



metrology

Measurement Uncertainty

Edited by

Simona Salicon

Printed Edition of the Special Issue Published in *Metrology*

Measurement Uncertainty

Measurement Uncertainty

Editor

Simona Salicone

MDPI • Basel • Beijing • Wuhan • Barcelona • Belgrade • Manchester • Tokyo • Cluj • Tianjin



Editor

Simona Salicone
Politecnico di Milano
Italy

Editorial Office

MDPI
St. Alban-Anlage 66
4052 Basel, Switzerland

This is a reprint of articles from the Topical Collection published online in the open access journal *Metrology* (ISSN) (available at: https://www.mdpi.com/journal/metrology/topical_collections/Measurement_Uncertainty).

For citation purposes, cite each article independently as indicated on the article page online and as indicated below:

LastName, A.A.; LastName, B.B.; LastName, C.C. Article Title. <i>Journal Name</i> Year , <i>Volume Number</i> , Page Range.
--

ISBN 978-3-0365-6608-5 (Hbk)

ISBN 978-3-0365-6609-2 (PDF)

© 2023 by the authors. Articles in this book are Open Access and distributed under the Creative Commons Attribution (CC BY) license, which allows users to download, copy and build upon published articles, as long as the author and publisher are properly credited, which ensures maximum dissemination and a wider impact of our publications.

The book as a whole is distributed by MDPI under the terms and conditions of the Creative Commons license CC BY-NC-ND.

Contents

About the Editor	vii
Preface to “Measurement Uncertainty”	ix
Simona Salicone New Frontiers in Measurement Uncertainty Reprinted from: <i>Metrology</i> 2022 , 2, 29, doi:10.3390/metrology2040029	1
Simona Salicone and Harsha Vardhana Jetti A General Mathematical Approach Based on the Possibility Theory for Handling Measurement Results and All Uncertainties Reprinted from: <i>Metrology</i> 2021 , 1, 6, doi:10.3390/metrology1020006	5
Harsha Vardhana Jetti and Simona Salicone A Possibilistic Kalman Filter for the Reduction of the Final Measurement Uncertainty, in Presence of Unknown Systematic Errors Reprinted from: <i>Metrology</i> 2021 , 1, 3, doi:10.3390/metrology1010003	23
Guglielmo Frigo and Marco Agostoni Calibration of a Digital Current Transformer Measuring Bridge: Metrological Challenges and Uncertainty Contributions † Reprinted from: <i>Metrology</i> 2021 , 1, 7, doi:10.3390/metrology1020007	37
Blair D. Hall and Annette Koo Digital Representation of Measurement Uncertainty: A Case Study Linking an RMO Key Comparison with a CIPM Key Comparison Reprinted from: <i>Metrology</i> 2021 , 1, 11, doi:10.3390/metrology1020011	51
Nikolay V. Kornilov, Vladimir G. Pronyaev and Steven M. Grimes Systematic Distortion Factor and Unrecognized Source of Uncertainties in Nuclear Data Measurements and Evaluations Reprinted from: <i>Metrology</i> 2022 , 2, 1, doi:10.3390/metrology2010001	67
Ian Smith, Yuhui Luo and Daniel Hutzschenreuter The Storage within Digital Calibration Certificates of Uncertainty Information Obtained Using a Monte Carlo Method Reprinted from: <i>Metrology</i> 2022 , 2, 3, doi:10.3390/metrology2010003	85
Blair D. Hall The GUM Tree Calculator: A Python Package for Measurement Modelling and Data Processing with Automatic Evaluation of Uncertainty Reprinted from: <i>Metrology</i> 2022 , 2, 9, doi:10.3390/metrology2010009	99
Christoph-Alexander Holst and Volker Lohweg Designing Possibilistic Information Fusion—The Importance of Associativity, Consistency, and Redundancy Reprinted from: <i>Metrology</i> 2022 , 2, 12, doi:10.3390/metrology2020012	121
Salvatore Dello Iacono, Giuseppe Di Leo, Consolatina Liguori and Vincenzo Paciello The Obtainable Uncertainty for the Frequency Evaluation of Tones with Different Spectral Analysis Techniques Reprinted from: <i>Metrology</i> 2022 , 2, 13, doi:10.3390/metrology2020013	157

Gorka Kortaberria, Unai Mutilba, Sergio Gomez and Brahim Ahmed
Three-Dimensional Point Cloud Task-Specific Uncertainty Assessment Based on ISO 15530-3
and ISO 15530-4 Technical Specifications and Model-Based Definition Strategy
Reprinted from: *Metrology* **2022**, 2, 24, doi:10.3390/metrology2040024 **171**

About the Editor

Simona Salicone

Simona Salicone (M.Sc. 2000, PhD 2004) is an Associate Professor of electrical and electronic measurements at Politecnico di Milano (DEIB). Her research activity deals with Digital Signal Processing applied to measurements with particular concern to the metrology characterization of the measurement systems based on complex algorithms and with the analysis of advanced mathematical methods for uncertainty representation and estimation. In this field, she has proposed an innovative approach to the expression and estimation of uncertainty in measurement framed within the mathematical theory of evidence, a more general theory than the probability theory. The new proposed approach has been applied in different field: mechanics, electrical, biomedical.

Her research activity is reflected in the over 130 papers published in international and national scientific journals, in the proceedings of national and international conferences in the field of instrumentation and measurements, and in two monographs, edited by Springer: "Measurement Uncertainty. An approach via the mathematical theory of evidence" (2007) and "Measuring Uncertainty within the Theory of Evidence" (2018).

In 2004 she has received the IEEE Instrumentation and Measurement **Outstanding Young Engineer Award**. In 2007, she is elected to the grade of **IEEE Senior Member**. Since 2014 until the end of 2016, she has been part of the **Editorial Board of the IEEE Instrumentation and Measurement Magazine**, and, since January 2016 until the end of 2017, she has been the **Associate Editor in Chief of the IEEE Instrumentation and Measurement Magazine**. In 2016 she has received the recognition as one of the Transactions Outstanding Reviewers of 2015. She is the winner of the **2016 IEEE Instrumentation and Measurement Society Faculty Course Development Award**. Since 2019, she is part of the **Selection Committee of the IEEE Faculty Course Development Award**. Since 2021, she has been part of the **Editorial Board of the MDPI Metrology journal**.

Preface to "Measurement Uncertainty"

This book treats about measurement uncertainty, both from the theoretical point of view and from the practical point of view. From the theoretical point of view, contributions are present which propose more general mathematical methods to handle measurement uncertainty. On the other hand, from the practical point of view, contributions are present which show how measurement uncertainty is considered in the specific applications. The aim of this book is to provide a wide overview on this very important topic.

Simona Salicone

Editor

New Frontiers in Measurement Uncertainty

Simona Salicone

Dipartimento di Elettronica, Informazione e Bioingegneria (DEIB), Politecnico di Milano, Building 7, Piazza Leonardo da Vinci 32, 20133 Milano, Italy; simona.salicone@polimi.it

1. Introduction

Metrology is the science of measurements. In our everyday life, we are constantly surrounded by measurements: from reading the time to weighing apples, we continuously measure something. However, measurements are also somehow embedded in objects, since, for example, the apple we buy has already been measured before its arrival at the greengrocer, in order to determine its calibre. In these measurements, uncertainty plays a very important rule. Metrologists know that no measurement makes sense without an associated uncertainty value. Without it, no decision can be taken; no comparisons can be made; no conformity can be assessed.

It is hence pivotal to know the meaning of measurement uncertainty, understand the contributions to measurement uncertainty, know how these contributions affect the final measurement uncertainty, have a mathematical tool to represent measurement uncertainty and propagate it through the measurement procedure, and consider measurement uncertainty in any application.

This Topical Collection “Measurement Uncertainty” started as a Special Issue, but many contributions have been submitted showing how metrology—and, in particular, measurement uncertainty—is an open, interesting, and important topic.

Therefore, with my great pleasure, the Special Issue has become a Topical Collection.

I invite Colleagues working on this issue to continue submitting papers, so that the Collection can grow and become a good place for a fruitful discussion.

2. Overview of Contributions

Following is a brief overview of the first ten contributions published in the Topical Collection.

The core of every contribution is represented by the “measurement uncertainty”, a concept introduced in 1995 by the “Guide to the expression of uncertainty in measurement”, generally known as GUM. The word “uncertainty” has a lexical meaning and reflects the lack of exact knowledge or lack of complete knowledge about something. Therefore, the value associated to a measured value (which should express the lack of exact knowledge about the value of the measurand) is called the “uncertainty value”. This value can be found, according to the suggestions of the GUM and following the mathematical probabilistic approaches therein proposed.

In the last decades, however, other methods have been proposed in the literature, which try to encompass the definitions of the GUM, while overcoming its limitations. Some of these methods are based on the possibility theory, such as the RFV (random-fuzzy variable) method. The authors of [1] briefly recall the RFV method, starting from the very beginning and the initial motivations and summarizing the most relevant obtained results.

Kalman filters, a concept that has been in existence for decades now, are widely used in numerous areas. The Kalman filter provides a prediction of the system states as well as the uncertainty associated to it. In [2], the same authors of [1] propose a new application of the RFV method on Kalman filters, with the specific aim of reducing the overall uncertainty associated to the state predictions. In particular, a possibilistic Kalman filter is defined,

Citation: Salicone, S. New Frontiers in Measurement Uncertainty. *Metrology* **2022**, *2*, 495–498. <https://doi.org/10.3390/metrology2040029>

Received: 28 October 2022
Accepted: 7 December 2022
Published: 12 December 2022

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

which uses random-fuzzy variables; not only does it consider and propagates both random and systematic contributions to uncertainty, but also reduces the overall uncertainty associated to the state predictions by compensating for the unknown residual systematic contributions.

In [3], measurement uncertainty is considered to be associated to measuring bridges for non-conventional instrument transformers with digital output. In this paper, the authors underline the necessity of synchronization between the analogue output and the digital one. They hence propose an ad hoc measurement setup that is able to monitor and quantify the main quantities of interest. The proposed measurement setup is laboratory implemented and the main sources of uncertainty are discussed and combined through a statistical analysis.

The authors of [4] yearn for a future scenario in which the digital reporting of measurement results is ubiquitous, and digital calibration certificates (DCCs) contain information about all the components of uncertainty in a measurement result. To show the benefits of this possible future scenario, the authors consider and compare the actual “international measurement comparisons” used by the International Committee for Weights and Measures (CIPM) and the regional metrology organization (RMO). They propose an uncertainty-number digital reporting format, which caters to all the information required and would simplify the comparison analysis, reporting, and linking; the format would also enable a more informative presentation of comparison results.

In [5], the authors deal with measurement uncertainty in Prompt Fission Neutron Spectra (PFNS) measurements, measurements of fission cross-sections, and measurements of Maxwellian spectrum-averaged neutron capture cross-sections for astrophysical applications. In particular, they demonstrate that these measurements are all subject to the presence of Systematic Distortion Factors (SDF). SDF may exist in any experiment: it leads to the bias of the measured value from an unknown “true” value. The SDF appears as a real physical effect if it is not removed with additional measurements or analysis. For a set of measured data with the best evaluated true value, their differences beyond their uncertainties can be explained by the presence of Unrecognized Source of Uncertainties (USU) in these data. The authors link the presence of USU in the data to the presence of SDF in the results of the measurements.

In [6], the topic of digital calibration certificates (DCC) is considered again. In calibration certificates, information about a quantity is frequently provided in the form of an estimate of the quantity and an associated standard or expanded uncertainty. Then, if the quantity must be used in another calculation, it is common—in the absence of any additional information—to assign a Gaussian probability distribution to the quantity. However, the true probability distribution of the quantity could be significantly different from the Gaussian one; therefore, this assignment may lead to unreliable results, when subsequent calculations are made. Even if the uncertainty evaluation has been made using a Monte Carlo simulation, only the summary information “estimate of the quantity and associated uncertainty” are generally reported in the calibration certificate, for the sake of brevity. Using two examples, the authors show how to present all the information derived from a Monte Carlo simulation in a fully machine-readable form and insert the whole information inside digital calibration certificates. In this way, no information is lost.

Technologies that can measure, analyse, and make critical decisions autonomously are beginning to emerge; hence, there is great interest in the digitalisation of metrology. In [7], the authors report on a Python package that implements algorithmic data processing using ‘uncertain numbers’, which satisfy the general requirements of the GUM for the expression of uncertainty. An uncertain number can represent a physical quantity that has not been exactly determined. Using uncertain numbers, measurement models can be expressed clearly and succinctly in terms of the quantities involved. The proposed algorithms provide an example of how metrological traceability can be supported in digital systems. In particular, uncertain numbers provide a format to capture and propagate

detailed information about quantities that influence a measurement along the various stages of a traceability chain.

One of the main challenges in designing information fusion systems is to decide on the structure and order in which information is aggregated. The key criteria by which topologies are constructed include the associativity of fusion rules as well as the consistency and redundancy of information sources. Fusion topologies regarding these criteria are flexible in design, produce maximal specific information, and are robust against unreliable or defective sources. In [8], an automated data-driven design approach for possibilistic information fusion topologies is detailed that explicitly considers associativity, consistency, and redundancy. The proposed design is intended to handle epistemic uncertainty and obtain robust topologies.

In [9], the authors analyse the measurement uncertainty associated to the evaluated frequencies of the spectral tones of signals created from superimposed sinusoids and white Gaussian noise, when different methods for the spectral analysis of the signals are applied. By comparing the obtained results, the authors draw some useful conclusions in order to guide a designer in choosing a method for the spectral analysis, according to the operating conditions.

Data-driven manufacturing in Industry 4.0 demands digital metrology to not only drive the in-process quality assurance of manufactured products, but also to supply reliable data to constantly adjust the manufacturing process parameters for zero-defect manufacturing processes. The better quality, improved productivity, and increased flexibility of manufacturing processes are obtained by combining intelligent production systems and advanced information technologies where in-process metrology plays a significant role. Today, the massive integration of 3D optical sensors occurs within manufacturing processes, replacing traditional Coordinate Measurement Machines (CMM) within the automotive, aerospace, and power generation industries. However, while the delivery of millions of points in a matter of seconds is assumed by 3D optical sensors, the process of automatically converting dense data into meaningful information and assuring the quality of these data remains a challenge. In [10], the authors present a practical approach to addressing both these challenges, based on ISO 15530-3 and ISO 15530-4 technical specifications and the application of MBD-based post-processing for the automatic processing of point clouds.

3. Conclusions

The above overview of the ten papers in this Topical Collection shows that measurement uncertainty is an important issue in many different fields. It can be stated that measurement uncertainty should be properly considered everywhere, in any industrial, economic, social, and legal application. My hope is for contributions from each field to be present in this Topical Collection.

Funding: This research received no external funding.

Data Availability Statement: Not applicable.

Conflicts of Interest: The author declares no conflict of interest.

