences, is widely understood, and facilitates rigorous analysis.

The structure of this article is as follows. The next section introduces some foundational assumptions and a notation for measurement modelling. Section 3 illustrates modelling in a variety of scenarios, including: ratio and difference measurements, international comparisons, intrinsic and quantum-based standards, and sensor networks. Section 4 examines the possibility of evaluating models and establishes a connection between modelling and the calculation of measurement uncertainty. Section 5 examines a way of structuring semantic model information to better align scientific concepts with digital systems development. This approach will facilitate the digitalisation of metrological processes and enhance interoperability. Section 6 discusses our argument that residual measurement error is central to a scientific understanding of traceability. It highlights examples from microwave metrology and optical goniometry where modelling, supported by software, has improved measurement accuracy and information flow along traceability chains. Our conclusions are summarised in Section 7, which is followed by three appendices containing supplementary material. Appendix A develops a model for calibrating a simple linear measuring system, Appendix B summarises a method for evaluating measurement uncertainty as described in the original GUM, and Appendix C outlines a method for evaluating uncertainty provided in supplements to the GUM.

2. Modelling

2.1. Basic Assumptions and Notation

A few simple assumptions provide a foundation for modelling. The first is that the quantity intended to be measured can, for all practical purposes, be uniquely defined. This quantity is called the *measurand* ([13] 3.1.3). A second assumption is that the measurand cannot be determined exactly; the value obtained by measurement is an approximation—an estimate—of the measurand. Consequently, since the exact value of the measurand is unknowable, the deviation of a measured value from the measurand cannot be determined either. Nevertheless, the best estimate of this deviation is usually zero.[1]

These assumptions suggest three entities for modelling measurements. Our notation will distinguish between them:[2]

1. a known value—such as the indication of a measuring system—will be denoted by a lower case italicised term (e.g., y);
2. an unknown value—such as a measurand—will be represented by an upper case italicised term (e.g., Y); and
3. an unknown residual value that has an estimate of zero—such as the difference between a measured value and the measurand—will be denoted by an upper case italic E and a suitable subscript to label the term (e.g., E_y).

In this notation, the second assumption above—that a measurand cannot be determined exactly—may be expressed simply as

$$Y = y - E_y \, , \tag{1}$$

where Y represents the measurand, y is its estimate, and the difference between them is the residual error term E_y. The subscript of E_y corresponds to the estimate of the measurand, because a different estimate would yield a different residual value.

2.2. Static Models

A *measurement model* is an equation that expresses a measurand in terms of the other quantities involved in a measurement. Equation (1) is the most general form of a measurement model, comprising only two quantities: the (known) measured value and the (unknown) difference between that value and the measurand. The equation expresses the fact that the measurand's value could be determined if the residual measurement error were known. It shows that E_y determines the accuracy of the measured value y as an approximation of the measurand Y. In practice, information about the likely magnitude of E_y will be attributed to the *measurement uncertainty* associated with the measured value, y.

It is important to note that this modelling provides a static view of what occurred during a measurement. Each term represents a definite quantity, although some values will not be known. This is analogous to a coin toss where the outcome remains covered by a hand: the result is definite but unknown to observers. A static representation does not inherently distinguish between systematic and random measurement errors, so careful consideration is sometimes required to avoid ambiguity when developing models and assigning identifiers to model terms. For example, if a measurement is affected by a random error, the corresponding term may include an instance identifier. The residual random error in the ith measurement could be denoted $E_{\text{rnd}\cdot i}$, where i serves to distinguish individual instances as needed.

---

[1] It is generally assumed that any known effect causing bias in a measurement will be corrected, ensuring that the measured value remains unbiased.

[2] If conventional notation in a technical discipline cannot be easily reconciled with the one proposed here, unknown-value terms may be underscored (e.g., $\underline{y}$) rather than capitalised.

2.3. Model Building

Models provide a quantitative framework to describe measurement results. Their construction is guided by the intended purpose. The level of detail required in a model depends on the information desired about a measurement. Since no measurand can be determined exactly, equation (1) serves as a generic model applicable to any measurement. However, it provides no insight into the quantities influencing the measurement result. An essential part of metrologists' work is to identify the influence quantities that affect a measurement and incorporate these factors into a model, which can be analysed to evaluate the measurement uncertainty. Such models usually have many terms. However, between the extremes of a completely generic model and one describing a specific measurement in detail, models with just a few terms that summarise a measurement result can also be employed. For instance, it may be convenient to represent the combined effect of numerous random influence factors using a single term, while modelling systematic errors individually [14]. For example, when two systematic effects, A and B, influence each measurement of Y, a suitable model could be

$$Y = y_i - E_{\text{sys·A}} - E_{\text{sys·B}} - E_{\text{rnd·}i} , \qquad (2)$$

where $E_{\text{rnd·}i}$ represents the combined effect of random errors in the ith measurement of Y.

2.4. Traceability

The accuracy of information obtained by measurement at the end of traceability chains is critical. It must be possible to assess whether the measurement accuracy available is fit for purpose in a given context. Along a chain, the accuracy of measuring systems must be accounted for to assess the likely magnitude of residual measurement error at each stage. That is why national metrology institutes invest significant resources in developing and maintaining realisations of measurement scales for SI base units and other related quantities, known as primary measurement standards. These standards provide the initial stage in the dissemination of units of measurement; they anchor traceability chains, so reliable information can be obtained later at their endpoints.

Calibration is the metrological process used to disseminate measurement units along a traceability chain. It characterises a measuring system and enables measured values to be traced back to calibration standards, and ultimately to primary standards. Calibration is carried out in two distinct phases ([10] 2.39). First, the response of the system being calibrated to certain reference standards is determined. This information is then used in the second phase to establish a relationship describing the system's response to an unknown measurand—in other words, to construct a measurement model. The two phases are distinct in their modelling requirements. The first focuses on the behaviour of the measuring system and requires a model that can explain non-ideal behaviour. During that phase, the model terms that characterise non-ideal behaviour are measured. The second phase establishes a measurement model which incorporates the terms measured in the first phase.

A simple example will help to illustrate these ideas. The behaviour of a mass balance is modelled by a systematic effect, O, which offsets balance readings, and a random error, $E_{\text{rnd·}i}$, which represents the inherent variability of repeated balance readings. The ith balance indication is

$$x_i = M + O + E_{\text{rnd·}i} , \qquad (3)$$

where M is the mass being measured. Equation (3) describes the behaviour of the measuring system, but it is not a measurement model. The system response is the focus of this equation, which is sometimes referred to as an observation equation.

In the first phase of calibration, a measure of the fixed offset can be obtained by looking at the response of the balance to a mass standard. This mass,

$$M_{\text{std}} = m_{\text{std}} - E_{m_{\text{std}}} , \qquad (4)$$

is not exactly known, but when placed on the balance the indication is ⟨171⟩

$$x_{\text{cal}} = M_{\text{std}} + O + E_{\text{rnd·cal}} \tag{5}$$

$$= m_{\text{std}} - E_{m_{\text{std}}} + O + E_{\text{rnd·cal}} \; . \tag{6}$$

Rearranging, we obtain a measurement model for the offset ⟨172⟩

$$O = x_{\text{cal}} - m_{\text{std}} + E_{m_{\text{std}}} - E_{\text{rnd·cal}} \; . \tag{7}$$

The measured value of O is the difference between the balance indication, x_{cal}, and a reported value of ⟨173⟩ the mass standard, m_{std}. This is associated with two components of error: one due to the repeatability ⟨174⟩ of the balance and another associated with the accuracy of the value, m_{std}, attributed to M_{std}. The ⟨175⟩ offset O may be summarised for modelling as ⟨176⟩

$$O = o_{\text{cal}} - E_{o_{\text{cal}}} \; , \tag{8}$$

where ⟨177⟩

$$o_{\text{cal}} = x_{\text{cal}} - m_{\text{std}} \tag{9}$$

$$E_{o_{\text{cal}}} = E_{\text{rnd·cal}} - E_{m_{\text{std}}} \; . \tag{10}$$

In the second phase of calibration, a measurement model for the calibrated balance can be established, ⟨178⟩ which incorporates the information about its offset. The model relates a balance indication, x_i, to the ⟨179⟩ measurand, ⟨180⟩

$$M = x_i - O - E_{\text{rnd·}i}$$
$$= x_i - o_{\text{cal}} + E_{o_{\text{cal}}} - E_{\text{rnd·}i}$$
$$= x_i - (x_{\text{cal}} - m_{\text{std}}) - E_{m_{\text{std}}} + (E_{\text{rnd·cal}} - E_{\text{rnd·}i}) \; . \tag{11}$$