References

1. Salicone, S.; Jetti, H.V. A general mathematical approach based on the possibility theory for handling measurement results and all uncertainties. *Metrology* **2021**, *1*, 76–92. [[CrossRef](#)]
2. Jetti, H.V.; Salicone, S. A possibilistic kalman filter for the reduction of the final measurement uncertainty, in presence of unknown systematic errors. *Metrology* **2021**, *1*, 39–51. [[CrossRef](#)]
3. Frigo, G.; Agustoni, M. Calibration of a digital current transformer measuring bridge: Metrological challenges and uncertainty contributions. *Metrology* **2021**, *1*, 93–106. [[CrossRef](#)]
4. 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. [[CrossRef](#)]
5. Kornilov, N.V.; Pronyaev, V.G.; Grimes, S.M. Systematic distortion factor and unrecognized source of uncertainties in nuclear data measurements and evaluations. *Metrology* **2022**, *2*, 1–18. [[CrossRef](#)]

6. Smith, I.; Luo, Y.; Hutzschenreuter, D. The storage within digital calibration certificates of uncertainty information obtained using a monte carlo method. *Metrology* **2022**, *2*, 33–45. [[CrossRef](#)]
7. 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. [[CrossRef](#)]
8. Holst, C.-A.; Lohweg, V. Designing possibilistic information fusion—The importance of associativity, consistency, and redundancy. *Metrology* **2022**, *2*, 180–215. [[CrossRef](#)]
9. Dello Iacono, S.; Di Leo, G.; Liguori, C.; Paciello, V. The obtainable uncertainty for the frequency evaluation of tones with different spectral analysis techniques. *Metrology* **2022**, *2*, 216–229. [[CrossRef](#)]
10. Kortaberria, G.; Mutilba, U.; Gomez, S.; Ahmed, B. Three-Dimensional point cloud task-specific uncertainty assessment based on ISO 15530-3 and ISO 15530-4 technical specifications and model-based definition strategy. *Metrology* **2022**, *2*, 394–413. [[CrossRef](#)]

Perspective

A General Mathematical Approach Based on the Possibility Theory for Handling Measurement Results and All Uncertainties

Simona Salicone ^{1,*} and Harsha Vardhana Jetti ²

¹ Dipartimento di Elettronica, Informazione e Bioingegneria (DEIB), Politecnico di Milano, Edificio 7 Piazza Leonardo da Vinci 32, 20133 Milano, Italy

² Department of Energy, Politecnico di Milano, Via Lambruschini 4, 20156 Milano, Italy; harshavardhana.jetti@polimi.it

* Correspondence: simona.salicone@polimi.it

Abstract: The concept of measurement uncertainty was introduced in the 1990s by the “Guide to the expression of uncertainty in measurement”, known as GUM. The word uncertainty has a lexical meaning and reflects the lack of exact knowledge or lack of complete knowledge about the value of the measurand. Thanks to the suggestions in the GUM and following the mathematical probabilistic approaches therein proposed, an uncertainty value can be found and be associated to the measured value. In the last decades, however, other methods have been proposed in the literature, which try to encompass the definitions of the GUM, thus overcoming its limitations. Some of these methods are based on the possibility theory, such as the one known as the RFV method. The aim of this paper is to briefly recall the RFV method, starting from the very beginning and the initial motivations, and summarize in a unique paper the most relevant obtained results.

Keywords: measurement uncertainty; random contribution; systematic contribution; probability density functions; possibility distributions; random-fuzzy variables; t-norms

Citation: Salicone, S.; Jetti, H.V. A General Mathematical Approach Based on the Possibility Theory for Handling Measurement Results and All Uncertainties. *Metrology* **2021**, *1*, 76–92. <https://doi.org/10.3390/metrology1020006>

Academic Editor: Han Haitjema

Received: 9 August 2021

Accepted: 30 August 2021

Published: 1 October 2021

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Background: The Concept of Measurement Uncertainty

In the 1990s, the “Guide to the expression of uncertainty in measurements”, known as GUM, introduced the concept of measurement uncertainty and provided some guidelines for its representation and propagation through the measurement function. In particular, the measurement uncertainty is defined as “a parameter, associated with the result of a measurement, that characterizes the dispersion of the values that could reasonably be attributed to the measurand” [1], as also recalled in [2].

This definition refers to a “dispersion of the values” because, as is widely known, when a quantity (the measurand) is measured more times, the measurement result generally varies, due to different contributions affecting the measurement procedure. This means that, because of the “dispersion of the values”, from a strict metrological point of view, the “true value” of the measurand cannot be known.

The uncertainty associated with a measured value has, therefore, the aim to provide information about how large this “dispersion of the values” is [1,2].

Therefore, from a strictly semantic point of view, it can be stated that the uncertainty value reflects the lack of exact knowledge or lack of complete knowledge about the value of the measurand. Hence, when one speaks about a measurement result, one always speaks about an incomplete information; this incomplete information must be somehow represented to provide validity of the measured value.

How can this representation be done? According to the GUM, the aim of the uncertainty evaluation is “to provide an interval about the measurement result that may be expected to encompass a large fraction of the distribution of values that could reasonably be attributed to the

quantity subject to measurement" [1]. Furthermore, it clearly states that "the ideal method for evaluating and expressing measurement uncertainty should be capable of readily providing such an interval, in particular, one with a coverage probability or level of confidence that corresponds in a realistic way to that required" [1].

As stated above, the "dispersion of the values" is due to different contributions affecting the measurement procedure. In particular, in the "International vocabulary of metrology", known as VIM [3], two contributions are defined: the random and the systematic contributions to uncertainty. (There is sometimes the mistake that the words random and systematic are substituted by the words "type A" and "type B", defined in the GUM, respectively. However, "type A" and "type B" refer to methods of evaluation of the uncertainty contribution and not explicitly to the nature of the uncertainty contribution itself.)

The random contribution is defined as the "component of measurement error that in replicate measurements varies in an unpredictable manner" [3], while the systematic one is defined as the "Component of measurement error that in replicate measurements remains constant or varies in a predictable manner" [3]. Therefore, due to the random contributions to uncertainty, the dispersion of the measured values may define an interval around the mean (of the measured values), and this interval might be indeed the "interval about the measurement result that may be expected to encompass a large fraction of the distribution of values that could reasonably be attributed to the quantity subject to measurement" required by the GUM [1]. In Figure 1, the blue dot represents the measurand value, while the pink asterisk is the mean of the measured values, and the purple line represents the interval that includes all the measured values. It can be easily seen that the purple interval also encompasses the measurand value, as generally happens if a proper coverage factor is applied. However, if a systematic contribution also affects the measurement result, then the interval that includes all the measured values is shifted on the right/left with respect to the previous interval. The direction of right or left depends on whether the systematic effect is positive or negative, as shown with the blue and orange intervals in Figure 1. It can be easily seen that these last intervals no longer represent the "interval about the measurement result that may be expected to encompass a large fraction of the distribution of values that could reasonably be attributed to the quantity subject to measurement" since the measurand value is completely outside these intervals.

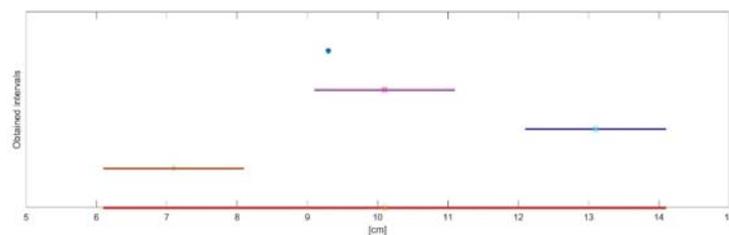


Figure 1. The effects of random and systematic contributions to uncertainty. Blue dot: unknown value of the measurand. Purple line: dispersion of the values and obtained interval when only random contributions affect the measured values. Blue and orange lines: obtained interval when a positive or negative systematic error affect the measured values. Red line: obtained interval when the effects of both random and unknown uncompensated systematic contributions are considered.

In the case that one wants to provide the interval, taking into account both the random and the systematic contributions to uncertainty, he/she should consider also the possible variability of the effect of the systematic contributions and, hence, should widen the uncertainty interval, as shown by the red line in Figure 1. Therefore, the purple interval is the uncertainty interval when only random contributions affect the measurement result, while the red interval is the uncertainty interval when systematic contributions also affect the measurement result.

The GUM states that “It is assumed that the result of a measurement has been corrected for all recognized significant systematic effects and that every effort has been made to identify such effects” [1]; in other words, the GUM requires that all efforts be made to identify, measure and correct for all the significant systematic effects. Under this assumption, only the random effects are present, and the uncertainty interval is reduced as shown in Figure 1.

2. The Authors’ Point of View

In the previous section, it is summarized the concept of measurement uncertainty, and it is recalled that the requirement of the GUM is that all the significant systematic effects are identified and compensated for. Satisfying this leads to the following important conclusions:

- The reduction in the overall uncertainty and hence, the reduction in the uncertainty interval.
- Only random contributions affect the measurement procedure, and therefore, the uncertainty contributions can be mathematically considered to be random variables and represented with probability density functions (pdf).

There are also some mathematical ways to also treat the systematic contributions to uncertainty in the mathematical framework of the probability theory, such as, for instance, a proper use of the correlation coefficients, but, in any case, the probability theory is born to handle the random phenomena and can correctly handle only random phenomena because of the way that pdfs combine with each other.

Furthermore, the GUM requires the compensation of the “significant systematic effects” [1] where the word “significant” is very important, bringing a crucial question: when is an effect (on the final measured result) significant?

Obviously, an effect can be significant in one topic and not significant in another. From the metrological point of view, the “significance” can be exploited by considering the “target uncertainty”, which is defined by the VIM as the “*measurement uncertainty specified as an upper limit and decided on the basis of the intended use of measurement results*” [3]. The target uncertainty is, therefore, a value that depends on the topic: it is a number that is generally as small as possible in primary metrology or in the industrial world in the limited case in which very precious objects are measured (such as diamonds, for instance). However, in most practical industrial situations, the target uncertainty is a trade-off between the cost of the uncertainty evaluation and the waste production; therefore, there is no need to set the target uncertainty to be as small as possible. In these situations, the correction for the “significant systematic effects” is generally not necessary for not exceeding the target uncertainty. Therefore, the industrial world is generally not interested in reducing the overall uncertainty by identifying and compensating for systematic effects.

In any case, compensation or not, to state whether a systematic effect is significant or not, it must be considered in the uncertainty evaluation. It becomes, therefore, an important issue to be able to mathematically determine the overall uncertainty in the best possible way.

Methods that employ a mathematical theory different from the probabilistic theory encompassed by the GUM have been proposed in the literature [4–8]. These methods are based on the possibility theory, as well as the RFV method recalled in this paper, which tries to encompass the definitions of the GUM, thereby overcoming its limitations.

The RFV method recalled in this paper can handle both random and systematic contributions to uncertainty in closed form. This is possible because, in this mathematical framework, many operators between the variables naturally defined in it are available. Therefore, different operators can be chosen, which can simulate the combination of the variables in a random or a nonrandom way. To introduce this method, the theory of evidence is shortly recalled in the next section, with the aim to provide a cornerstone to the method, rather than giving the mathematical details, for which the readers are referred to [9–11].

3. Shafer's Theory of Evidence

The mathematical theory of evidence was defined by Glen Shafer in the 1970s to generalize the probability theory [9]. In particular, if probability functions are considered, they obey the additivity rule, so that the following holds:

$$Pro(U) + Pro(\bar{U}) = 1 \tag{1}$$

where U and \bar{U} are complementary sets.

However, in Shafer's (and also the authors') opinion, the additivity rule is not able to handle correctly all possible situations of knowledge/unknowledge. Therefore, he generalizes this rule, and to do this, he defines the belief functions Bel , for which the superadditivity rule applies:

$$Bel(U) + Bel(\bar{U}) \leq 1 \tag{2}$$

Given a certain statement A , the degree of belief $Bel(A)$ is a judgment. This means that, given A , different individuals with different levels of expertise regarding A might provide different judgments. In his book, Shafer writes explicitly:

"Whenever I write of the 'degree of belief' that an individual accords to the proposition, I picture in my mind an act of judgment. I do not pretend that there exists an objective relation between given evidence and a given proposition that determines a precise numerical degree of support. Rather, I merely suppose that an individual can make a judgment ... he can announce a number that represents the degree to which he judges that evidence supports a given proposition and, hence, the degree of belief he wishes to accord the proposition" [9]

In his book, Shafer also provides two examples to show that belief functions are more suitable to handle knowledge/unknowledge with respect to probability functions: the example of the Ming vase and the example of Sirius star are here briefly recalled.

3.1. The Ming Vase

A person is shown a Chinese vase and is asked whether the vase is a real vase of the Ming dynasty or a counterfeit. Sets A and B are assigned to the two possibilities, as shown in Table 1.

Table 1. The Chinese vase and the two considered sets.

	Case	Event
	A	The vase is a real Ming vase.
	B	The vase is a counterfeit.

Of course, looking at the vase, there could be different situations that also depend on the interviewed person, i.e., whether the person is an expert or not:

1. The evidence suggests the authenticity of the vase.
2. The evidence suggests that the vase is a counterfeit.
3. Some evidence suggests the authenticity, while other evidence, the counterfeit:
 - a. Substantial evidence on both sides.
 - b. Little evidence on both sides.
4. The observer is not an expert and has no evidence to say whether the vase is true or false.

Let us now consider how these different situations can be handled with the probability and the belief functions.

In the first two cases, the same numerical values are given to both the probability and the belief functions (as shown in the first two cases of Table 2) since probably the interviewed person is an expert, and hence can recognize whether the vase is true or false.

Table 2. Assignments given to probability and belief functions in the considered cases.

Case	Pro		Bel	
1	$Pro(A) = 1$	$Pro(B) = 0$	$Bel(A) = 1$	$Bel(B) = 0$
2	$Pro(A) = 0$	$Pro(B) = 1$	$Bel(A) = 0$	$Bel(B) = 1$
3A	$Pro(A) = 0.5$	$Pro(B) = 0.5$	$Bel(A) = 0.5$	$Bel(B) = 0.5$
	or $Pro(A) = 0.6$	$Pro(B) = 0.4$	$Bel(A) = 0.34$	or $Bel(B) = 0.4$
3B	$Pro(A) = 0.5$	$Pro(B) = 0.5$	$Bel(A) = 0.1$	$Bel(B) = 0.2$
	or $Pro(A) = 0.6$	$Pro(B) = 0.4$		
4	$Pro(A) = 0.5$	$Pro(B) = 0.5$	$Bel(A) = 0$	$Bel(B) = 0$

On the other hand, the other two situations are treated in a different way by the probability and the belief functions since probability functions must obey the additivity rule, while belief functions need not.

Therefore, when cases 3A and 3B are considered, probability functions can take the values, for instance, given in Table 2, but no lower values can be assigned, even if little evidence is present on both *A* and *B*. On the other hand, when belief functions are considered, the person can indicate two numbers, which more precisely represent his/her idea about *A* and *B*.

In case 3A, it may happen that the same numbers are assigned to probability and belief functions (according to the degree of belief about *A* and *B*), but it may also happen that different numbers are assigned since, for belief functions, it is not necessary to satisfy the additivity rule (see Table 2). Furthermore, in case 3B, where there is little evidence on both sides, it is not possible to assign a small number to both *A* and *B* with probability functions, while this can be done with belief functions (see Table 2).

The different behavior of the probability and the belief functions is even more emphasized when Case 4 is considered, where the person is not an expert and therefore declares his/her ignorance about the vase. This is the classical situation, called, by Shafer, *total ignorance*, in which a zero value is assigned to all possible sets (and a unitary value is assigned only to the entire universal set, which include all possibilities). Therefore, as shown in Table 2, $Bel(A) = 0$ and $Bel(B) = 0$ in the case of total ignorance (Case 4). The probability functions, on the other side, must always obey the additivity rule, and therefore, even in the case of total ignorance (as in the case of equal evidence on both *A* and *B*) $Pro(A) = 0.5$ and $Pro(B) = 0.5$ are assigned, not to give preference to either *A* or *B*.

Total ignorance is, therefore, treated in a completely different way by the probability and the belief functions; an interesting question is determining which method is the better one. It seems that the belief functions are more suitable to represent total ignorance at least for two reasons. First, with probability functions, it is not possible to distinguish the two different cases where there is an equal degree of belief on both cases *A* and *B*, and there is no evidence about either *A* or *B*. In fact, in both these cases, $Pro(A) = 0.5$ and $Pro(B) = 0.5$ must be assigned. Second, probability functions may lead to incongruent results when more than two sets are considered, as in the following example of the Sirius star [9].

3.2. The Sirius Star

Are there or are there not living beings on the planet in orbit around star Sirius? Let us only consider the case where the interviewed person is not an expert at all, so the case of Shafer’s total ignorance, and let us consider the two different situations given in Tables 3 and 4. In the first case, total ignorance is admitted on only two sets, while in the second case, total ignorance is professed on three sets, and the two ways of forming the sets are independent.

Table 3. The Sirius star and the two considered sets.

	Case	Event
	A	There is life.
	B	There is not life.

Table 4. The Sirius star and the three considered sets.

	Case	Event
	C	There is life.
	D	There are planets but not life.
	E	There are not even planets.

Table 5 shows the values assigned to the belief function for the sets defined in Table 3 (first column) and for the sets defined in Table 4 (second column). It is, however, possible to compare the two cases since, by considering the sets defined in Tables 3 and 4, it can be stated that $A = C$ and $B = D \cup E$. The last column is the comparison of the two previous columns and shows that the assigned values in the two cases are coherent with each other.

Table 5. The Sirius star and total ignorance represented with the belief functions.

Case of Table 3		Case of Table 4		Comparison
A	$Bel(A) = 0$	C	$Bel(C) = 0$	$Bel(A) = Bel(C)$
B	$Bel(B) = 0$	D	$Bel(D) = 0$	$Bel(D \cup E) = Bel(B)$
		E	$Bel(E) = 0$	

On the other hand, Table 6 shows the results for the probability functions and, when the two cases of Tables 3 and 4 are compared, it follows that there is no consistency at all. In fact, set A defined in the case of Table 3 is exactly set C defined in the case of Table 4, but as shown in Table 6, $Pro(A) \neq Pro(C)$. Furthermore, set B defined in the case of Table 3 is exactly set $D \cup E$ defined in the case of Table 4, but $Pro(D \cup E) \neq Pro(B)$ since the following holds:

$$Pro(D \cup E) = Pro(D) + Pro(E) - Pro(D \cap E) = \frac{1}{3} + \frac{1}{3} - 0 = \frac{2}{3} \neq Pro(B) = \frac{1}{2}$$

Table 6. The Sirius star and total ignorance represented with the probability functions.

Case of Table 3		Case of Table 4		Comparison
A	$Pro(A) = \frac{1}{2}$	C	$Pro(C) = \frac{1}{3}$	$Pro(A) \neq Pro(C)$
B	$Pro(B) = \frac{1}{2}$	D	$Pro(D) = \frac{1}{3}$	$Pro(D \cup E) \neq Pro(B)$
		E	$Pro(E) = \frac{1}{3}$	

Then, it can be concluded that belief functions are more suitable than probability functions to handle total ignorance, that is, all situations where an individual has no evidence/no knowledge about the considered topic and about the considered given sets.

This great interest in total ignorance is due to the fact that total ignorance is mostly present in the field of measurements, as shown in the simple practical example in the next section.

4. Total Ignorance in Measurements

Let us here consider a simple example to show how, in measurement procedures, the situation called total ignorance by Shafer is very often present.

A calibrator provides a reference voltage of 24 V, and some multimeters of the same typology (4 1/2 Leader 856) are employed to measure this voltage. The instrument data sheet provides the measuring accuracy as ± % of reading ± number of digits, and the value of each digit is given by the resolution in the considered range. According to the data sheet, Table 7 provides the resolution and the measuring accuracy in the different ranges. For the measurand $V_x = 24$ V, which is the reference voltage in the proposed example, the range is 30 V and therefore, according to the specifications, the measurement accuracy is $\pm(0.05\% V_x + 2 \text{ mV}) = \pm 0.014$ V.

Table 7. The multimeter data sheet.

Multimeter 4 1/2 LEADER 856			
Range	Full Scale	Resolution	Measuring Accuracy
300 mV	29,999 mV	0.01 mV	$\pm(0.03\% V_x + 0.02 \text{ mV})$
3 V	29,999 V	0.1 mV	$\pm(0.05\% V_x + 0.2 \text{ mV})$
30 V	29,999 V	1 mV	$\pm(0.05\% V_x + 2 \text{ mV})$
300 V	29,999 V	10 mV	$\pm(0.05\% V_x + 20 \text{ mV})$
1000 V	10,000 V	0.1 V	$\pm(0.05\% V_x + 0.2 \text{ V})$

Two different measurement procedures are considered:

1. All multimeters are employed to measure the reference voltage.
2. Only one multimeter is employed to measure the reference voltage.

Figure 2 shows, with the orange line the reference voltage and with the pink crosses the value measured by 10 different multimeters. Since different instruments are employed, it is likely to happen that the measured values fall around the reference value. In this situation, it could be possible to apply a probabilistic approach by considering the following: the mean of the measured values; an uncertainty interval around the evaluated mean, and a pdf over this interval (but only if a high number of different instruments are employed).

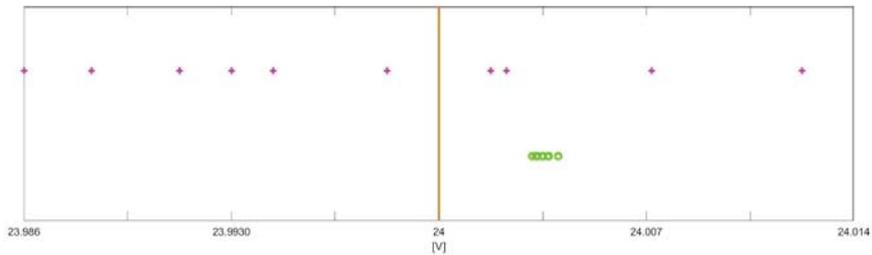


Figure 2. Up: values measured by different multimeters. Down: values measured by a single multimeter.

This situation represents very well the calibration procedure that is performed by the instruments' manufacturer to provide the accuracy interval, which reflects the behavior of all instruments of the same typology. However, this situation is very seldom met in practice, because generally only one instrument is available and employed. Under this more common situation, when only one multimeter is employed, the value measured by the multimeter will be shifted with respect to the reference value. Moreover, if different measurements were taken, this would not help to better estimate the reference value since all measured values would be shifted more or less the same amount with respect to the reference value, as shown by the green circles in Figure 2. In fact, all measured values are taken, in this case, by the same instrument and, therefore, are affected by the same systematic error, even if a small variation can be observed, due to the presence of also random phenomena.

In this last case, even if the mean of the measured values is taken, no better estimate of the measurand can be obtained. Additionally, even if an interval is built, according to the dispersion of the measured values, this interval would not contain the value of the measurand. Therefore, to provide a good uncertainty interval, it is necessary to refer to the accuracy interval provided by the data sheet. The data sheet does not provide any pdf associated with this interval, and therefore, no pdf can be assigned to the obtained interval.

When we have a pdf over a given support, it is possible to assign a confidence interval (or degree of belief) to any subintervals of the support. However, when no pdf is assigned and no knowledge is available to assign a specific pdf, it is not possible to associate any confidence interval (or degree of belief) to any subintervals of the support. We are, therefore, perfectly in the case of Shafer's total ignorance, where a degree of belief can be assigned to the support (or universal set), but no degree of belief can be assigned to the subintervals (to the subsets of the universal set).

It clearly follows that total ignorance is present in the measurement field. Since belief functions better represent total ignorance, it is worth exploring these functions and the theory of evidence to find an alternative, more general way to handle measurement uncertainty and measurement results. It is not the aim of this paper to provide all definitions and mathematical details, for which the readers are referred to the published literature [10–14]. The next section will, therefore, give only some introduction to come to the possibility distributions (PD) and the random-fuzzy variables (RFV).

5. The Random-Fuzzy Variables

In the previous sections, belief functions are introduced and it is shown how they can suitably represent the available knowledge, including total ignorance. It is interesting to observe that belief functions are a generalization of the probability functions and the necessity functions. In this respect, it is first necessary to know what a focal element is.

Let us first define the basic probability assignment function:

$$\begin{aligned}
 m : P(X) &\rightarrow [0, 1] \\
 m(\emptyset) &= 0 \\
 \sum_{A \in P(X)} m(A) &= 1
 \end{aligned}
 \tag{3}$$

where X is the universal set, $P(X)$ is the power set of X and \emptyset is the empty set. According to (3), $m(A)$ represents the degree of belief that an element x belongs to set A (only to set A and not to its subsets).

Set A for which $m(A) > 0$ is called the focal elements of X . When the focal elements are singletons, then it can be proved [9–12] that belief functions are probability functions, and the theory of evidence enters in the particular case of the probability theory. This shows that the belief functions are, as wanted by Shafer, a generalization of the probability functions. However, it is also interesting to consider another particular case of belief functions, which are called necessity functions and are obtained when the focal elements are all nested, as shown in Figure 3.

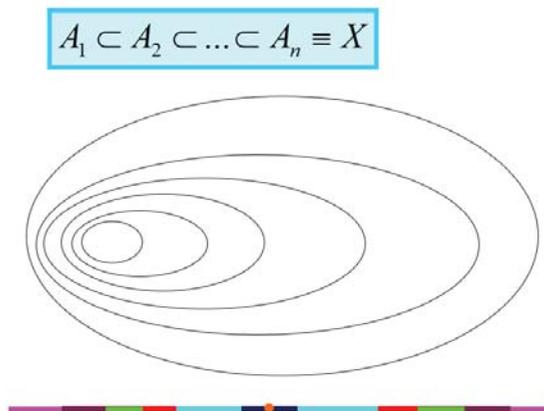


Figure 3. Example of nested focal elements, for sets and for intervals.

The upper plot in Figure 3 clearly shows that, when sets are considered, all sets can be ordered in such a way that $A_1 \subset A_2 \subset \dots \subset A_n \equiv X$. When, instead of sets, intervals are considered, the lower plot can be drawn, which still satisfies $A_1 \subset A_2 \subset \dots \subset A_n \equiv X$. This case is very interesting from the metrological point of view because there could be an analogy between these nested intervals and the confidence intervals of a given pdf at different, increasing levels of confidence.

The necessity function is defined as follows:

$$Nec(A_j) = \sum_{k=1}^j m(A_k)$$

and represents the degree of belief that an element x belongs to set A and to all its subsets. When the belief functions are necessity functions, then the theory of evidence enters the particular case of the possibility theory.

In the same way that probability density functions are defined in probability, possibility distribution functions (PD) are defined in possibility as follows:

$$r : X \rightarrow [0, 1]$$

where:

$$\max(r(x)) = 1$$

when $x \in X$.

It can be proved [10,11] that the nested intervals of Figure 3, together with their corresponding necessity functions $Nec(A_j)$, represent confidence intervals at specific levels of confidence, coverage probability, or degree of belief $Nec(A_j)$. Therefore, remembering the GUM words that “the ideal method for evaluating and expressing measurement uncertainty should be capable of readily providing such an interval, in particular, one with a coverage probability or level of confidence that corresponds in a realistic way to that required” [1], it can be stated that the possibility theory, which provides all confidence intervals at all confidence levels, is perfectly GUM compliant.

If the intervals of Figure 3 are not overlapped with each other but are positioned at different vertical levels α , such as $\alpha_j = 1 - Nec(A_j)$, then a fuzzy variable is obtained, as in the example in Figure 4.

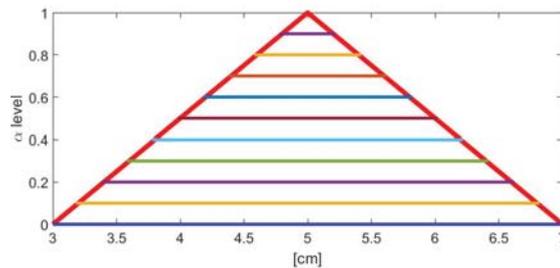


Figure 4. Example of possibility distributions and confidence intervals.

The fuzzy variable is commonly defined by its membership function which is, from the strict mathematical point of view, a PD.

Since a fuzzy variable (a PD) represents confidence intervals at all levels of confidence, a fuzzy variable can be used to represent in a very immediate way the result of a measurement [10–12]. Moreover, since different kinds of uncertainty contributions may affect the measurement procedure, the best way to represent the result of a measurement is the use of a fuzzy variable of type 2 and, in particular, a random-fuzzy variable (RFV). An RFV provides two PDs and can, hence, represent separately the effects on the measurement result of the different contributions to uncertainty. An example of RFV is given in Figure 5, with the red and violet lines. In an RFV, the uncompensated systematic contributions are represented by the internal PD $r_{int}(x)$ (violet line), while the random contributions are represented by the random PD $r_{ran}(x)$ (green line). The external PD $r_{ext}(x)$ (red line) is obtained by the combination of the two PDs $r_{int}(x)$ and $r_{ran}(x)$ [10–13].

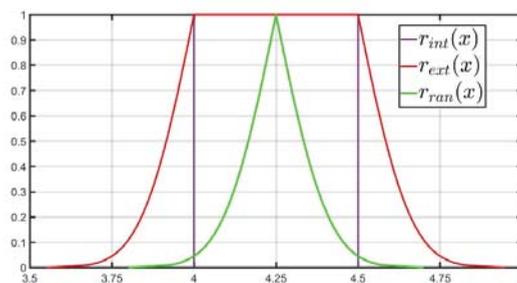


Figure 5. Example of RFV.