Equation (11) involves two measurement stages along a traceability chain: one is a determination ⟨181⟩ of the reference mass, M_{std}; the other is a measurement of some mass, M, using the calibrated balance. ⟨182⟩ Calibration of the system makes a connection between these two stages. The residual error in the value ⟨183⟩ $x_i - (x_{\text{cal}} - m_{\text{std}})$, which is an estimate of M, has three components, one of which is $E_{m_{\text{std}}}$ from the ⟨184⟩ previous stage. Thus, to properly account for measurement error in M we should know about the ⟨185⟩ measurement error in the calibration standard used too. ⟨186⟩

Although not shown in the example, traceability requires links to be created all the way up the ⟨187⟩ chain to a primary standard or suitable reference. Here, $E_{m_{\text{std}}}$ inherently depends on other measuring ⟨188⟩ systems, unless the measurement of M_{std} was performed using a primary mass standard. These other ⟨189⟩ measuring systems must be calibrated too. This reflects the nature of a traceability chain, in which ⟨190⟩ links are forged by successive calibrations. ⟨191⟩

2.5. Topology ⟨192⟩

Traceability chains are not necessarily simple linear structures linking one measurement stage to ⟨193⟩ the next. Chains can divide into multiple branches, and continue to grow through independent calibra- ⟨194⟩ tions and measurement processes. They may also merge, either as independent chains converging or ⟨195⟩ as previously divided branches recombining. ⟨196⟩

With careful labelling of terms, modelling can represent these structures, enabling thorough analysis. For instance, consider two masses, M_a and M_b, weighed using the same calibrated balance. The individual measurements can be modelled as:

$$M_a = x_a - (x_{cal} - m_{std}) - E_{m_{std}} + (E_{rnd \cdot cal} - E_{rnd \cdot a}) \,, \tag{12}$$

$$M_b = x_b - (x_{cal} - m_{std}) - E_{m_{std}} + (E_{rnd \cdot cal} - E_{rnd \cdot b}) \,. \tag{13}$$

These are independent branches; however, if we are interested in the mass difference, $M_a - M_b$, the branches recombine when the difference is evaluated. Subtracting (13) from (12) and simplifying, we obtain a model for the difference measurement,

$$M_a - M_b = x_a - x_b - (E_{rnd \cdot a} - E_{rnd \cdot b}) \,. \tag{14}$$

The individual mass readings are correlated by terms associated with calibration; however, equation (14) shows these shared effects do not influence the mass difference measurement.

Modelling can capture variations of this scenario. Suppose, the same mass reference is used to calibrate the balance before each measurement (see Figure 1). To distinguish the influence quantities arising during these calibrations, we write:

$$M_a = x_a - (x_{cal \cdot 1} - m_{std}) - E_{m_{std}} + (E_{rnd \cdot cal \cdot 1} - E_{rnd \cdot a}) \,, \tag{15}$$

$$M_b = x_b - (x_{cal \cdot 2} - m_{std}) - E_{m_{std}} + (E_{rnd \cdot cal \cdot 2} - E_{rnd \cdot b}) \,, \tag{16}$$

where 'cal · 1' labels quantities from one calibration, and 'cal · 2' labels those from the other. The mass difference is now

$$M_a - M_b = x_a - x_b - (x_{cal \cdot 1} - x_{cal \cdot 2}) + (E_{rnd \cdot cal \cdot 1} - E_{rnd \cdot cal \cdot 2}) - (E_{rnd \cdot a} - E_{rnd \cdot b}) \,. \tag{17}$$

The quantities arising in balance calibrations no longer cancel; however, the mass difference measurement remains insensitive to the mass calibration standard.
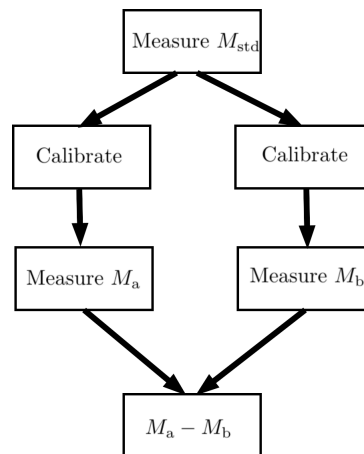
Figure 1. A measurement of mass difference offers an example of a traceability chain that divides and later recombines. The chain bifurcates when a mass balance is calibrated using the same mass, M_{std}, before each measurement. Distinct influence quantities arise during these calibrations and the mass measurements that follow, but the influences associated with the determination of M_{std} are shared and correlate the individual mass measurements. The chains combine when the mass difference is evaluated.

An intriguing outcome in these examples is the apparent insensitivity of the difference measurement to the mass standard used for calibrating the balance: it seems that a traceable calibration standard is unnecessary. However, this conclusion arises from the simple model adopted for the

measuring system, which provides a very basic representation of the balance. This issue will be   215
revisited in Section 3.2.   216

3. Different Traceability Scenarios   217

The principal argument of this article is that physical modelling can describe the relationship   218
between a measurand and a measured value of that measurand. The measurand is a defined quantity   219
whereas a measured value is the outcome of a real-world process (a measurement) influenced by   220
various physical factors. The challenge lies in finding a suitable model to represent the relationship   221
between a measured value and the measurand in different situations.   222

3.1. Quantity Ratios   223

Mathematical measurement models must adhere to the rules of quantity calculus and the laws of   224
physics. One such rule is that, for terms to be legitimately added or subtracted, they must represent   225
quantities of the same kind. Therefore, only quantities of the same kind can be compared by difference.   226
On the other hand, there is no such restriction on the multiplication and division of quantities. A   227
product of quantities is understood to be proportional to its various factors. So, the ratio of two   228
quantities of the same kind is sometimes said to be 'dimensionless', meaning that the ratio does not   229
dependent on the units used to express values in the numerator and denominator.   230
This raises an interesting question about traceability: what is an appropriate reference to establish   231
traceability when a measurand is the ratio of two quantities of the same kind? Before addressing   232
this, it is important to clarify a potential source of confusion. In dimensional analysis, dimensionless   233
quantities are treated as pure numbers, which is convenient for mathematical analysis of dimensional   234
problems; however, this is not appropriate when modelling measurements. Quantities defined as   235
ratios of the same kind of quantity should be regarded as distinct quantities in their own right. For   236
example, linear scale factor, which is a ratio of lengths, is clearly not the same as emissivity, which is a   237
ratio of energies.   238
When a measurand is defined as the ratio of quantities, there are two possible approaches   239
to establish traceability to an appropriate reference, thereby anchoring the traceability chain. One   240
approach is to trace back to a primary standard for the ratio quantity itself; the other is to trace back to   241
standards for the individual quantities in the numerator and denominator.[3] The first case is no different   242
from establishing traceability to any quantity for which a reference of the same kind is available; the   243
traceability chain can be straightforwardly modelled back to a primary standard.   244
The second case can be affected by common factors associated with the traceability chains of the   245
numerator and the denominator. For instance, using the same system to measure both quantities may   246
result in the cancellation of some terms. Nevertheless, it is important to consider the situation carefully.   247
Appendix A describes an example where a linear system measures the ratio of two quantities, Q_1 and   248
Q_2. The measurement model, given in Equation (A6), is   249

$$Q = \frac{x_i - O - E_{\text{rnd} \cdot i}}{G} \, ,$$

where G is a gain factor and O is a fixed offset. The gain factor cancels in the model of the ratio   250
measurement (A13):   251

$$\frac{Q_1}{Q_2} = \frac{x_1 - O - E_{\text{rnd} \cdot 1}}{x_2 - O - E_{\text{rnd} \cdot 2}} \, .$$

However, during calibration, the measured value of the offset (A5) is actually influenced by the system   252
gain. Equation (A14), in the appendix, provides an expression for the ratio measurement model   253
that shows how traceability is derived from the standards used to calibrate the system. Thus, the   254

---

[3] This is not limited to ratios of the same kind of quantity. For instance, measurements of speed can be traced back to standards of length and time.

traceability of a measurement of the ratio of Q is established through traceability chains that refer back to Q for both the numerator and the denominator.