Extending the considerations made for the fuzzy variables, it can be stated that the cuts X_α at levels α of the RFV are the confidence intervals associated to the measurement result at the confidence levels $Nec(X_\alpha) = 1 - \alpha$ (as shown in Figure 6). In particular, the internal interval of each confidence interval is due to the effect on the measured value of the systematic contributions to uncertainty, while the external intervals are due to the effect of the random contributions.

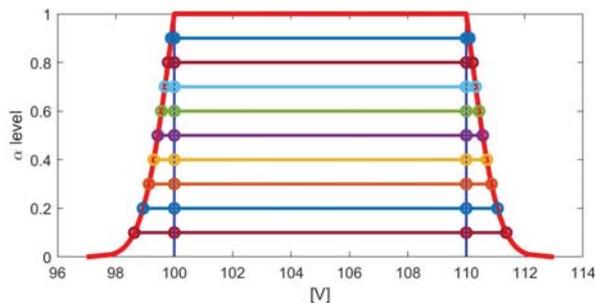


Figure 6. Example of RFV and some of the confidence intervals, which show the systematic effects and the random effects on the final measured result.

If RFVs can suitably represent measurement results, then it is important to understand how an RFV can be built and how two RFVs can be combined with each other, as will be briefly explained below; we refer the readers to the literature for more details [10–14].

5.1. RFV Construction

To build an RFV, it is necessary to define the shape of the PDs $r_{int}(x)$ and $r_{ran}(x)$, whose construction is different [10–14] since they represent different kinds of contributions.

As far as $r_{int}(x)$ is concerned, this PD represents the uncompensated systematic contributions to uncertainty. As shown in the example of the multimeter in the previous Section 4, generally, the only available knowledge is, in this case, the accuracy interval given by the manufacturer of the employed instrument in the data sheet. Therefore, the available knowledge can be represented by Shafer’s total ignorance. As is also shown in Section 3, total ignorance is mathematically represented by the belief function [9–11]:

$$Bel(X) = 1$$

$$Bel(A) = 0 \quad \forall A \subset X$$

and by the rectangular PD, such as the one in violet line in Figure 5. It follows that $r_{int}(x)$ is rectangular in most situations, even if situations may exist that could lead to different shapes [10–14].

On the other hand, $r_{ran}(x)$ must represent the random contributions to uncertainty and therefore, in most cases, a pdf is known or can be supposed. In this case, the corresponding PD can be easily obtained by applying the suitable probability–possibility transformation (different probability–possibility transformations are available in the literature to transform pdfs into PDs. The suitable transformation when PDs are used to represent measurement results is the maximally specific probability–possibility transformation, which preserves all confidence intervals and corresponding confidence levels) [10,15].

As an example, when the pdf is uniform, then the corresponding PD is triangular; when the pdf is triangular, then the corresponding PD is the orange one in Figure 7; when the pdf is Gaussian, then the corresponding PD is the blue one in Figure 7.

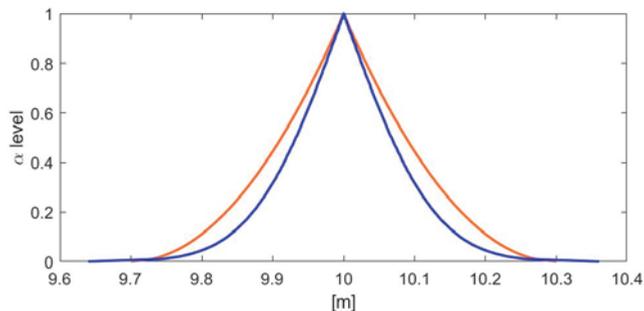


Figure 7. Example of PDs coming from given pdfs. Blue line: PD from a Gaussian pdf. Orange line: PD from a triangular pdf.

5.2. RFV Combination

When the measurement results are represented by RFVs and they must be combined, it is possible to take into account all the available metrological information about the nature of the contributions to be combined and the way these contributions combine in the specific measurement procedure. According to that, since PDs can be combined using many different mathematical operators, the most proper one can be chosen.

Without entering the details, for which the readers are referred to [10,16,17], it can be stated that the random contributions to uncertainty always compensate with each other during the combination, and therefore, an operator that simulates this typical probabilistic compensation should be chosen. On the other hand, the systematic contributions to uncertainty could compensate or not with each other during the combination, according to the specific contributions and the specific measurement procedure. Therefore, there should be the possibility to choose between a mathematical operator that simulates compensation and another one which does not compensate.

Let us first consider the evaluation of the joint PD, starting from two PDs. As an example, Figure 8 shows the results obtained by combining the same two PDs with the use of two different t-norms (for the definition of the mathematical t-norms, the readers are addressed to [15]): the min t-norm (on the left) and the Frank t-norm (on the right). In the upper plots, the two-dimensional joint PDs are shown, while in the lower plots, the corresponding α -cuts are shown. It can be easily seen how compensation applies when the Frank t-norm is employed, while no compensation applies when the min t-norm is employed.

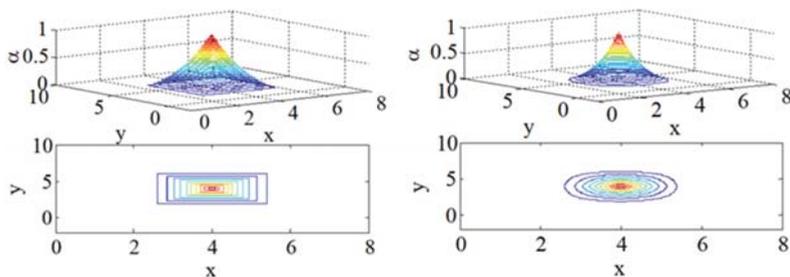


Figure 8. Combination of uncorrelated contributions. The same initial PDs are considered on both the left and right figures. On the left, the min t-norm is applied; on the right the Frank t-norm is applied. In the upper plots, the joint possibility distributions are shown. In the lower plots, the corresponding α -cuts are shown.

Figure 8 refers to the combination of uncorrelated contribution. Without entering the details, the correlation can also be considered, as shown, as an example, in Figures 9 and 10.

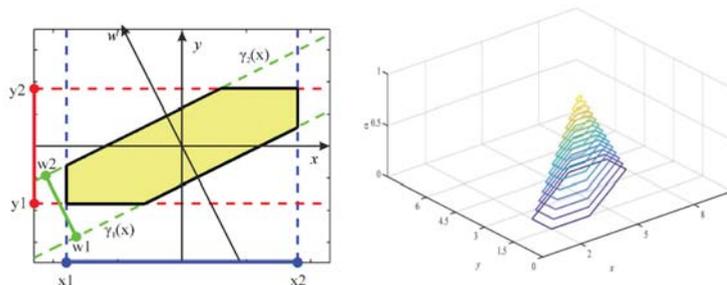


Figure 9. Combination of correlated contributions when the min t-norm is applied. The same initial PDs as in Figure 8 are considered. **Right:** joint PD. **Left:** shape of its α -cuts.

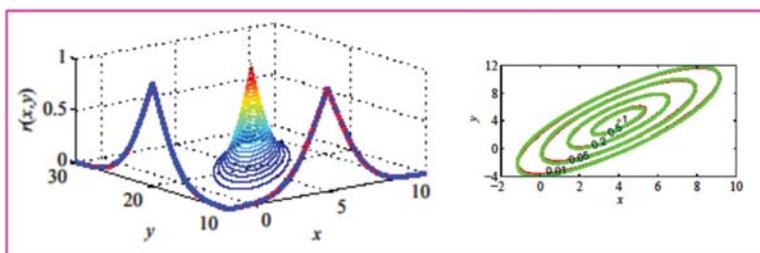


Figure 10. Combination of correlated contributions when the Frank t-norm is applied. The same initial PDs as in Figure 8 are considered. **Left:** joint PD. **Right:** its α -cuts.

From Figures 9 and 10, it can be easily seen how correlation modifies the joint PDs and the corresponding α -cuts.

Once the joint PDs $r_{ran}(x, y)$ and $r_{int}(x, y)$ are obtained, it is possible to evaluate the joint PD $r_{ext}(x, y)$ and the final RFV (this is obtained by applying the famous Zadeh extension principle. The readers are referred to [9,10] for the details) [9,10,16,17].

6. Example

To show the potentiality of the RFV approach, a simple example is here reported, where the RFV approach is compared with the GUM approach [1] and the Monte Carlo approach, as suggested by [18].

The GUM approach consists of the application of the law of propagation of the uncertainty [1], while random and systematic contributions to uncertainty are combined applying the quadratic law. The results given by the GUM approach are provided in terms of two specific confidence intervals: the ones at coverage probabilities 95.45% and 68.27%. These intervals are compared with the corresponding α -cuts at the same level of confidence of the RFVs obtained with the RFV approach.

The Monte Carlo approach consists of taking extractions from the given pdfs (in a way to agree with the available information) and combining the extractions to obtain a final histogram. Then, the histogram is converted in a pdf, and the pdf is converted to a PD (through the probability–possibility transformation mentioned above) for an immediate comparison with the RFVs given by the RFV approach.

Let us come to the example. A teacher measures the length and width of her desk with a wooden ruler and evaluates the area of the desk. She/he also asks her/his pupils to take the same measurements (and the area evaluation) with measuring tapes that they have built with some white cloth and a pencil to mark the cloth every half centimeter. The measurements are taken under different assumptions about both the measurement procedure and the uncertainty contributions, as shown in Table 8.

Table 8. The considered case studies.

Case	Procedure	Random	Systematic
A	Known measuring tape	uniform <i>pdf</i> ± 0.25 cm	Compensated
B	1 unknown measuring tape	uniform <i>pdf</i> ± 0.25 cm	uniform <i>pdf</i> ± 0.5 cm
C	2 unknown measuring tapes	uniform <i>pdf</i> ± 0.25 cm	uniform <i>pdf</i> ± 0.5 cm
D	1 unknown measuring tape	uniform <i>pdf</i> ± 0.25 cm	Interval ± 0.5 cm
E	2 unknown measuring tapes	uniform <i>pdf</i> ± 0.25 cm	Interval ± 0.5 cm

As far as the procedures are concerned, “Known measuring tape” means that the measuring tapes are somehow characterized, and therefore, the systematic error introduced by each of them is known; since the pupil uses their own tape, the systematic error is known and can be compensated. “1 unknown measuring tape” means that both length and width are measured with the same tape taken randomly among the tapes; the systematic error introduced by the tape is not known and it cannot be compensated but, since the same tape is used for the two measurements, the two measurements are correlated with each other. “2 unknown measuring tapes” means that length and width are measured with two different tapes taken randomly among the tapes; the systematic errors introduced by the tapes are not known and they cannot be compensated and since two different tapes are used for the two measurements, and therefore, the two measurements are uncorrelated with each other.

As far as the uncertainty contributions are concerned, the random contributions are supposed to be uniformly distributed; the systematic contributions are compensated (case A), uniformly distributed (case B and C) or without any other knowledge rather than the given interval (case D and E), as in Shafer’s total ignorance situation.

The uncertainty contributions reported in Table 8 are related to the pupils’ measuring tapes, while no uncertainty is assumed to affect the teacher’s measurements, realized with the wooden ruler, so that the teacher’s measured values are considered to be the reference values $l_{ref} = 90$ cm for the length and $w_{ref} = 60$ cm for the width, while $A_{ref} = 5400$ cm² is the reference area.

This means that, when the Monte Carlo approach is followed, extractions from the given pdfs in Table 8 are considered; when the RFV method is applied, the given pdfs in Table 8 are transformed into the corresponding PDs by applying the probability–possibility transformation; when the GUM approach is followed, the standard uncertainties are derived from the given pdfs in Table 8, that is, since the pdfs are uniform, the standard uncertainties are equal to the semi-width of the support of the pdfs divided by a factor $\sqrt{3}$ [1,2].

Without entering the details, for which the readers are referred to [10], the obtained results are shown in the following Figures 11–13. When only random contributions to uncertainty are present because the systematic ones are compensated for, the three approaches provide exactly the same results, showing the validity of the RFV method in simulating the presence of the random contributions. When both random and systematic contributions are present and their associated pdfs are known, the GUM approach underestimates the final measuring uncertainty, while the RFV and the Monte Carlo approaches provide very similar results. In this case, the RFV approach has the advantages of being faster and distinguishing, in the final measurement result, the effects due to the two different kinds of contributions. Finally, in the case of total ignorance, neither the GUM or the Monte Carlo approach can represent it in a different way with respect to cases B and C; therefore, they provide incorrect results.

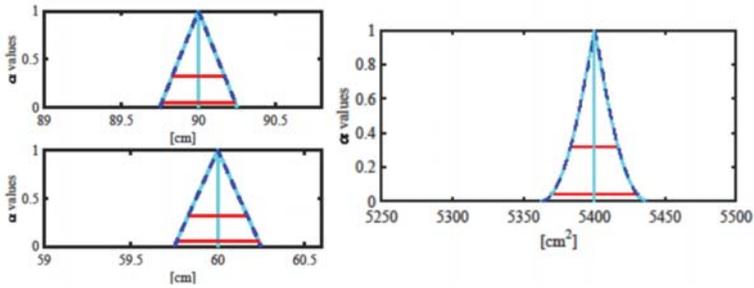


Figure 11. Obtained results when case study A is considered: GUM approach (red lines), Monte Carlo approach (blue lines), RFV approach (cyan lines). On the left: the length (**upper plot**) and the width (**lower plot**). On the right: the evaluated area of the desk.

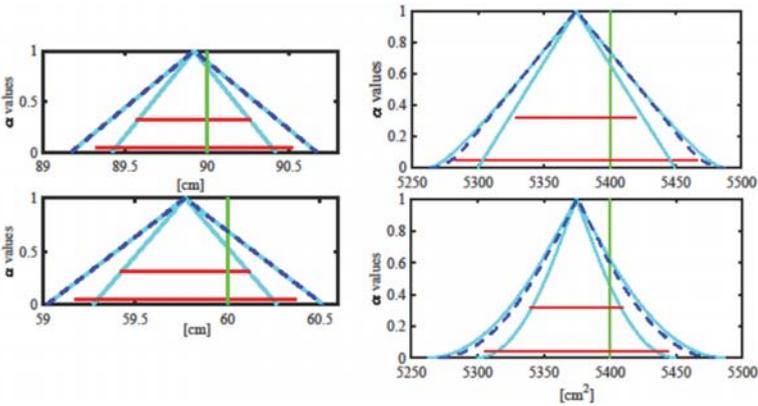


Figure 12. Obtained results when case studies B and C are considered: GUM approach (red lines), Monte Carlo approach (blue lines), RFV approach (cyan lines). On the left: the length (**upper plot**) and the width (**lower plot**). On the right: the evaluated area of the desk in case studies B (**upper plot**) and C (**lower plot**).