3.2. Quantity Differences

In section 2.5, a simple mass balance was used to show how calibration creates links between the stages of a traceability chain. However, we noted that the chain was abruptly truncated when the balance was incorporated into a model of a mass difference measurement. In fact, the balance's representation was quite simple: it accounted for a fixed offset and a repeatability error, but assumed a perfect response to mass increments. This model allowed the balance to be calibrated by just measuring its offset. Since an offset does not affect the measurement of a mass difference, the representation effectively truncated the traceability chain for a mass difference.

A better representation of a balance could take account of the sensitivity, or gain, of its response to different masses. One way to do this is to introduce a gain factor, G, to the original model. The observation equation for an indication is then

$$x_i = G\,M + O + E_{\text{rnd}\cdot i}\,, \tag{18}$$

where M, O, and $E_{\text{rnd}\cdot i}$ represent the measurand, offset, and repeatability, respectively, as before. The parameters, G and O, must be determined by calibration, which will require at least two calibration standards, M_a and M_b.[4]

The mass difference measurement model is now

$$M_1 - M_2 = \frac{x_1 - x_2 - (E_{\text{rnd}\cdot 1} - E_{\text{rnd}\cdot 2})}{G}\,, \tag{19}$$

which does not depend on the balance offset. However, the measurement model of G used during calibration depends on M_a and M_b. Thus, the traceability of mass difference measurements using this model would be established in terms of the balance response to changes in mass.

3.3. Measurement Comparisons

As noted in the Introduction, the CIPM MRA [2] provides reliable quantitative information on the metrological compatibility of similar calibration and measurement services at different NMIs. To maintain entries in the MRA, NMIs are expected to participate regularly in international measurement comparisons relevant to their services [15]. In these *CIPM key comparisons*, a group of NMIs each measure a suitable measurand—often a stable property of an artefact—and then the comparison coordinator will apply an agreed method of analysis to the results.

Comparison analyses assess the reliability of participants' measurement capabilities. For a measurand Y, participant 'a' submits a result y_a, participant 'b' submits a result y_b, and so on, with each result accompanied by a corresponding statement of uncertainty. In modelling terms, it may be assumed that $Y = y_a - E_a$, $Y = y_b - E_b$, and so forth, where the residual errors represent the extent to which each result deviated from the measurand. A comparison analysis estimates the residuals E_a, E_b, etc., which are called degrees of equivalence (DoE). DoEs can be compared with the uncertainty information provided for each result by the participants. If a DoE is larger than can be explained by a statistical interpretation of the uncertainty, it suggests a problem with the participant's measurement analysis. Likely, some sources of measurement error have not been properly accounted for in the measurement model.

After a CIPM key comparison has been completed, similarly structured *RMO key comparisons* can be carried out by regional metrology organisations (RMOs). Participation in RMO comparisons gives many more NMIs the opportunity to register and maintain claims under the MRA. To evaluate NMI performance across all comparisons, each RMO comparison must include several participants

---

[4] Appendix A gives the calibration equations for a generic linear system measuring Q, rather than M.

from the initial CIPM comparison. The systematic effects associated with measurements made by these linking participants introduce correlations in the results. However, such linking effects can be accounted for when measurement models are combined with the algorithms used for comparison analysis. This is described in a report on the analysis methods recommended by the Consultative Committee on Photometry and Radiometry (CCPR) [16]. Succinct measurement models, using one residual error term to represent combined random effects and one to represent combined systematic effects, were used in that analysis. More detailed measurement models, representing the various influence quantities of each comparison participant, can be handled using specific data processing that is designed to evaluate a digital representation of measurement models [17]. Doing so provides more insights into the comparison analysis outcomes.

3.4. Traceability of Intrinsic and Quantum-Based Standards

Intrinsic measurement standards are standards based on an inherent and reproducible property of a phenomenon or substance ([10] clause 5.10). They serve as metrological references with assigned consensus values, such as the triple-point-of-water.

From a modelling perspective, a measurement involving an intrinsic standard can be represented by the generic model (1),

$$Y = y - E_y \, ,$$

where the single term E_y represents the combined effects of all influence quantities. This term can be expanded into an appropriately detailed model of the realisation of a particular standard. It is recognised that there are generally two sources of uncertainty in the realisation of an intrinsic standard: one associated with the consensus value (common to all realisations) and the other associated with the specific implementation. In modelling terms, this means that an intrinsic standard may have an error that is common to all standards of the same type, arising from the process that fixed a consensus value (an estimate). A second, implementation-specific error, results from influence factors in the realisation of a system.

Comparison with other standards provides evidence that the system in question is adequately represented by its model—that no unaccounted-for influences introduce bias or reduce precision—and that it is metrologically compatible with the measurand. The need to compare systems is fundamentally the same as for the metrology comparisons, described in section 3.3. To claim traceability, the metrological compatibility of realisations must be demonstrated. Without such verification, traceability cannot be assured [18].

Quantum-based standards are a type of intrinsic measurement standard. These standards rely on the realisation of certain physical properties with well-defined values determined by quantum mechanical phenomena, such as the Josephson effect or the quantum Hall effect. The need for calibration to establish traceability in such systems is seemingly unnecessary. However, their realisation is not immune to influence factors, and so careful characterisation is required and comparison with independently verified systems is necessary [19].

3.5. Traceability in Sensor Networks

Sensor networks are multiple interconnected systems of sensors, which monitor physical parameters in their environment and transmit data to a central system for aggregation and analysis. The complex topology of many sensor networks presents challenges in establishing metrological traceability. The typical hierarchical structures, where measurement standards and calibrations are used to disseminate traceability, are not easily applied to sensor networks.

Measurement modelling can help by representing sensor measurement results. For example, when sensors are manufactured in large numbers as batches, the characteristics of a batch can be estimated from a smaller sample. This characterisation can be done using traceable measuring systems. The data obtained could be incorporated in sensor models, with batch characteristics represented by common terms for all sensors, sample variability between sensors represented by sensor-dependent fixed-effect

terms and, lastly, variability between repeat observations with individual sensors represented by random effect terms. Errors due to sampling variation could also be included. Modelling in this way allows data processing to account for systematic effects, enabling the extraction of more precise information from aggregated measurements [20]. Many modern sensors are equipped with data processing capabilities, so sensor models could be integrated in the devices to provide plug-and-play functionality.

4. Model Evaluation

Having mathematical models that represent measurement results raises an important question: can the models be evaluated? Although all terms are assumed to have definite values, some remain unknown, making the computational task seem ill-defined. Nevertheless, terms that are not exactly known can be expressed as the sum of a (known) estimate and an (unknown) residual error. In this form, models can be evaluated by taking zero as the estimate for all residual error terms. This calculation will yield the measured value—the best estimate of the measurand.

However, residual measurement errors determine the ultimate accuracy of measured values, which must be accounted for when reporting traceable measurement results. So, these error terms should not be overlooked. In relation to the VIM definition of traceability, they are the subjects of an evaluation of measurement uncertainty.

In metrology, measurement uncertainty is quantified using probabilistic concepts. However, since probability can be interpreted in different ways, so too can uncertainty. The frequentist interpretation, where probability represents the long-run relative frequency of events, is likely familiar to most readers and is consistent with the GUM [13]. However, the Bayesian view, which is based on a state-of-knowledge, has been emphasised in later supplements to the GUM [21,22].

While exploring the consequences of adopting different definitions of probability—and, by extension, uncertainty—is beyond the scope of this article, this section will focus on how static measurement models can be used as a foundation for uncertainty calculations. The modelling approach described here is grounded in general physical principles, providing a robust framework for the description of measurements. Decisions about data handling—such as the choice of an appropriate probability interpretation—can be made at the end of a traceability chain based on the type of information required.

4.1. Traceability Chains

A traceability chain consists of a succession of carefully linked stages. While we may say that a 'measurement' is performed at each stage, it is the combination of stages that constitutes the actual traceable measurement. The hierarchy of stages along a traceability chain is often depicted in the form of a pyramid or triangle. At the summit are the formal definitions of quantities, implemented directly below by NMIs. Second-tier calibration laboratories then provide calibration services to testing laboratories and industry organisations. As one moves down the hierarchy, traceability chains divide into more and more branches.

At the base, information is gathered about a quantity of interest and used—rather than being passed on—to make a decision. Typically the final stage compares measured values with other measured values, or with nominal quantities. Comprehensive modelling captures the potentially complex branching structures that can arise—for example, the cases of ratio and difference measurements already mentioned and the complexities of measurement comparison analysis. Representing the complete traceability chain is essential for properly assessing the accuracy of results.

To discuss the staged nature of traceability chains, we will adopt specific notation. The interpretation of upper- and lower-case terms follows the meanings provided earlier. To represent a model at a given stage, we use functional notation with a stage-index parameter, such as

$$Y_k = f_k(\Lambda_k) \, ,$$

where k labels the stage, Y_k is the measurand at that stage, and the elements of the set Λ_k are the arguments of $f_k(\cdot)$. These arguments may include known and unknown terms evaluated at earlier stages, as well as individually identified quantities or quantity estimates. A complete measurement model is represented by iteration through the stages:

$$Y_k = f_k(\Lambda_k), \quad \text{for } k = 1, \ldots, N, \tag{20}$$

with the final stage delivering the measurand, $Y = Y_N$. It is worth noting that the mathematical composition of stage functions is implicit in this description: later stages will be composed of earlier stages.

For instance, if the first and second stage results feed into a third stage, which, in turn, feeds the fourth (Figure 2), the model could be expressed at the fourth stage as

$$Y_4 = f_4(f_3(f_1(\cdots), f_2(\cdots), \cdots), \cdots), \tag{21}$$

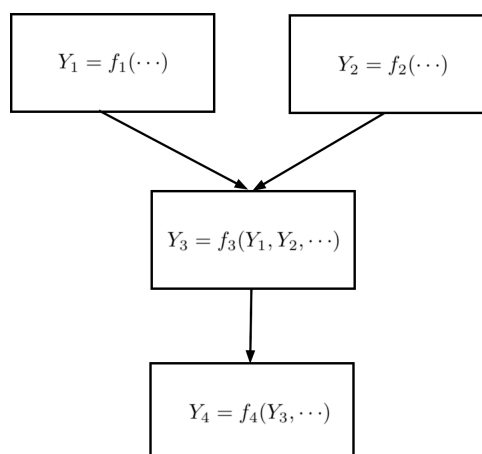which explicitly shows the composition, while omitting other arguments.

Figure 2. In general, the results from earlier stages of a traceable measurement are the arguments of later stage models. In this example, the first two stage results are needed during the third stage.

4.2. GUM Evaluation of Uncertainty

The GUM describes a method for evaluating measurement uncertainty when a function describing the measurement is available of the form

$$Y = f(X_1, X_2, \cdots, X_L), \tag{22}$$

where the output quantity Y is the measurand and $X_1, X_2, \cdots, X_L$ are input quantities on which Y depends. A summary of this method is given in Appendix B. GUM notation is compatible with the notation employed in this paper but is less strict, because capital letters can also denote random variables—an abstract mathematical concept.

Uncertainty arises because the values of some terms in the model are not know exactly. These terms correspond to the input quantities in equation (22). To evaluate the uncertainty of y as an estimate of Y, additional information about the input quantities is required. Specifically, $u(x_i)$ denotes the *standard uncertainty* of the estimate, x_i, of X_i, expressed as a standard deviation, while ν_i represents the *degrees of freedom* associated with $u(x_i)$. Furthermore, if there is correlation between estimates, the correlation coefficient $r(x_i, x_j)$ must be provided. Models developed using the approach described in this paper can be handled by the GUM methodology when values for these attributes are documented with the model.

Application of the GUM treatment of uncertainty to static models implies a frequentist interpretation of probability, where the variability of data is associated with influence quantities. The GUM

classifies methods for evaluating the uncertainty of input quantities into two groups: type-A, which applies statistical evaluation methods to data, and type-B, for evaluations that do not involve statistical analysis. Standard methods for evaluating sample statistics are typically used in type-A data processing and are rooted in classical frequentist approaches. Type-B evaluation draws on other sources of information, such as a physical model of a process that influences the measurement. This type-B analysis aligns with the view that terms represent observable quantities that vary due to physical effects. Consequently, a frequentist interpretation of probability is also applicable.

While it is sometimes argued that systematic errors complicate the frequentist analysis of a measurement, static models do not differentiate between random and systematic errors. Indeed, static modelling of the complete traceability chain effectively avoids the problem alluded to. A static model represents all the quantities that contributed to a measurement result, with each having a definite value that may or may not be known. Influences typically classified as systematic are represented by single terms with enduring values, whereas influences typically classified as random are represented by multiple terms, each with a value associated with a different part of a measurement. In this way, static models capture the effects typically ascribed to systematic errors, as demonstrated in the simple cases of difference and ratio measurements in Sections 3.2 and 3.1.

4.3. The Monte Carlo Method for Evaluating Uncertainty

Monte Carlo simulation has many applications in computational science. It is commonly used to model dynamic stochastic behaviour, providing insights into how a system evolves over time. The Joint Committee on Guides in Metrology (JCGM) has issued supplements to the GUM that describe a method of computing uncertainty called the *Monte Carlo Method* (MCM) [21,22]. However, this method does not model dynamic system behaviour ([21] Note 2, p 10). Instead, random number generation is used to evaluate Bayesian probability distributions associated with unknown fixed quantities. A brief summary of the MCM is given in Appendix C.

As explained in Section 4.2, GUM notation is compatible with the notation employed in this paper. So, models developed using the approach described here can serve as precursors to MCM uncertainty calculations. Additional information about the input quantities in equation (22) is required. These terms are associated with probability distributions representing the state of knowledge about the inputs, and the result of an MCM calculation represents the state of knowledge distribution for the output quantity Y. Applying the MCM treatment of uncertainty to these models implies a Bayesian interpretation of probability, which is not based on the variability of data, i.e., relative frequency [21, Note 4, p viii]. The interpretation of uncertainty statements produced by the MCM is therefore different from uncertainty statements produced using the GUM method.

As with the GUM method, static modelling of the full traceability chain will facilitate the formulation of MCM calculations because the distinction between random and systematic errors is unnecessary. Static modelling uniquely identifies all quantities instead of classifying some as random and others as systematic. Within the context of an MCM calculation, model terms correspond to specific realisations of random variables. Static modelling identifies common influences that affect multiple stages, enabling them to be addressed appropriately.

4.4. The Need for Documentation

The VIM definition of traceability, cited in the Introduction, refers to a "documented unbroken chain of calibrations, each contributing to the measurement uncertainty". This phrase can now be better understood in the context of model evaluation. On one hand, calibration forges links between stages in a traceability chain. An "unbroken chain" is necessary to track influences back to their origins. On the other hand, uncertainty calculations can be preformed when a model is available and when probabilistic information about terms in the model is provided. This information, along with evidence of its accuracy, must be collected and "documented".

The documentation requirements for GUM and MCM uncertainty calculations differ, and eliciting the necessary information is beyond the scope of this article. However, an important underlying aspect

of traceability, which has not been mentioned so far, pertains to the quality of information used in calculations and, consequently, the reliability of the results. One definition of traceability considered by Belanger included the sentence [5]:

> Measurements have traceability to the designated standards if and only if scientifically rigorous evidence is produced on a continuing basis to show that the measurement process is producing measurement results (data) for which the total measurement uncertainty relative to national or other designated standards is quantified.

Although this sentence never became part of VIM definitions, a shortened version was, for a time, included in NIST traceability policy [23,24]:

> It is noted that traceability only exists when scientifically rigorous evidence is collected on a continuing basis showing that the measurement process is producing documented results for which the total measurement uncertainty is quantified.

The idea here is that measurements must be performed with the measuring system under statistical control, ensuring that the unpredictable effects of influence quantities can be objectively described in terms of probability, and that those descriptions remain valid for a reasonable period before and after the measurement. This consistency is essential to ensure the reliability of results.

In practice, meeting this expectation involves verifying the performance of a measuring system to produce evidence of satisfactory operation, which is then documented. Ehrlich and Rasberry expand on this idea in their description of *Metrological Timelines in Traceability* [24].