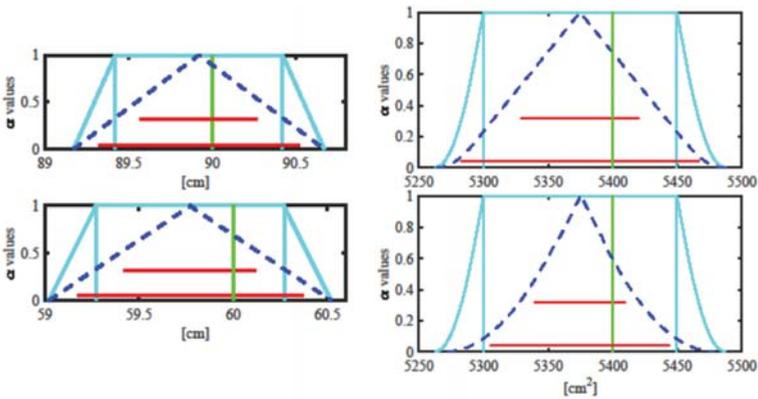


Figure 13. Obtained results when case studies D and E are considered: GUM approach (red lines), Monte Carlo approach (blue lines), RFV approach (cyan lines). On the left: the length (**upper plot**) and the width (**lower plot**). On the right: the evaluated area of the desk in case studies D (**upper plot**) and E (**lower plot**).

7. Conclusions

This paper represents a review paper of the RFV approach, proposed in the literature in the last decades.

It has been shown the potentiality of this approach, which is able to represent and propagate measurement results in closed form, by simulating the way the uncertainty contributions propagate through the measurement procedure.

Other more specific applications are present in the more recent literature, like for instance the generalization of Bayes' theorem in the possibility domain [19,20] or the realization of a possibilistic Kalman filter [21,22], thus showing the versatility of the RFV approach.

Author Contributions: Conceptualization, S.S.; methodology, S.S.; software, H.V.J.; validation, H.V.J.; formal analysis, H.V.J. and S.S.; investigation, H.V.J. and S.S.; resources, H.V.J. and S.S.; data curation, H.V.J. and S.S.; writing—original draft preparation, S.S.; writing—review and editing, H.V.J. and S.S.; visualization, H.V.J. and S.S.; supervision, S.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

- JCGM 100:2008. Joint Committee for Guides in Metrology, Evaluation of Measurement Data—Guide to the Expression of Uncertainty in Measurement (GUM 1995 with Minor Corrections). 2008. Available online: <https://www.bipm.org/en/publications/guides> (accessed on 29 August 2021).
- EA 4/02: Evaluation of the Uncertainty of Measurement in Calibration, European Cooperation for Accreditation, Amsterdam. 2013. Available online: <https://european-accreditation.org/wp-content/uploads/2018/10/ea-4-02-m-rev01-september-2013.pdf> (accessed on 29 August 2021).
- VIM—International Vocabulary of Metrology—Basic and General Concepts and Associated Terms, BIPM JCGM 200:2008. Available online: <https://www.bipm.org/en/publications/guides> (accessed on 29 August 2021).
- Mauris, G. Expression of Measurement Uncertainty in a Very Limited Knowledge Context: A Possibility Theory-Based Approach. *IEEE Trans. Instrum. Meas.* **2007**, *56*, 731–735. [CrossRef]
- Mauris, G. Representing and Approximating Symmetric and Asymmetric Probability Coverage Intervals by Possibility Distributions. *IEEE Trans. Instrum. Meas.* **2009**, *58*, 41–45. [CrossRef]
- Mauris, G.; Lasserre, V.; Foulloy, L. A fuzzy approach for the expression of uncertainty in measurement. *Measurement* **2001**, *29*, 165–177. [CrossRef]
- De Capua, C.; Romeo, E. A Comparative Analysis of Fuzzy t-Norm Approaches to the Estimation of Measurement Uncertainty, 13th IMEKO TC4 Symposium. 2004. Available online: <http://home.mit.bme.hu/~kollar/IMEKO-profiles-for-web/tc4/TC4-13rd-Athens-2004/Topic%204/D18.pdf> (accessed on 29 August 2021).
- Urbanski, M.; Wasowsky, J. Fuzzy approach to the theory of measurement inexactness. *Measurement* **2003**, *34*, 67–74. [CrossRef]
- Shafer, G. *A Mathematical Theory of Evidence*; Princeton University Press: Princeton, NJ, USA, 1976.
- Salicone, S.; Prioli, M. *Measurement Uncertainty within the Theory of Evidence*. Springer Series in Measurement Science and Technology. 2018. Available online: <https://doi.org/10.1007/978-3-319-74139-0> (accessed on 29 August 2021).
- Salicone, S. *Measurement Uncertainty: An Approach Via the Mathematical Theory of Evidence*; Springer Series in Reliability Engineering; Springer: New York, NY, USA, 2007; Available online: <https://www.springer.com/gp/book/9780387306551> (accessed on 29 August 2021).
- Ferrero, A.; Salicone, S. Uncertainty: Only one mathematical approach to its evaluation and expression? *IEEE Trans. Instrum. Meas.* **2012**, *61*, 2167–2178. [CrossRef]
- Ferrero, A.; Salicone, S. The construction of random-fuzzy variables from the available relevant metrological information. *IEEE Trans. Instrum. Meas.* **2009**, *58*, 1149–1157. [CrossRef]
- Ferrero, A.; Salicone, S. A comparison between the probabilistic and possibilistic approaches: The importance of a correct metrological information. *IEEE Trans. Instrum. Meas.* **2018**, *67*, 607–620. [CrossRef]
- Klir, G.J.; Yuan, B. *Fuzzy Sets and Fuzzy Logic. Theory and Applications*; Prentice Hall PTR: Hoboken, NJ, USA, 1995.
- Ferrero, A.; Prioli, M.; Salicone, S. Processing dependent systematic contributions to measurement uncertainty. *IEEE Trans. Instrum. Meas.* **2013**, *65*, 1015–1021. [CrossRef]

17. Ferrero, A.; Prioli, M.; Salicone, S. Joint random-fuzzy variables: A tool for propagating uncertainty through nonlinear measurement functions. *IEEE Trans. Instrum. Meas.* **2016**, *62*, 720–731. [[CrossRef](#)]
18. JCGM 101:2008. Joint Committee for Guides in Metrology, Evaluation of Measurement Data—Supplement 1 to the Guide to the Expression of Uncertainty in Measurement—Propagation of Distributions Using a Monte Carlo Method. 2008. Available online: <https://www.bipm.org/en/publications/guides> (accessed on 29 August 2021).
19. Jiang, W.; Salicone, S.; Ferrero, A.; Zhang, Q. A possible way to perform recursive Bayesian estimate in the possibility domain. *IEEE Trans. Instrum. Meas.* **2017**, *66*, 3218–3227. [[CrossRef](#)]
20. Ferrero, A.; Jetti, H.V.; Salicone, S. A Modified Bayes' Theorem for Reliable Conformity Assessment in Industrial Metrology. *Measurement* **2021**. in Press. [[CrossRef](#)]
21. Ferrero, A.; Jetti, H.V.; Salicone, S. The Possibilistic Kalman Filter: Definition and Comparison with the Available Methods. *IEEE Trans. Instrum. Meas.* **2021**, *70*, 1–11. [[CrossRef](#)]
22. Jetti, H.V.; Salicone, S. A possibilistic Kalman filter for the reduction of the final measurement uncertainty, in presence of unknown systematic errors. *Metrology* **2021**, *1*, 39–51. [[CrossRef](#)]

Article

A Possibilistic Kalman Filter for the Reduction of the Final Measurement Uncertainty, in Presence of Unknown Systematic Errors

Harsha Vardhana Jetti ¹ and Simona Salicone ^{2,*}

¹ Department of Energy, Politecnico di Milano, Via Lambruschini 4, 20156 Milano, Italy; harshavardhana.jetti@polimi.it

² Department of Electronics, Information and Bioengineering, Politecnico di Milano, P.za Leonardo da Vinci 32, 20133 Milano, Italy

* Correspondence: simona.salicone@polimi.it

Abstract: A Kalman filter is a concept that has been in existence for decades now and it is widely used in numerous areas. It provides a prediction of the system states as well as the uncertainty associated to it. The original Kalman filter can not propagate uncertainty in a correct way when the variables are not distributed normally or when there is a correlation in the measurements or when there is a systematic error in the measurements. For these reasons, there have been numerous variations of the original Kalman filter, most of them mathematically based (like the original one) on the theory of probability. Some of the variations indeed introduce some improvements, but without being completely successful. To deal with these problems, more recently, Kalman filters have also been defined using random-fuzzy variables (RFVs). These filters are capable of also propagating distributions that are not normal and propagating systematic contributions to uncertainty, thus providing the overall measurement uncertainty associated to the state predictions. In this paper, the authors make another step forward, by defining a possibilistic Kalman filter using random-fuzzy variables which not only considers and propagates both random and systematic contributions to uncertainty, but also reduces the overall uncertainty associated to the state predictions by compensating for the unknown residual systematic contributions.

Keywords: random-fuzzy variables; Kalman filter; systematic uncertainty contributions; styling

Citation: Jetti, H.V.; Salicone, S. A Possibilistic Kalman Filter for the Reduction of the Final Measurement Uncertainty, in Presence of Unknown Systematic Errors. *Metrology* **2021**, *1*, 39–51. <https://doi.org/10.3390/metrology1010003>

Academic Editor: Richard Leach

Received: 1 July 2021

Accepted: 10 August 2021

Published: 17 August 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The Kalman filter (KF) is an algorithm that has long been in existence. It filters the noise on the measured values of the states and provides an estimation of the system states based on the state equations. The classical KF algorithm requires that the states are free from any systematic errors and that the state variables are independent from each other and can be represented by Gaussian distributions [1]. But in most practical situations, the systematic error can not be compensated perfectly and there is a residual systematic error. In this case, the classical formulations of the KF underestimate the uncertainty associated to the state estimates, because the systematic error is not propagated in a correct mathematical way. To deal with this, attempts have been made to develop KF algorithms that are also able to consider systematic contributions to uncertainty [2–5]. For instance, in [5], the authors try to use a Schmidt KF that considers the systematic error as a separate state in the state equations and a noise covariance matrix of the possible systematic errors is built and propagated.

More recently, the theory of possibility has been proposed in the literature to represent and propagate both systematic and random contributions to uncertainty. The theory of possibility has been proven by numerous applications in the literature [6–11] to be an effective alternative to the theory of probability when both random and systematic contributions to uncertainty are present in the measurement procedure.

Some attempts to define KFs based on the theory of possibility are already present in the literature [12,13]. However, in [12,13], as far as understood, they consider uncertainty in a fuzzy way that is not compatible with the recommended guidelines in metrology, as specified in [14,15]. In metrology, uncertainty must be considered according to the definitions given in [15].

Within the framework of the theory of possibility, quantities are represented by possibility distributions [16–21]. In particular, as shown in [16–18], where measurement results are considered to be affected by both random and systematic contributions to uncertainty, measured quantities are represented by random-fuzzy variables (RFVs). RFVs consist of an internal membership function which represents the systematic contribution to uncertainty in the quantity and an external membership function which represents the overall uncertainty due to both the systematic and random contributions. As shown in [16,18], this way of representation is perfectly compatible with the metrological definitions given in [14,15]. So, to be able to utilize all the advantages of RFVs, the KF should be able to process them as well.

Possibilistic KFs based on RFVs are available in the literature [22,23]. In [22], a KF using RFVs is defined but there is a high noise in the state predictions given by the KF. In [23], the authors define a possibilistic KF that also uses RFVs and make a comparison with a few other existing KFs, including the Schmidt KF, clearly showing the advantages of the defined possibilistic KF.

Starting from the possibilistic KF defined in [23], this paper proposes an alternative version, which also allows reducing the effects of the systematic contributions to uncertainty, thereby reducing the overall uncertainty associated to the system state predictions. While the possibilistic KF defined in [23] is useful when we are only interested in propagating the residual systematic uncertainty to evaluate the total uncertainty associated to the state predictions from both the random and systematic contributions, the KF defined in this paper can be used to reduce the effects of the systematic contributions to uncertainty and thereby also reduce the overall uncertainty associated to the state predictions.

The rest of the paper has been organized in six sections. Section 2 describes the case study used for the simulation results for an initial validation of the alternative possibilistic KF. Section 3 describes the construction of the RFVs and the algorithm of the modified possibilistic KF described in [23]. Section 4 describes the algorithm for the alternative possibilistic KF proposed in this paper. Section 5 describes more simulations that have been performed to further validate the alternative possibilistic KF. Section 6 describes the experimental case study that has been performed to prove the effectiveness of the alternative possibilistic KF. Section 7 summarizes the paper and gives a conclusion.

To facilitate an easy comparison between the proposed possibilistic KF and the original one defined in [23], the same simulated case study as in [23] is considered here, as briefly described in Section 2.

2. The Case Study

The considered case study is quite simple. A vehicle is moving at a velocity $v_{\text{ref}}(t)$ with an acceleration $a_{\text{ref}}(t)$, as shown in Figure 1.

The state equations of the vehicle can be written as:

$$\begin{aligned} v_k &= v_{k-1} + \tau \cdot a_{k-1} + w_k^v \\ a_k &= a_{k-1} + w_k^a \end{aligned} \quad (1)$$

- v_k and a_k are velocity and acceleration of the vehicle at time k ;
- w_k^v and w_k^a are the standard deviation of the noise in velocity and acceleration respectively at time k ;
- τ is the time period within two successive measurements

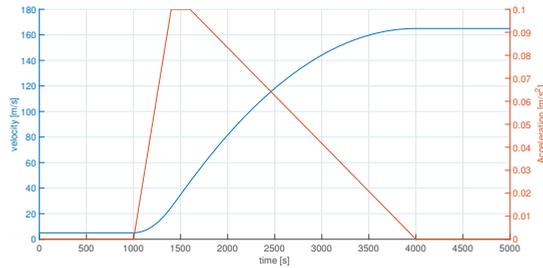


Figure 1. Reference values of velocity (blue line) and acceleration (red line) over time.

It is assumed that the noises are random in nature and belong to Gaussian distributions that do not vary with time (Gaussian distributions are considered as in [23], for a direct comparison). So, $w_k^v = w^v$ and $w_k^a = w^a$ are the standard deviations of the constant normal distributions with zero mean.

w^v is assumed to be 0.003 m/s. This value has been derived by considering the accuracy of a GPS which has been reported in the official GPS website [24], which is usually quite accurate compared to the speedometer of the vehicle. Whereas, w^a is assumed to be 0.0005 m/s² and is supposed to be due to some noise in the circuit or to the driver applying force on the accelerator.