5. Modelling for Digitalisation in Metrology?

A simple and general approach to measurement modelling has been introduced based on the scientific concepts of a physical quantity and quantity calculus. The notation, with clearly defined semantics, emphasises the distinction between known and unknown quantities, thereby facilitating the use of probability to describe unknown quantities.

In this section, we explore the modelling approach in the context of the principles adopted by the Object Management Group's Model-Driven Architecture (OMG MDA) [25]. Our motivation stems from the potential we perceive in MDA principles for developing digital representations tailored to metrological information.

Core metrological concepts, such as traceability, can be expressed in well-defined scientific language anchored to scientific principles. By explicitly defining the semantic structure of these representations—illustrated here through the modelling of calibration scenarios using a defined notation—models of specific situations maintain strict logical relationships to foundational concepts.

In parallel, domain-specific models can be developed to represent these scenarios in digital systems, supporting such functionalities as evaluating numerical results and their associated uncertainties. If digital tools have been designed to conform to the same scientific principles, their semantic consistency will facilitate interoperability across diverse systems. This section provides a broad overview of these ideas. We believe this is an area that would benefit from further research efforts.

5.1. Measurement Modelling from the MDA Perspective

The MDA approach uses a model hierarchy to achieve a clear separation of semantic structures and provide flexibility for development. This has been explained by Bézivin, whose work has been influential within the field of model-driven engineering [26]. The approach distinguishes four abstraction levels:

M0: A system (i.e., the real-world entity or phenomenon being represented).

M1: A model of the system, capturing its structure and behaviour in a specific context.

M2: A meta-model, which defines the elements, relationships, and rules used to construct models at the M1 level.

M3: A meta-meta-model, which provides the foundational concepts for defining meta-models.

This 4-level structure aligns with the modelling approach in this paper, illustrated in Figure 3. The real-world system of interest, at M0 in the MDA hierarchy, is a measurement result. At the opposite end, M3 in the hierarchy represents the conceptualisation of a measurement result, grounded in the concept of a physical quantity and the rules of quantity calculus, which define the operations applicable to such quantities. These concepts form the basis for the meta-models at the M2 level, where the entities and structures used to describe measurements are formalised. The entities used for modelling include known and unknown quantities, as well as arithmetic operators, as introduced in Section 2. These elements are identified by specific notation to facilitate the expression of quantity equations.
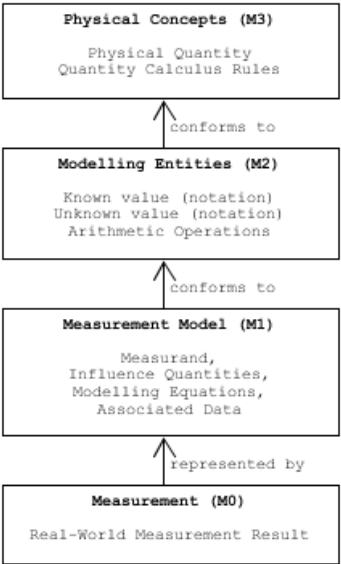

Figure 3. The four-level hierarchy applied to measurement modelling. A model at level M1 represents the outcome of a measurement in the real world (M0). The elements and rules for constructing this model are defined in the meta-model at level M2. At the top level (M3), fundamental concepts rooted in basic scientific principles form the foundation for all levels. Each M1 model represents a measurement result at M0, conforms to the meta-model at defined M2, and aligns with the high-level concepts established at M3.

Models constructed at the M1 level provide concrete representations of specific measurements. Our focus is on the processes that generate a traceable result. This can be viewed as a sequence of stages, each involving definite quantities that collectively determine the final outcome. Factors such as aleatory behaviour and time dependence of quantities are not relevant.

M1 models represent specific aspects of measurement results to enable reasoning about their properties, such as traceability. In some cases, M1 models may impose additional domain-specific constraints not defined in the M2 meta-model. For example, while quantity calculus allows for the addition of quantities, intensive quantities like temperature cannot be meaningfully added.

5.2. Parallel Hierarchies

Parallel model hierarchies can be developed, extending from the same foundational M3 model, but defining different software entities at the M2 level. Programs at the M1 level, built on these M2 definitions, may serve as models representing different aspects of real-world measurement results.

For instance, distinct M2-level meta-models could support the two types of uncertainty calculation discussed in Sections 4.2 and 4.3 (Figure 4). A meta-model tailored for GUM uncertainty calculations would define the data type for unknown quantities with attributes such as a value, standard uncertainty, and degrees of freedom. Conversely, a meta-model for MCM calculations would represent unknown quantities with attributes for a value, the state-of-knowledge distribution type, and its associated parameters. The GUM meta-model would also describe how uncertainty is propagated and evaluated using the LPU (A16) and Welch-Satterthwaite formulae (A17). In contrast, the MCM meta-model

would define control structures for Monte Carlo simulations and methods for summarising results. {539}
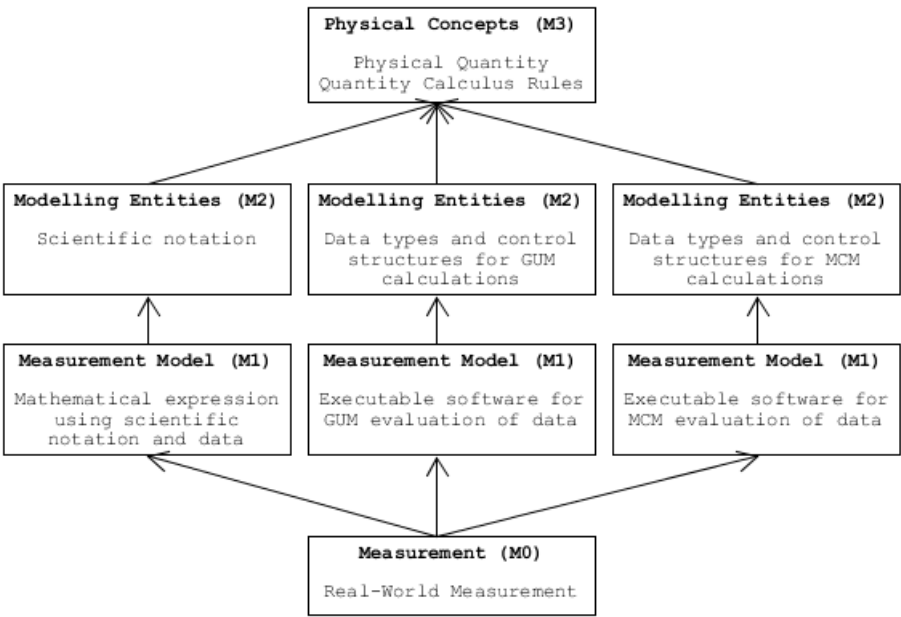M2 meta-models would also identify the notion of probability applicable to M1 models. {540}

Figure 4. Three parallel branches of measurement modelling: one using scientific notation, one employing software designed for GUM uncertainty evaluations, and one using software tailored to MCM uncertainty calculations. Each branch originates from the same conceptual basis and addresses the same measurement result in the real-world. While models at a given level are not necessarily directly equivalent between branches, their derivation from shared concepts ensures that logical transformation between models is possible.

5.3. Can the MDA Approach Help in Metrology? {541}

The MDA approach can support the digitalisation of metrology by facilitating a clear separation {542}
of concerns. The methodology will allow metrologists to focus on describing their problems using {543}
metrological principles in scientific terms, while information technologists can concentrate on the {544}
design and implementation of robust digital systems. This will help to maintain metrological under- {545}
standing and scientific integrity throughout the digitalisation process, without placing unnecessary {546}
cross-disciplinary demands on either metrologists or IT developers. It will also insulate the new digital {547}
metrology infrastructure from problems associated with rapid evolution of digital technologies. Flater {548}
has expressed similar ideas, observing that the conceptual foundations of a documentary standard {549}
often evolve at a different pace than the technologies used in its implementation—a challenge that {550}
MDA principles could help address [27]. {551}

The key feature of the MDA approach that enables this separation of concerns is its establishment {552}
of a common conceptual foundation. This paper has illustrated the approach using a simple conceptual {553}
model applied to measurement result modelling. Although simple, this conceptual foundation is {554}
actually quite effective. Being grounded in fundamental scientific concepts, such as quantities, quantity {555}
calculus, and probability, it captures the essential elements required for modelling. Its simplicity is a {556}
strength, facilitating clear and unambiguous communication of ideas while offering the flexibility to {557}
develop more complex meta-models tailored to specific applications. {558}

Early efforts in the digitalisation of metrology have not exploited hierarchical modelling. Instead, {559}
developers have created a variety of digital systems where the conceptual foundations are implied {560}
rather than explicitly defined. This reliance on a tacit understanding of underlying concepts introduces {561}
a risk of ambiguity and inconsistency, particularly when different groups of people or systems interact. {562}

By adopting the MDA approach, however, the metrology community can identify and apply {563}
appropriate and well-defined conceptual foundations. This clarity will foster consistency and col- {564}

laboration between disciplines, reducing misunderstandings and facilitating the creation of more interoperable systems.

6. Discussion

Another perspective on the central theme of this paper is that all measured values are inherently wrong; they inevitably contain some degree of error. Metrologically traceable measurements, however, provide an objective means to quantify the magnitude of that error, enabling the accuracy of measurement to be evaluated so that the usefulness of a result can be considered. Ehrlich and Rasberry suggested that the primary use of traceability is to answer the questions ([24] section 1.1):

What correction should be applied to a measurement result [...] to match the result that would be obtained using the instrument (standard) to which traceability is desired? What is the uncertainty of this corrected result?

These questions apply between adjacent stages in a traceability chain, where the correction required is the (unknown) value of residual error contributed by the latest stage. However, by iteratively asking these questions back along a traceability chain, we see that the idea aligns with our simple generic model

$$Y = y - E_y \; .$$

The term uncertainty in the VIM definition of traceability serves a pragmatic purpose. Metrologists are accustomed to evaluating the accuracy of measurements in their specialist fields, using methods described in the GUM or its supplements. The results of these evaluations can be reported in a few different formats described in the GUM. So, the VIM guides people to use familiar processes that harmonise the way in which information is shared.

Nevertheless, the concept of residual measurement error—not measurement uncertainty—is central to the scientific understanding of traceability. These errors are physical quantities that represent small deviations from nominal or estimated quantity values. Consequently, they can be modelled and analysed as elements of physical systems. Although residual errors give rise to measurement uncertainty in the actual values of quantities, errors and uncertainties are not interchangeable. In metrology, uncertainty broadly refers both to probabilistic descriptions of terms in measurement models (e.g., standard uncertainty) and to the evaluation of statistical inferences about quantity values (e.g., expanded uncertainty) [12]. The meaning of uncertainty is further complicated by the various interpretations of probability employed within the metrological community.

For these reasons, we contend that static modelling of traceability chains captures the essence of the scientific nature of metrological traceability by highlighting the fundamental role of residual errors. The language of science is appropriate for this purpose, and the strict logic of physical quantity equations allows for rigorous analysis of measurement scenarios while mathematical notation avoids some of the challenges inherent in verbal descriptions.

Several pieces of evidence support this contention. The most striking example comes from the microwave metrology community, which has adopted a modelling approach to describe measurements of complex-valued transmission line components [28]. Rigorous modelling has significantly enhanced the information available through traceable measurements [29]. Modelling, supported by software, can unravel complicated effects due to common influence quantities involved in instrument calibration procedures. It is possible to pass detailed information between stages along the traceability by adopting digital file formats that contain information about the uncertainty components due to residual errors at each stage. This enables down-stream users to benefit from the detailed modelling done by an NMI [30].

Another recent example is a study that applied modelling, again supported by specialised software, to a four-axis goniometric system measuring optical reflectance [31]. The model of this system has many configuration terms that are not known exactly and must be considered to account for the accuracy of the result. By modelling the set-up, the system performance in different configurations

could be examined enabling a better understanding of the correlations between various measurement errors. This understanding significantly improved the accuracy obtainable for certain measurements.

A common feature of these examples is the use of special software for model evaluation. Several NMIs have developed packages that implement the GUM method for uncertainty calculation applied to a traceability chain [32,33]. The software uses an abstract data type to represent unknown quantities [34]. Tailored to the requirements of the GUM method, the data type is designed to build a digital representation of the model during data processing, incorporating the uncertainty components associated with influence quantities, which are uniquely identified. Propagating these records between stages allows a full evaluation of combined uncertainty and degrees of freedom at the end of a traceability chain (see Appendix B).

The entities in the software code for model evaluation correspond to terms in the mathematical measurement model (illustrating the correspondence discussed in Section 5.2). They possess attributes enabling the GUM calculation of uncertainty (Appendix B). However, the overarching conceptual model remains unspecified—tacitly implied rather than explicitly stated. We contend that the usefulness of these software packages highlights the potential benefits of adopting a more formal semantic structure, in line with MDA principles.

7. Conclusions

The key feature of a metrologically traceable measurement result is its ability to be meaningfully compared with another traceable result or a nominal value expressed on the same measurement scale. Traceable results include an objective quantitative assessment of accuracy, enabling informed critical decisions to be made when physical quantities are involved.

Metrological traceability incorporates an assessment of the accuracy of measurements by considering the residual measurement errors involved. The accuracy of a traceable result is inevitably limited by the accumulation of residual errors during the measurement. However, the distinctive feature of traceable measurements is that the physical effects that give rise to these errors are identified, and their impact on the final result is modelled to evaluate the accuracy.

This paper has introduced a structured approach to modelling, incorporating a straightforward style of notation. We have emphasised the importance of modelling the entire traceability chain, from the realisation of primary reference standards through to the end user. This ensures that the accuracy of a final result can be traced back to the errors that occurred in the measuring processes. We have also emphasised the requirement for a static model of the quantities involved. A static model consists of terms representing definite values of quantities contributing to the final result—like a snapshot of the physical situation.

The described modelling approach is grounded in fundamental physical principles, including the concepts of physical quantities and quantity calculus. It does not involve notions of probability or measurement uncertainty; however, it serves as a useful precursor to uncertainty calculations, whether performed using the original GUM method or the GUM Supplement Monte Carlo Method. These methods of evaluating uncertainty require a static model of the measurement. Static modelling of the entire chain has the advantage of eliminating the need to explicitly distinguish between random and systematic errors, as both are inherently accounted for within the model.

The importance of identifying a formal semantic basis for modelling has been discussed in relation to modern methods of digital system design. We argue that creating hierarchical modelling structures, anchored in well-defined conceptual models, holds promise for the further development of digitalisation in metrology. This approach offers a clearer separation of concerns between metrology and technology, both during the development of new digital systems and their subsequent maintenance. It is well-suited to a more formal, structured, enterprise-based framework for the overarching architectural plan of digital transformation in metrology.

Abbreviations

The following abbreviations are used:

CIPM	International Committee for Weights and Measures
CGPM	General Conference on Weights and Measures
ILAC	International Laboratory Accreditation Cooperation
GUM	Guide to the Expression of Uncertainty in Measurement
VIM	International Vocabulary of Metrology
ISO	International Standards Organisation
MRA	Mutual Recognition Agreement
NMI	National Metrology Institute
NIST	National Institute of Standards and Technology
QI	Quality Infrastructure

Appendix A. Calibrating a Linear Measuring System

The observation equation of simple system for measuring a quantity Q is

$$x_i = G\,Q + O + E_{\text{rnd}\cdot i}\;. \tag{A1}$$

The system produces an indication x_i for each measurement. Its response is characterised by a gain, G, and an offset, O, both of which must be determined to calibrate the system. Additionally, measured values exhibit variability, which is modelled by a term $E_{\text{rnd}\cdot i}$ for each indication.