The measured values of the velocity and the acceleration are supposed to have been obtained from the on board sensors of the vehicle. The accuracies of the onboard sensors are in general one or two magnitudes less accurate than a GPS based measurement. So, the following is considered:

- For the velocity, the random contribution is assumed to be normally distributed with a standard deviation of $\sigma_m^v = 0.16$ m/s. It has also been assumed that there is a residual systematic error in the measurement with an estimated value of 0.3 m/s. However, this is unknown and only an interval of possible values is known: $e_{sys} = [-0.32 \text{ m/s}, +0.32 \text{ m/s}]$ has been assumed.
- For the acceleration, it has been assumed that there is no systematic error in the measurements and the random error is supposed to be normally distributed with a standard deviation of $\sigma_m^a = 0.005$ m/s².

3. Construction of the RFVs and the Possibilistic Kalman Filter

Although this has been explained in detail in [23], it has been recalled in this paper as the construction of the RFVs is the same also for the alternative possibilistic KF defined in this paper.

In the possibilistic KF defined in [23], all the states are RFVs and the algorithm is as shown in Figure 2 [23].

$$\text{According to Equation (1): } \mathbf{A}_k = \mathbf{A} = \begin{bmatrix} 1 & \tau \\ 0 & 1 \end{bmatrix} \text{ and } \mathbf{H}_k = \mathbf{H} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Matrix \mathbf{Q}^{POS} considers the model uncertainties and is a matrix of RFVs. According to the assumptions given in Section 2, we define \mathbf{Q}^{POS} where:

- The element related to velocity is an RFV obtained by transforming the velocity noise variable into the possibility domain. Since there is no systematic error in the noise and the random part is assumed to be Gaussian, there is no internal possibility distribution (PD) in the RFV and the random PD is obtained by applying the probability-possibility transformation [16] on the zero mean normal probability density function (pdf) with standard deviation w^v in the possibility domain;
- Similarly, the element related to acceleration is also an RFV in which there is no internal PD and the random PD is obtained by applying the probability-possibility transformation [16] on a zero mean normal pdf with standard deviation w^a in the possibility domain;

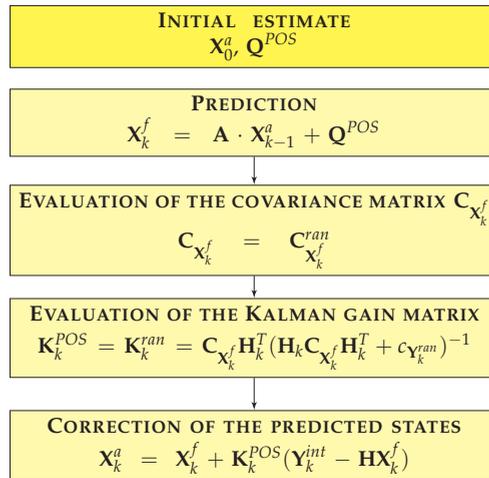


Figure 2. The possibilistic Kalman filter algorithm [23].

As for the initial state vector X_0^a , it is assumed that there are no systematic contributions to uncertainty. So, the RFV is obtained by just the random PD as follows:

- The initial velocity is an RFV consisting of just the random PD which is obtained by using the probability-possibility transformation [16] on a normal pdf with mean equal to the first measured value for velocity (v_{m1}) and standard deviation w^v ;
- Similarly, the initial acceleration is an RFV consisting of just the random PD obtained by using the probability-possibility transformation [16] on a normal pdf with mean equal to the first measured value for acceleration (a_{m1}) and standard deviation w^a .

As for the measured values in each step k , matrix Y_k is the matrix of the RFVs of the velocity and acceleration measurements. The RFV associated with the simulated measured velocity is centered on the simulated measured velocity at step k (v_{mk}) and

- The internal PD is a rectangular PD with width $\pm e_{sys}$ around v_{mk} ;
- the random PD is obtained by using the probability-possibility transformation [16] on a zero mean normal pdf, with standard deviation σ_m^v .

On the other hand, the acceleration has no systematic error. So, the RFV associated to the simulated measured acceleration is centered on the simulated measured acceleration at step k (a_{mk}) and

- the internal PD is zero;
- the random PD is obtained by using the probability-possibility transformation [16] on a normal pdf, with mean a_k and standard deviation u_{ran}^a .

Matrix $C_{X_k^f}$ is the noise covariance matrix of the velocity and acceleration RFVs. However, as it is shown in the equations in Figure 2 and explained in [23], $C_{X_k^f} = C_{X_k^f}^{ran}$. So, the possibilistic variances and covariances are evaluated from only the random contributions to uncertainty in both the velocity and acceleration RFVs.

Similarly, $C_{Y_k} = C_{Y_k^{ran}}$ which means that the possibilistic variances and covariances of the noise covariance matrix associated with the measurements are evaluated from just the random uncertainty contributions in the velocity and acceleration measurements.

The described KF has been applied to the case study described in Section 2. The results obtained from the simulations are presented in Figures 3 and 4.

The predicted values of the velocity and acceleration from the KF are obtained by evaluating the mean values of the a posteriori RFVs in matrix X_k^a . In both Figures 3 and 4, the blue lines represent the differences in the predicted values given by the KF and the true values of the velocity and acceleration respectively.

The uncertainty limits associated to the state predictions (red lines) are the α -cut at $\alpha = 0.01$ of the velocity and acceleration RFVs predicted by the KF. The α -cut can be considered as the confidence interval at the confidence level $1-\alpha$ [16]. For $\alpha = 0.01$, these intervals correspond to the 99% confidence interval in the corresponding pdf.

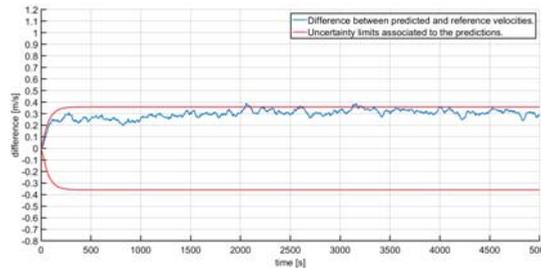


Figure 3. Difference in the reference and predicted velocity values (blue line) provided by the possibilistic Kalman filter together with the predicted uncertainty interval (red lines).

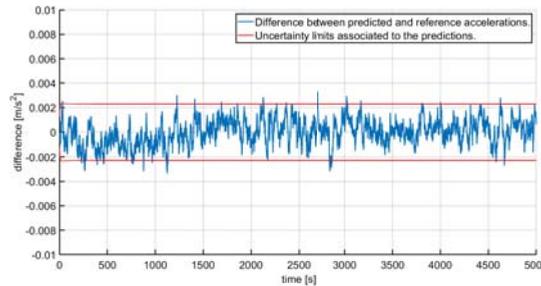


Figure 4. Difference in the reference and predicted acceleration values (blue line) provided by the possibilistic Kalman filter together with the predicted uncertainty interval (red lines).

4. The Alternative Kalman filter Algorithm

In this paper, an alternative version of the KF algorithm described in Section 3 is presented, which allows for the reduction of the residual systematic error. As can be seen in the results in Figures 3 and 4, the possibilistic KF algorithm described in Section 3 estimates the uncertainty intervals associated with the predictions very accurately in the presence of a systematic error. However, it does not compensate for the systematic error.

The alternative possibilistic KF which is proposed in this paper makes use of the above uncertainty interval to partially compensate for the systematic error. The new algorithm is synthetically shown in Figure 5. With respect to the algorithm in Figure 2, it can be seen that all steps are equal, except the last one, which corresponds to the “correction of the predicted states”.

In particular, a new RFV \mathbf{Y}_k^{comp} is considered, which tries to compensate for the residual systematic error. At each step k , \mathbf{Y}_k^{comp} consists of just the internal PD which is centered at the positive uncertainty limit evaluated by the KF at the previous iteration (step $k - 1$) and with the same width and shape as the internal membership function of the RFVs of the state variables estimated by the KF in the previous iteration (\mathbf{X}_{k-1}^{int}).

$\mathbf{Y}_k^{int_modified}$ is then obtained by adding or subtracting the RFV \mathbf{Y}_k^{comp} from \mathbf{Y}_k^{int} , depending on if the systematic error is positive or negative:

$$\mathbf{Y}_k^{int_modified} = \begin{cases} \mathbf{Y}_k^{int} + \mathbf{Y}_k^{comp} & \text{if systematic error} < 0 \\ \mathbf{Y}_k^{int} - \mathbf{Y}_k^{comp} & \text{if systematic error} > 0 \end{cases} \quad (2)$$

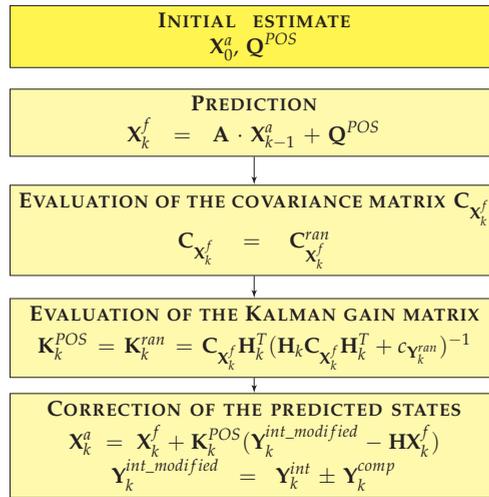


Figure 5. The alternative possibilistic Kalman filter algorithm.

It is exactly like a negative feedback loop: the effects of the systematic contributions to uncertainty predicted by the KF is used as a feedback to compensate for a possible systematic error and the systematic error is partially compensated for. The intrinsic requirement for applying this method is that we know the direction of the systematic error i.e., it should be known if the error is positive or negative.

The obtained results are shown in Figures 6 and 7. Again, the predicted values for the velocity and acceleration given by the KF are the mean values of the velocity and acceleration RFVs in matrix X_k^a .

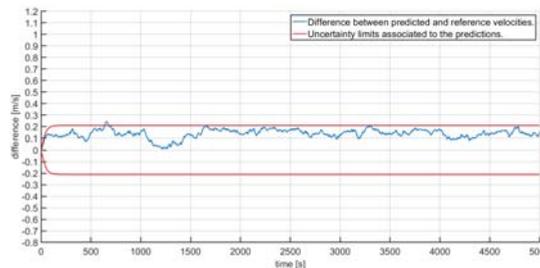


Figure 6. Difference in the reference and predicted velocity values (blue line) provided by the possibilistic Kalman filter defined in this paper, together with the predicted uncertainty interval (red lines).

As in Figures 3 and 4, also in Figures 6 and 7 the blue lines represent the differences in the predicted values given by the KF and the true values of the velocity and acceleration respectively. The uncertainty limits associated the state predictions (red lines) are the $\alpha - cut$ at $\alpha = 0.01$ of the velocity and acceleration RFVs predicted by the KF.

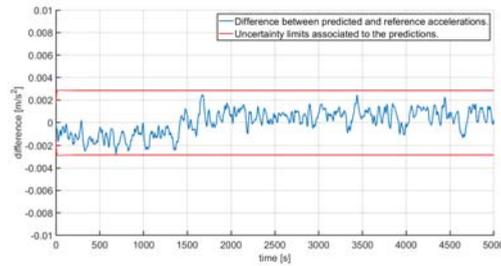


Figure 7. Difference in the reference and predicted acceleration values (blue line) provided by the possibilistic Kalman filter defined in this paper, together with the predicted uncertainty interval (red lines).

In Figure 6, with respect to Figure 3, it can be clearly seen that the uncertainty limits have been significantly reduced along with the residual systematic error in the velocity estimate. Table 1 gives a comparison with synthetic indexes for the velocity of the possibilistic KF and the alternative possibilistic KF.

Table 1. Comparison of synthetic indexes for the velocity.

KF	Possibilistic	Alternative Possibilistic
Convergence(s)	151	138
Steady-state error	0.3024	0.1696
Variation of error	0.0220	0.0257
Uncertainty limits	± 0.3589	± 0.2106
Variation of uncertainty limits	0	0
Percentage inside the uncertainty limits	99.00	95.88

5. Further Simulations

Further simulations have been performed in order to verify the effectiveness of the alternative possibilistic KF in all situations. In particular, we want to verify whether the algorithm still works in a good way when it is applied, but no residual systematic error is present.

In fact, the result of the introduction of the “feedback” loop is that the residual systematic error is compensated by the maximum possible value since the uncertainty limit of the RFVs evaluated in each step (which is the value of the α -cut at $\alpha = 0.01$ of the RFV) is considered. This means that it is possible that the residual systematic error could be overcompensated as the magnitude of this is unknown.

So, it is important that even if the residual systematic error happens to be zero (which is the limiting case), the overcompensation should not be so high that the predictions of the state variables obtained from the KF fall out of the evaluated uncertainty limits. To verify this, the same example described in the Section 2 is considered except that the systematic error is considered to be zero (instead of 0.3 m/s).

In this case, the results in Figures 8 and 9 are obtained.

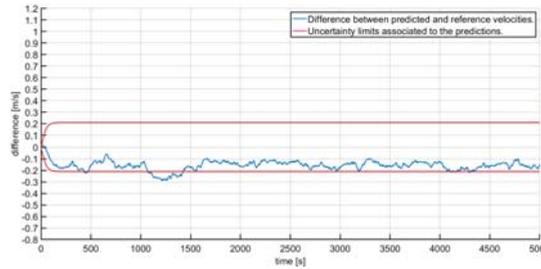


Figure 8. Difference in the reference and predicted velocity values (blue line) provided by the alternative possibilistic KF, together with the predicted uncertainty interval (red lines) when residual systematic error is zero.

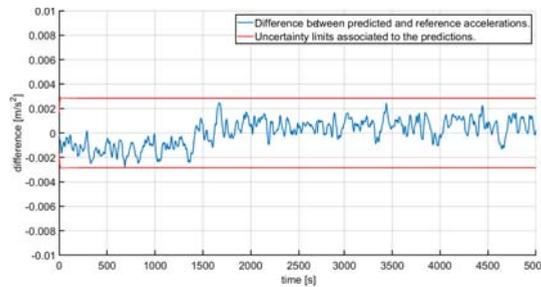


Figure 9. Difference in the reference and predicted acceleration values (blue line) provided by the alternative possibilistic KF, together with the predicted uncertainty interval (red lines) when residual systematic error is zero.

As expected, as can be seen in Figure 8, the systematic error in the velocity has been overcompensated, but it is still mostly inside the evaluated uncertainty limits. This demonstrates that the alternative possibilistic KF algorithm successfully decreases the uncertainty associated to the state predictions provided by the KF in all situations. In fact, the average uncertainty in Figure 8 is in any case smaller than the one in Figure 3.