A minimum of two calibration standards are needed to measure G and O. We assume that standards for $Q_a = q_a - E_{q_a}$ and $Q_b = q_b - E_{q_b}$ are available, and that these quantities are not exactly known. The system responses to the calibration standards are

$$x_a = G\,Q_a + O + E_{\text{rnd}\cdot a}\,, \tag{A2}$$

$$x_b = G\,Q_b + O + E_{\text{rnd}\cdot b}\;. \tag{A3}$$

Solving for the gain and offset we obtain

$$G = \frac{x_a - x_b - (E_{\text{rnd}\cdot a} - E_{\text{rnd}\cdot b})}{Q_a - Q_b}\;, \tag{A4}$$

$$O = \frac{[x_a + x_b - (E_{\text{rnd}\cdot a} + E_{\text{rnd}\cdot b}) - G(Q_a + Q_b)]}{2}\;. \tag{A5}$$

The measurement model of the system is

$$Q = \frac{x_i - O - E_{\text{rnd}\cdot i}}{G}\;. \tag{A6}$$

The explicit influence of factors involved in the calibration can be seen by substituting for G, O, Q_a, and Q_b and rearranging terms

$$Q = \frac{x_i - E_{\text{rnd}\cdot i}}{G} - \frac{x_a + x_b - (E_{\text{rnd}\cdot a} + E_{\text{rnd}\cdot b})}{2G} + \frac{Q_a + Q_b}{2} , \tag{A7}$$

$$= \frac{x_i - E_{\text{rnd}\cdot i}}{G} - \frac{x_a + x_b - (E_{\text{rnd}\cdot a} + E_{\text{rnd}\cdot b})}{2G} + \frac{q_a + q_b - (E_{q_a} + E_{q_b})}{2} , \tag{A8}$$

$$= \left[\frac{q_a - q_b - (E_{q_a} - E_{q_b})}{x_a - x_b - (E_{\text{rnd}\cdot a} - E_{\text{rnd}\cdot b})} \right] \left[x_i - E_{\text{rnd}\cdot i} - \frac{x_a + x_b - (E_{\text{rnd}\cdot a} + E_{\text{rnd}\cdot b})}{2} \right]$$
$$+ \frac{q_a + q_b - (E_{q_a} + E_{q_b})}{2} . \tag{A9}$$

It is not uncommon for one of the calibration standards to represent the absence of Q (e.g., $q_a = 0$), corresponding to a nominal value of zero, and another to represent a nominal value of unity (e.g., $q_b = 1$). However, this does not significantly simplify the model (A9), as it does remove any of the influence terms.

Appendix A.1. Difference Measurement

If the system is used to measure a difference, Q_1 and Q_2, the offset term cancels. However, the approximate values attributed to calibration standards continue to influence the measurement:

$$Q_1 - Q_2 = \frac{x_1 - x_2 - (E_{\text{rnd}\cdot 1} - E_{\text{rnd}\cdot 2})}{G} , \tag{A10}$$

$$= [x_1 - x_2 - (E_{\text{rnd}\cdot 1} - E_{\text{rnd}\cdot 2})] \frac{Q_a - Q_b}{x_a - x_b - (E_{\text{rnd}\cdot a} - E_{\text{rnd}\cdot b})} , \tag{A11}$$

$$= [x_1 - x_2 - (E_{\text{rnd}\cdot 1} - E_{\text{rnd}\cdot 2})] \frac{q_a - q_b - (E_{q_a} - E_{q_b})}{x_a - x_b - (E_{\text{rnd}\cdot a} - E_{\text{rnd}\cdot b})} . \tag{A12}$$

Appendix A.2. Ratio Measurement

If the system is used to measure a ratio, Q_1/Q_2, the gain term cancels out in the measurement model. However, the offset remains as an influence factor and is correlated with the gain measurement during calibration

$$\frac{Q_1}{Q_2} = \frac{x_1 - O - E_{\text{rnd}\cdot 1}}{x_2 - O - E_{\text{rnd}\cdot 2}} , \tag{A13}$$

$$= \frac{x_1 - [x_a + x_b - (E_{\text{rnd}\cdot a} + E_{\text{rnd}\cdot b}) - G(Q_a + Q_b)]/2 - E_{\text{rnd}\cdot 1}}{x_2 - [x_a + x_b - (E_{\text{rnd}\cdot a} + E_{\text{rnd}\cdot b}) - G(Q_a + Q_b)]/2 - E_{\text{rnd}\cdot 2}} . \tag{A14}$$

By incorporating more detailed expressions for G, Q_a, and Q_b, equation (A14) can be expanded to reveal specific residual errors associated with calibration, as in equation (A12).

Appendix B. GUM Evaluation of Uncertainty—Summary

The GUM describes a method for evaluating measurement uncertainty based on a function of the form given in (22):

$$Y = f(X_1, X_2, \cdots, X_L) ,$$

where the output quantity Y is the measurand and $X_1, X_2, \cdots, X_L$ are input quantities on which Y depends. A measured value, y, for Y, is obtained by evaluating $f(\cdot)$ using the best estimates available for each of the input quantities, $x_1, x_2, \cdots, x_L$.

The method of evaluating the uncertainty is known as the *Law of Propagation of Uncertainty* (LPU) ([13] clause 5). To evaluate the uncertainty of y as an estimate of Y, the following information about input estimates is needed:

- each input estimate has a *standard uncertainty*, denoted $u(x_i)$,

- each standard uncertainty has a number of *degrees of freedom*, denoted ν_i,   703
- a correlation coefficient, $r(x_i, x_j)$, must be provided if input estimates are correlated.   704

The LPU considers components of uncertainty, which are defined in terms of the sensitivity of Y 705
to changes in a particular input. The component of uncertainty for X_i is   706

$$u_i(y) = \frac{\partial Y}{\partial X_i}\, u(x_i) \; . \tag{A15}$$

The LPU evaluates the *combined standard uncertainty* of y as   707

$$u(y) = \left[\sum_{i=1}^{L}\sum_{j=1}^{L} u_i(y)\, r(x_i, x_j)\, u_j(y) \, , \right]^{1/2} \tag{A16}$$

where $r(x_i, x_j) = 1$ when $i = j$ and $r(x_i, x_j) = 0$ in the absence of correlation. There is also an expression   708
called the Welch-Satterthwaite formula to evaluate ν_y, the *effective degrees of freedom* of $u(y)$ ([13] clause   709
G.4),   710

$$\frac{u^4(y)}{\nu_y} = \sum_{i=1}^{L} \frac{u_i^4(y)}{\nu_i} \; . \tag{A17}$$

Thus, the estimate y of Y is associated with a standard uncertainty that has a number of degrees of   711
freedom, just like the input quantity estimates.   712

We take the view that equation (22) should describe the whole traceability chain, which raises the   713
question: can the LPU still be implemented if calculations are performed at each stage and the results   714
are passed from stage to stage? The answer is yes [35]. Evaluation of equations (A16) and (A17) is only   715
required at the end of the chain. However, to enable this, a set of uncertainty components must be   716
determined at each stage, along with the corresponding measured value. So, the intermediate results   717
propagated from one stage to the next should consist of the stage value, y_k, and the set of components   718
of uncertainty in y_k for all inputs. These components of uncertainty are of the form   719

$$u_i(y_k) = \frac{\partial Y_k}{\partial X_i}\, u(x_i) \; .$$

Appendix C. The Monte Carlo Method of Uncertainty Evaluation—Summary   720

The first supplement to the GUM describes a *Monte Carlo Method* (MCM) for evaluating mea-   721
surement uncertainty ([21] clause 5.9). The approach also applies to a function of the form given in   722
(22):   723

$$Y = f(X_1, X_2, \cdots, X_L) \, ,$$

where the output quantity Y is the measurand and $X_1, X_2, \cdots, X_L$ are input quantities on which Y 724
depends. The second supplement to the GUM, which addresses the evaluation of uncertainty in   725
multivariate quantities, describes the extension of the MCM to multivariate measurands ([22] clause 7).   726

To evaluate the uncertainty of y as an estimate of Y, a joint probability density function for the   727
input quantities to equation (22) must be determined ([22] clause 6). This task is potentially challenging;   728
however, in practice a selection of common distributions is available to choose from ([21] Table 1).   729

The MCM implements an approximate form of propagation of distributions. It samples from   730
random deviates generated by probability distributions associated with the input quantities. Repeating   731
this many times accumulates a set of values, which is characteristic of the probability distribution   732
function, $G_Y(\eta)$, for the output Y. The MCM data set can be used to evaluate the expectation and   733
suitable quantile information for $G_Y(\eta)$ that can be reported as a measured value and its uncertainty,   734
respectively.   735

Our view is that equation (22) should describe the full traceability chain, which raises the question: 736
can the MCM still be implemented if calculations are performed at each stage and the results are   737
passed from stage to stage? It is more difficult to give a clear answer to this question than it was for the   738

LPU calculation in Appendix B. Often it will be sufficient to pass the MCM data set, evaluated for Y_k at the kth stage, to the next stage. However, when the traceability chain has a branching structure that is not a simple tree, common influence factors may become important. If MCM data for a common influence must be accessed at several stages, it will complicate the information-passing requirements between stages.