6. Experimental Case Study

To validate the simulation results, a parrot AR drone has been used for the experimental case study. The drone has the following technical specifications as given by the manufacturer:

- 1 GHz 32 bit ARM Cortex processor with 800 MHz video DSP.
- 1 Gbit DDR2 RAM at 200 MHz.
- Wi-Fi b/g/n.
- 3 axis accelerometer ± 50 mg precision.
- 3 axis gyroscope 2000° second precision.
- Pressure sensor ± 10 Pa precision.
- 60fps vertical QVGA camera.
- 3 axis magnetometer 6° precision
- Ultrasound sensors.

The parrot AR drone has been developed as a low cost drone by parrot company and is quite customizable. The code is open source and can be modified according to the necessity. It has a variety of sensors and the data can be obtained from them and processed as needed. For the present case study, the velocity and acceleration measurements have been considered. For information about the algorithm used by the drone to calculate its speed, the readers are suggested to refer to [25].

The employed drone has been observed to have a negative systematic error in the velocity measurements obtained from the sensors present in the drone itself. So, the velocity is being underestimated by the sensors of the drone. It has also been observed that the systematic error is not constant for all runs. Each individual run had a systematic error that may be different from the other runs. So, only an interval of values can be estimated and the error can not just be compensated.

By performing a large number of runs of the drone, the interval for the systematic error has been estimated and this was used to construct the internal membership function of the RFV for the measured velocity. The constructed RFV assumed to be centered at zero velocity can be seen in Figure 10.

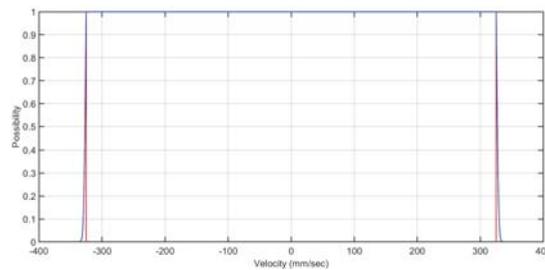


Figure 10. RFV of the velocity constructed from the data. The blue line represents the external membership function and the red line represents the internal membership function.

The measured acceleration, on the other hand, does not have any systematic contributions to uncertainty. Hence, the RFV can be constructed by simply using a probability-possibility transformation on the probability distribution of the acceleration.

The drone was made to fly for a few seconds to cover a distance of approximately 4 m. The velocity and acceleration data from the sensors is obtained from the drone every 5 ms using a software program that links the computer with the drone using the Wi-Fi network. The alternative possibilistic KF described in Section 4 was used to provide the filtered velocity and acceleration predictions with their respective uncertainties as well as compensate partially for the systematic error in the velocity measurements provided by the drone.

The velocity estimates provided by the KF were integrated to get the estimated distance traveled by the drone. Similarly, the velocity measurements directly obtained by the drone were integrated as well, to get the distance that the drone traveled according to the sensors present in the drone.

At the end of every run, the actual distance from the starting point was been measured. Measuring tape was used to do this since the error in the distance calculated using the velocity data from the sensors is quite high and the precision of the measuring tape is enough to be deemed negligible. Several runs were made and the distances estimated by the KF and those estimated according to the sensor data were compared with the actual distance traveled by the drone. To facilitate a comparison between the alternative KF defined in this paper and the possibilistic KF defined in [23], the sensor data was processed using both the KFs separately.

The results using the possibilistic KF defined in [23] can be seen in Figure 11. The green line represents the distances estimated according to the velocity measurements obtained directly from the sensors in the drone. The blue line represents the distance obtained from the velocity estimates of the defined possibilistic KF. The black line represents the actual distance traveled by the drone. Finally, the red lines represent the upper and lower bounds for the uncertainty.

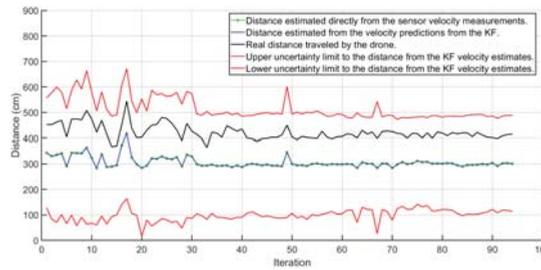


Figure 11. Distances obtained from the velocity estimates of the possibilistic KF (blue line). The predicted uncertainty intervals (red lines). Actual distance traveled by the drone (black line) and distances estimated according to the velocity measurements obtained directly from the sensors in the drone (green line). Green line and blue line are almost the same.

It can be seen that the distances estimated by the possibilistic KF are quite close to the distances from the sensors. The blue line and the green line in Figure 11 are almost the same and that is why only the green dots and the blue line can be seen in the figure. However, the real measurements lie inside the uncertainty limits of the distances provided by the KF.

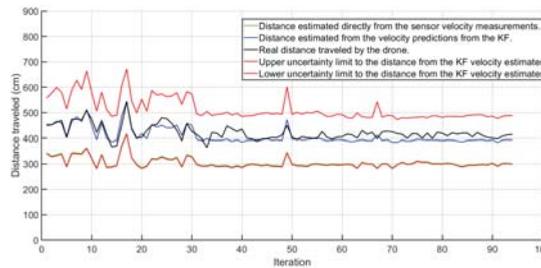


Figure 12. Distances obtained from the velocity estimates of the defined alternative possibilistic KF (blue line). The predicted uncertainty intervals (red lines). Actual distance traveled by the drone (black line) and distances estimated according to the velocity measurements obtained directly from the sensors in the drone (green line).

The results using the alternative KF defined in this paper can be seen in Figure 12. Again, the green line represents the distances estimated according to the velocity measurements obtained directly from the sensors in the drone. The blue line represents the distance obtained from the velocity estimates of the defined possibilistic KF. The black line represents the actual distance traveled by the drone. Finally, the red lines represent the upper and lower bounds for the uncertainty.

For an easier comparison, Figure 13 shows again the distances obtained using the modified possibilistic KF (green line) and those obtained using the alternative possibilistic KF (blue line) along with the actual distance traveled by the drone (red line).

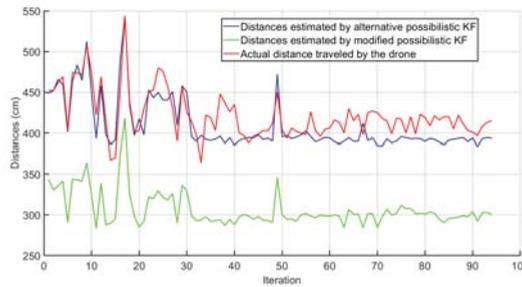


Figure 13. Distances obtained from the velocity estimates of the defined alternative possibilistic KF (blue line). The distances obtained from the velocity estimates of the modified alternative possibilistic KF (green lines). Actual distance traveled by the drone (red line).

A comparison of the results obtained from the two KFs has also been given in Table 2. From Table 2, it can be clearly seen that the distance obtained using the alternative KF defined in this paper is much more accurate and closer to the real measurements than the distances obtained from the sensor measurements or those obtained from the possibilistic KF defined in [23].

Table 2. Comparison of the distance estimates of the drone obtained from the two KFs.

KF	Possibilistic	Alternative Possibilistic
Average error between the real distance and estimated distance	115.3273	12.4821
Mean width of the uncertainty band	±416.30	±210.63

Additionally, it can be easily seen that the width of the uncertainty limits associated with the distance (red lines) are also smaller in Figure 12 compared to that in Figure 11. The same can be verified from Table 2.

This confirms that the systematic error in the velocity is being compensated quite efficiently using the defined alternative possibilistic KF and the overall uncertainty associated to the predictions is being decreased as well.

7. Conclusions

The modified possibilistic KF defined in [23] is capable of propagating the systematic contributions to uncertainty effectively. This paper defines an alternative possibilistic KF which also decreases the effects of the systematic uncertainty contributions on the final measurement and therefore can be considered an improved version of the KF defined in [23].

The same simulated case study as in [23] has been considered to facilitate an easy comparison and the results obtained using the KF defined in this paper have been shown along with the results obtained by using the KF defined in [23]. The obtained results show that the proposed KF provides a compensation of the systematic uncertainty and decreases the overall uncertainty associated to the predictions.

The only requirement to use this method is that the direction of the residual systematic error should be known. This requirement is not so difficult to be satisfied in the era of big data. In any case, if not satisfied, the modified possibilistic KF defined in [23] is still valid and can be successfully applied. A possible area of application of the alternative possibilistic KF proposed in this paper could be in PTP networks where the network traffic is being monitored and thereby it can be evaluated if the transmission delay is higher from master to slave or from slave to master, thus identifying the direction of the systematic error

in the calculation of the offset. Therefore, this method could be used to further decrease the uncertainty associated with the time predictions provided by the KF.

Author Contributions: Conceptualization, H.V.J. and S.S.; methodology, H.V.J. and S.S.; software, H.V.J.; validation, H.V.J.; formal analysis, H.V.J. and S.S.; investigation, H.V.J. and S.S.; resources, H.V.J. and S.S.; data curation, H.V.J. and S.S.; writing—original draft preparation, H.V.J.; writing—review and editing, S.S.; visualization, H.V.J. and S.S.; supervision, S.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Faragher, R. Understanding the basis of the kalman filter via a simple and intuitive derivation. *IEEE Signal Process. Mag.* **2012**, *29*, 128–132. [CrossRef]
- Wei, W.; Gao, S.; Zhong, Y.; Gu, C.; Subic, A. Random weighting estimation for systematic error of observation model in dynamic vehicle navigation. *Int. J. Control Autom. Syst.* **2016**, *14*, 514–523. [CrossRef]
- Yang, Y.; Rees, N.; Chuter, T. Reduction of encoder measurement errors in ukirt telescope control system using a kalman filter. *IEEE Trans. Control Syst. Technol.* **2002**, *10*, 149–157. [CrossRef]
- Noack, B.; Klumpp, V.; Hanebeck, U.D. State estimation with sets of densities considering stochastic and systematic errors. In Proceedings of the 12th International Conference on Information Fusion, Seattle, WA, USA, 6–9 July 2009.
- Novoselov, R.Y.; Herman, S.M.; Gadaleta, S.M.; Poore, A.B. Mitigating the effects of residual biases with schmidt-kalman filtering. In Proceedings of the 2005 7th International Conference on Information Fusion, Philadelphia, PA, USA, 25–28 July 2005; Volume 1, p. 8.
- Zhu, Q.; Jiang, Z.; Zhao, Z.; Wang, H. Uncertainty estimation in measurement of micromechanical properties using random-fuzzy variables. *Rev. Sci. Instrum.* **2006**, *77*, 035107. [CrossRef]
- Pertile, M.; Cecco, M.D.; Baglivo, L. Uncertainty evaluation in two-dimensional indirect measurement by evidence and probability theories. *IEEE Trans. Instrum. Meas.* **2010**, *59*, 2816–2824. [CrossRef]
- Pertile, M.; Cecco, M.D. Uncertainty evaluation for complex propagation models by means of the theory of evidence. *Meas. Sci. Technol.* **2008**, *19*, 1–10. [CrossRef]
- Lou, C.W.; Dong, M.C. A novel random fuzzy neural networks for tackling uncertainties of electric load forecasting. *Electr. Power Energy Syst.* **2015**, *73*, 34–44. [CrossRef]
- Tu, Y.; Zhou, X.; Gang, J.; Liechty, M.; Xu, J.; Lev, B. Administrative and market-based allocation mechanism for regional water resources planning. *Resour. Conserv. Recycl.* **2015**, *95*, 156–173. [CrossRef]
- Chen, D.; Chu, X.; Sun, X.; Li, Y. A new product service system concept evaluation approach based on information axiom in a fuzzy-stochastic environment. *Int. J. Comput. Integr. Manuf.* **2015**, *28*, 1–19. [CrossRef]
- Matia, F.; Jiménez, A.; Al-Hadithi, B.M.; Rodríguez-Losada, D.; Galán, R. The fuzzy kalman filter: State estimation using possibilistic techniques. *Fuzzy Sets Syst.* **2006**, *157*, 2145–2170. [CrossRef]
- Oussalah, M.; Schutter, J.D. Possibilistic kalman filtering for radar 2d tracking. *Inf. Sci.* **2000**, *130*, 85–107. [CrossRef]
- JCGM 100:2008. *Evaluation of Measurement Data—Guide to the Expression of Uncertainty in Measurement*, (GUM 1995 with Minor Corrections); Joint Committee for Guides in Metrology. 2008. Available online: <http://www.bipm.org/en/publications/guides/gum.html> (accessed on 12 June 2021).
- JCGM 200:2012. *International Vocabulary of Metrology—Basic and General Concepts and Associated Terms* (VIM 2008 with Minor Corrections); Joint Committee for Guides in Metrology. 2012. Available online: <http://www.bipm.org/en/publications/guides/vim.html> (accessed on 12 June 2021)
- Salicone, S.; Prioli, M. *Measurement Uncertainty within the Theory of Evidence*; Springer Series in Measurement Science and Technology; Springer: New York, NY, USA, 2018.
- Ferrero, A.; Salicone, S. A comparison between the probabilistic and possibilistic approaches: The importance of a correct metrological information. *IEEE Trans. Instrum. Meas.* **2018**, *67*, 607–620. [CrossRef]
- Ferrero, A.; Salicone, S. Uncertainty: Only one mathematical approach to its evaluation and expression? *IEEE Trans. Instrum. Meas.* **2012**, *61*, 2167–2178. [CrossRef]
- Mauris, G.; Lasserre, V.; Foulloy, L. A fuzzy approach for the expression of uncertainty in measurement. *Measurement* **2001**, *29*, 165–177. [CrossRef]
- Mauris, G.; Berrah, L.; Foulloy, L.; Haurat, A. Fuzzy handling of measurement errors in instrumentation. *IEEE Trans. Instrum. Meas.* **2000**, *49*, 89–93. [CrossRef]
- Urbanski, M.; Wasowsky, J. Fuzzy approach to the theory of measurement inexactness. *Measurement* **2003**, *34*, 67–74. [CrossRef]

22. Ferrero, A.; Ferrero, R.; Salicone, S.; Jiang, W. The kalman filter uncertainty concept in the possibility domain. *IEEE Trans. Instrum. Meas.* **2019**, *68*, 4335–4347. [[CrossRef](#)]
23. Ferrero, A.; Jetti, H.V.; Salicone, S. The possibilistic kalman filter: Definition and comparison with the available methods. *IEEE Trans. Instrum. Meas.* **2020**, *77*, 035107.
24. GPS Accuracy. Available online: <https://www.gps.gov/systems/gps/performance/accuracy/> (accessed on 28 May 2021).
25. Bristeau, P.J.; Callou, F.; Vissière, D.; Petit, N. The navigation and control technology inside the ar. drone micro uav. *IFAC Proc. Vol.* **2011**, *44*, 1477–1484. [[CrossRef](#)]

Article

Calibration of a Digital Current Transformer Measuring Bridge: Metrological Challenges and Uncertainty Contributions [†]

Guglielmo Frigo * and Marco Agustoni

Swiss Federal Institute of Metrology METAS, Lindenweg 50, 3007 Bern-Wabern, Switzerland; marco.agustoni@metas.ch

* Correspondence: guglielmo.frigo@metas.ch

[†] Any mention of commercial products within this paper is for information only; it does not imply recommendation or endorsement by METAS.