References

1. BIPM.; OIML.; ILAC.; ISO. Joint BIPM, OIML, ILAC, and ISO Declaration on Metrological Traceability.
2. CIPM. Mutual recognition of national measurement standards and of calibration and measurement certificates issued by national metrology institutes, 1999. https://www.bipm.org/documents/20126/43742162/CIPM-MRA-2003.pdf (accessed 17 April 2023).
3. ILAC. ILAC Mutual Recognition Arrangement: Scope and Obligations, 2022. https://ilac.org/publications-and-resources/ilac-policy-series/ (accessed 17 April 2023).
4. Rasberry, S.D.; Belanger, B.; Garner, E.; Brickencamp, C.; Ehrlich, C.D., Traceability: An Evolving Concept. In *A Century of Excellence in Measurements, Standards, and Technology*; National Institute of Standards and Technology, 2001; Vol. SP 958, pp. 167–171.
5. Belanger, B.C. Traceability: an evolving concept. *ASTM Standardization News* **1980**, *8*, 22–28. Republished in *Bulletin OIML* **1980**, *78*, 21–25.
6. Nicholas, J.V.; White, D.R. *Traceable temperatures: an introductory guide to temperature measurement and calibration*; Science Information Division, DSIR: Wellington, New Zealand, 1982. (subsequent editions published by John Wiley & Sons).
7. BIPM.; IEC.; ISO.; OIML. *Vocabulary of metrology, Part 1, basic and general terms (international)*, 1st ed.; ISO: Genève, Switzerland, 1984.
8. BIPM.; IEC.; IFCC.; IUPAC.; IUPAP.; ISO.; OIML. *International vocabulary of basic and general terms in metrology*, 2nd ed.; ISO: Genève, Switzerland, 1993.
9. BIPM.; IEC.; IFCC.; ISO.; IUPAC.; IUPAP.; OIML. *Guide to the Expression of Uncertainty in Measurement*, 1st ed.; ISO: Switzerland, Genève, 1993.
10. BIPM.; IEC.; IFCC.; ILAC.; IUPAC.; IUPAP.; ISO.; OIML. *International vocabulary of metrology – Basic and general concepts and associated terms (VIM)*, 3rd ed.; BIPM Joint Committee for Guides in Metrology: Paris, Sèvres, 2012.
11. CGPM. Resolution 2 — On the global digital transformation and the International System of Units, 2022. https://www.bipm.org/en/cgpm-2022/resolution-2 (accessed 17 April 2023).
12. White, D.R.; Hall, B.D.; Saunders, P. The purposes of measurement uncertainty. *AIP Conf. Proc.* **2024**, *3220*, 150001. https://doi.org/10.1063/5.0234063.
13. BIPM.; IEC.; IFCC.; ISO.; IUPAC.; IUPAP.; OIML. *Evaluation of measurement data – Guide to the expression of uncertainty in measurement JCGM 100:2008 (GUM 1995 with minor corrections)*, 1st ed.; BIPM Joint Committee for Guides in Metrology: Paris, Sèvres, 2008.
14. Willink, R. A formulation of the law of propagation of uncertainty to facilitate the treatment of shared influences. *Metrologia* **2009**, *46*, 145–153. https://doi.org/10.1088/0026-1394/46/3/001.
15. CIPM. Measurement comparisons in the CIPM MRA. Technical report, Bureau International des Poids et Mesures, Pavillon de Breteuil, F-92312 Sèvres Cedex, 2021. version 1.0.
16. Koo, A.; Hall, B.D. Linking an RMO or bilateral comparison to a primary CCPR comparison. Technical Report 0776, Callaghan Innovation, Lower Hutt, New Zealand, 2020. https://doi.org/10.5281/zenodo.3958381.
17. Hall, B.D.; Koo, A. Digital Representation of Measurement Uncertainty: A Case Study Linking an RMO Key Comparison with a CIPM Key Comparison. *Metrology* **2021**, *1*, 166–181. https://doi.org/10.3390/metrology1020011.
18. Wallard, A.J.; Quinn, T.J. "Intrinsic" standards – are they really what they claim? *Cal Lab: The International Journal of Metrology* **1999**, *Nov/Dec*, 28–30.
19. Pettit, R.; Jaeger, K.; Ehrlich, C. Issues in Purchasing and Maintaining Intrinsic Standards. *Cal Lab: The International Journal of Metrology* **2000**.
20. Hall, B.D. An Opportunity to Enhance the Value of Metrological Traceability in Digital Systems. In Proceedings of the 2019 IEEE International Workshop on Metrology for Industry 4.0 and IoT (MetroInd4.0&IoT). IEEE, 2019, pp. 16–21. https://doi.org/10.1109/METROI4.2019.8792841.

21.   BIPM.; IEC.; IFCC.; ISO.; IUPAC.; IUPAP.; OIML. *Evaluation of measurement data—Supplement 1 to the "Guide to the expression of uncertainty in measurement"—Propagation of distributions using a Monte Carlo method JCGM 101:2008*, 1st ed.; BIPM Joint Committee for Guides in Metrology: Paris, Sèvres, 2008. https://doi.org/10.59161/JCGM101-2008.

22.   BIPM.; IEC.; IFCC.; ISO.; IUPAC.; IUPAP.; OIML. *Evaluation of measurement data—Supplement 2 to the "Guide to the expression of uncertainty in measurement"—Extension to any number of output quantities JCGM 102:2011*, 1st ed.; BIPM Joint Committee for Guides in Metrology: Paris, Sèvres, 2011. https://doi.org/10.59161/JCGM102-2011.

23.   Hebner, R.E. Calibration traceability: a summary of NIST's view, 1996.

24.   Ehrlich, C.D.; Rasberry, S.D. Metrological timelines in traceability. *J. Res. Natl. Inst. Stand. Technol.* **1998**, *103*, 93. https://doi.org/10.6028/jres.103.005.

25.   Object Management Group. Model Driven Architecture (MDA), 2003. https://www.omg.org/mda/ (accessed 17 January 2025).

26.   Bézivin, J. On the Unification Power of Models. *Software and Systems Modeling* **2005**, *4*, 171–188. https://doi.org/10.1007/s10270-005-0079-0.

27.   Flater, D. Impact of model-driven standards. In Proceedings of the 35th Annual Hawaii International Conference on System Sciences, 2002, pp. 3706–3714. https://doi.org/10.1109/HICSS.2002.994500.

28.   Zeier, M.; Allal, D.; Judaschke, R. *Guidelines on the Evaluation of Vector Network Analysers (VNA)*, 3rd ed.; Vol. cg-12, *EURAMET Calibration Guide*, EURAMET, 2018.

29.   Wollensack, M.; Hoffmann, J.; Ruefenacht, J.; Zeier, M. VNA Tools II: S-parameter uncertainty calculation. In the ARFTG Microwave Measurement Conference Digest, Montreal, QC, Canada, June 2012; Vol. 79. DOI: 10.1109/ARFTG79.2012.6291183.

30.   Zeier, M.; Rüfenacht, J.; Wollensack, M. VNA Tools – a software for metrology and industry. In *METinfo*; METAS, 2020; Vol. 27/2, pp. 4–6.

31.   Molloy, E.; Saunders, P.; Koo, A. Effects of rotation errors on goniometric measurements. *Metrologia* **2022**, *59*, 025002.

32.   Zeier, M.; Wollensack, M.; Hoffmann, J. METAS UncLib – a measurement uncertainty calculator for advanced problems. *Metrologia* **2012**, *49*, 809–815.

33.   Hall, B.D. The GUM tree calculator: A python package for measurement modelling and data processing with automatic evaluation of uncertainty. *Metrology* **2022**, *2*, 128–149.

34.   Hall, B.D. Computing with Uncertain Numbers. *Metrologia* **2006**, *43*, L56—L61.

35.   Hall, B.D. Propagating uncertainty in instrumentation systems. *IEEE Trans. Instrum. Meas.* **2005**, *54*, 2376–2380.