Abstract: In this paper, we consider the calibration of measuring bridges for non-conventional instrument transformers with digital output. In this context, the main challenge is represented by the necessity of synchronization between analog and digital outputs. To this end, we propose a measurement setup that allows for monitoring and quantifying the main quantities of interest. A possible laboratory implementation is presented and the main sources of uncertainty are discussed. From a metrological point of view, technical specifications and statistical analysis are employed to draw up a rigorous uncertainty budget of the calibration setup. An experimental validation is also provided through the thorough characterization of the measurement accuracy of a commercial device in use at METAS laboratories. The proposed analysis proves how the calibration of measuring bridges for non-conventional instrument transformers requires ad hoc measurement setups and identifies possible space for improvement, particularly in terms of outputs' synchronization and flexibility of the generation process.

Keywords: measuring bridge; calibration; non-conventional instrument transformer; sampled values; digital output; synchronization

Citation: Frigo, G.; Agustoni, M. Calibration of a Digital Current Transformer Measuring Bridge: Metrological Challenges and Uncertainty Contributions. *Metrology* **2021**, *1*, 93–106. <https://doi.org/10.3390/metrology1020007>

Academic Editor: Simona Salicone

Received: 31 August 2021

Accepted: 15 September 2021

Published: 3 October 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In view of reducing greenhouse gas emissions and carbon dependence, modern power systems are experiencing an ever-increasing integration of renewable energy sources and distributed generation [1,2]. Such resources are typically connected via dedicated inverters whose power electronics-based control can not guarantee any rotational inertia or regularization of the energy generation profile [3,4]. As a consequence, power systems are expected to face much faster dynamics, as proven by recent adverse events in South Australia and California [5,6].

In order to address such challenges, also the measurement infrastructure needs to undergo a significant renovation, both in terms of instrumentation and control strategies [7]. In particular, the transition from traditional to digital electrical substations paves the way to more sophisticated and optimized approaches for the collection and aggregation of the quantities of interest, e.g., voltage and current levels at the transformer secondary windings [8]. In this context, the recent IEC Std 61869-9:2016 [9] defines the operational and communication requirements for instrument transformer with digital output. Due to their capability of converting the output signal directly in a digital form (and thus compatible with many processing and storage applications), such transformers are typically referred to as non-conventional instrument transformers, briefly NCIT [10].

In terms of communication protocol, the IEC Std 61850-9-2:2011 [11] introduces the Sampled Values (SV): a publisher/subscriber protocol for information exchange between Stand Alone Merging Units (SAMUs) and Intelligent Electronic Devices (IEDs) over the Ethernet. Originally conceived just as an efficient way to concentrate the outputs of NCITs

and SAMUs [12–14], the SV is now directly applied to more sophisticated processing applications, e.g., phasor measurements [15] and protection schemes [16].

The recent EMPIR project FutureGrid II has been investigating the measurement needs and potential of SVs in modern electrical substations. In particular, dedicated calibration infrastructures for transmitting and receiving SVs have been developed and thoroughly characterized [17,18]. However, a rigorous and well-established procedure for the metrological characterization of NCITs is not yet available. The measurement setup typically includes a transformer measuring bridge capable of processing both analog and digital inputs [19]. The calibration of such device, though, is not straightforward and requires a precise assessment of the several uncertainty sources involved in the measurement process. Indeed, the comparison between purely analog quantities and time-stamped digital values represents a non-negligible challenge, especially in terms of synchronization and phase angle uncertainty.

In this paper, we consider the problem of calibrating a measuring bridge for NCITs from a metrological point of view [20,21]. In particular, we describe a novel measurement setup and discuss the implementation challenges and requirements as well as the possible uncertainty contributions. A preliminary calibration campaign confirms the feasibility and reliability of the proposed approach, and sets a realistic performance target for the uncertainty budget of the calibration infrastructure.

The paper is organized as follows: in Section 2, we present the measurement principle inherent in measuring bridges for traditional and non-conventional transformers. Section 3 outlines the measurement setup for the bridge calibration and describes the actual implementation in METAS laboratories. In Section 4, we discuss the main uncertainty contributions and derive a preliminary uncertainty budget based on technical specifications and statistical analysis. In Section 5, we provide an experimental validation by presenting the results of a measurement campaign on a commercial device. Finally, Section 6 provides some closing remarks and outlines the next steps of the research.

2. Measuring Bridge: Configurations and Measurement Principles

In this section, we briefly describe the measurement principle of measuring bridges for instrument transformers, focusing on the transition from the traditional analog approach to the non-conventional approach based on SV communication protocol. In the following, we refer to the specific case of current transformers but similar considerations apply as well to voltage transformers.

Traditional measuring bridges for instrument transformers rely on the well-known difference method [22]. As shown in Figure 1a, the same current source I_s is supplied to two current transformers: a standard reference transformer, typically referred to as normal (channel N), and the transformer under test (channel X). For the sake of comparability, the two transformers adopt the same transformation ratio. As a consequence, they should produce the same current output at the secondary winding.

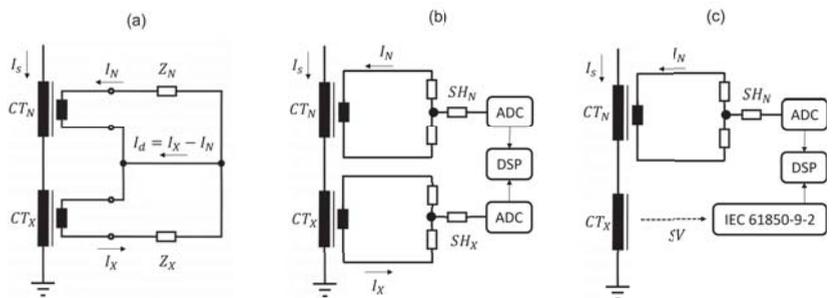


Figure 1. Typical configuration of a measuring bridge based on: difference method (a), digital signal processing (b), and IEC 61850-9-2 protocol (c).

It is worth noticing that, in a calibration context, the current source and the normal transformers are subject to periodic and thorough metrological characterization campaigns: systematic errors are suitably compensated, whereas random contribution determine the source stability and the transformer uncertainty, whose levels are guaranteed to be much lower than the expected performance of the device under test.

By means of current sensors (typically, a calibrated shunt and a voltmeter or a digitizer), the measuring bridge determines the current flowing at the secondary winding of the two transformers, I_N and I_X , respectively, as well as their difference $I_d = I_X - I_N$. In a vector space rotating at the nominal system frequency (e.g., 50 Hz), it is possible to represent these quantities as rotating vectors, whose magnitude and phase depend on the characteristics of the transformer under test.

Based on these measurements, the measuring bridge calculates the complex transformer or excitation error $\Delta E = I_d / I_N$ (The excitation error is not necessarily included in a calibration report as it depends on the accuracy and stability of the selected current source and reference standard transformer. In this paper, we report also ΔE as it is one of the measurement values commonly output by a measurement bridge, and thus it might be interesting to associate it with a measurement uncertainty), the transformer ratio error $\Delta \varepsilon$, and the phase displacement $\Delta \varphi$ (In this paper, the test waveforms consist of sine waveforms. Therefore, a negative phase displacement corresponds to a current I_X that is *delayed* with respect to the reference current I_N).

In Figure 1b, we present an example of new generation of measuring bridges. With the emergence of integrated circuits and fixed-point microprocessors, also measuring bridges have been equipped with Analog-to-Digital Converters (ADCs) and Digital Signal Processing (DSP) units for a more sophisticated treatment of the digitizer outputs. Instead of considering their difference in an analog circuit, each channel is processed independently: by means of a Discrete Fourier Transform (DFT), it is possible to define the complex coefficient associated with the nominal system rate. The comparison between these complex quantities allow for quantifying the excitation and ratio errors and the phase displacements. Moreover, by differentiating the phase information, it is also possible to determine the signal frequency and detect possible distortion introduced in the transformation.

Finally, Figure 1c represents the configuration of a measuring bridge for NCITs. As the transformer under test outputs the current at the secondary winding directly in a digital format, the X channel has to be supplied with an Ethernet board responsible of capturing the SV data packets and aligning them with the samples provided by the ADC on the N channel.

First, the captured SV data packets are queued in a First-In-First-Out (FIFO) buffer. Then, the time-stamp information is extracted and compared with the internal time of the measuring bridge: in the presence of high discrepancies (e.g., delayed transmission), the comparison with the reference channel values is unfeasible and the measuring bridge outputs an error message due to synchronization loss. Otherwise, the analog quantities are extracted from the SV data packets and transmitted to the DSP for the DFT processing and the error computation.

In this regard, it is reasonable to assume that the excitation error ΔE is mostly dependent on the accuracy and stability of the current measurement at the N-channel. In the absence of synchronization or packet loss, the SV data stream is characterized by a constant amplitude whose accuracy depends only on the quantization error and on possible numerical errors in the bridge DSP. On the contrary, I_N is an analog quantity that might vary as function of time, depending on the stability of the current source and on the characteristic of the standard transformer CT_N .

3. Measurement Setup

In this section, we present the measurement setup for the metrological characterization of a measuring bridge for non-conventional instrument transformers. Indeed, a detailed

analysis of the employed instruments and measurement techniques is crucial in view of the uncertainty analysis in the following section.

As shown in Figure 2, the setup consists of six main components: a time reference, a calibrator, a transconductance amplifier, a calibrated shunt, a set of synchronized voltmeters, and the Device Under Test (DUT), i.e., the measuring bridge.

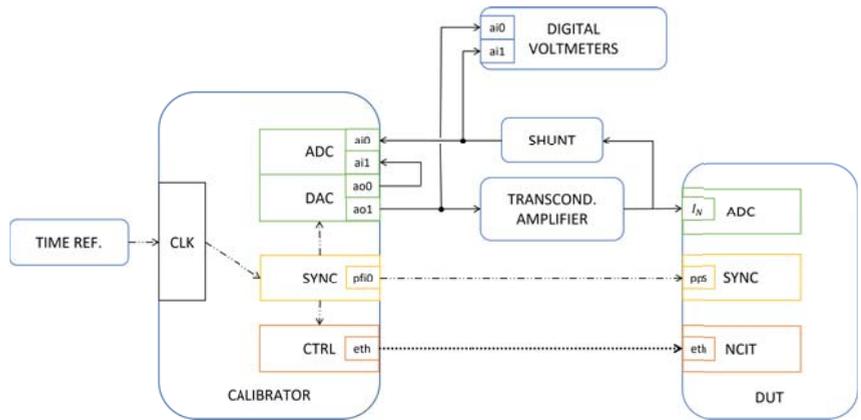


Figure 2. Measurement setup employed for the calibration of measuring bridges for non-conventional instrument transformers.

The time reference is responsible for providing the calibrator with a refined and stable time-base. To this end, a 10-MHz signal overrides the internal clock of the calibrator. It is worth noticing that, in such application, the traceability to Universal Time Coordinate (UTC) time is not mandatory. The only constraint is the exact synchronization between the calibrator analog and digital outputs, as well as between the calibrator and the measuring bridge.

The calibrator consists of three main units:

- A digital acquisition unit with a Digital-to-Analog Converter (DAC) and an ADC that operate in simultaneous mode, i.e., share the same time-base and sampling rate. The DAC is responsible for generating the analog test waveform to be supplied at the transconductance amplifier and then the DUT, whereas the ADC simultaneously re-acquires the same waveform to make it available for further processing and defining the actual reference values. In this context, it should be noticed that both DAC and ADC are typically equipped with two channels. One pair of channels (ai0 and ao1 in Figure 2) is dedicated to the test waveform generation and re-acquisition, whereas the other one (ai1 and ao0) is intended for self-calibration tasks, namely for the definition of the DAC phase offset [18], as further discussed in the following section;
- A synchronization unit locked to the internal clock, and thus to the external time reference. The synchronization board is responsible for two main tasks: distributing the triggers for the other units within the calibrator, and providing the measuring bridge with a Pulse-Per-Second (PPS) signal that is aligned with the time-stamp of the SV data packets. As regards the first task, the main difficulty is represented by the necessity of simultaneously triggering a purely hardware unit, namely the DAQ, and a purely software process, namely the SV transmission. To this end, software defined triggers are programmed as future time events, i.e., in correspondence of the first rising edge of the internal time base after a given time instant. As regards the second task, instead, the PPS is generated as a Transistor-Transistor Logic (TTL) signal, disciplined at the same rising edge as the software triggers;
- A controller unit with sufficient memory and processing capabilities, and an Ethernet board. On one side, the controller supervises the DAQ unit: it defines the test

waveform to be generated as a sample series at the given sampling rate, stores the acquired samples, and processes them in order to estimate (in quasi real-time) the DAC phase offset. On the other side, the controller is responsible for publishing the SV data packets, from the encapsulation of the SV to the actual transmission through a dedicated Ethernet board.

The DAC outputs a low-voltage sinusoidal signal, in the range of ± 2 V. The amplitude, frequency and initial phase of the signal can be customized to specific test conditions. The conversion to the current levels expected by the N channel of the measuring bridge is carried out by a transconductance amplifier. In this sense, the amplifier ratio represents a further degree of freedom in view of a finer control of the current level, and thus of the excitation. The transconductance amplifier is not an ideal current source and introduces non-negligible uncertainty contributions on both the amplitude and phase of the signal supplied to the measuring bridge.

The ADC re-acquires the amplifier output by means of a calibrated high-precision shunt whose input range is suitably adapted to the specific test configuration. Typically, the shunt output is scaled such that a full input range corresponds to an output range of 0.8 V. Given the calibration context and the high-accuracy of the employed shunts, their contribution in terms of amplitude and phase uncertainty can be reasonably considered as negligible, as further discussed in the next section.

The time-series acquired at the two ADC input channels are processed via a DFT-based routine (further details in [18]) and the phase associated with the fundamental frequency is retrieved. In particular, channel ai1 is representative of the contribution of ADC only, whereas channel ai0 is representative of the entire measurement chain. By properly differentiating these terms, it is possible to define the actual phase of the signal supplied to the measuring bridge.

In the top-centre part of the scheme, a pair of Digital Voltmeters (DVMs) monitors the input and output signal of the series of transconductance amplifier and shunt. The DVMs are employed as high-precision sampling systems that operate in simultaneous mode: the acquired time series are processed via a sine fitting method that allows for accurately estimating the amplitude, frequency and initial phase of the signals under analysis. The DVMs' trigger is not synchronous with the PPS of the synchronization unit, neither is it disciplined to the time reference. As a consequence, the phase information cannot be related to the phase measured on the calibrator. Nevertheless, the difference between the phase measured on each DVM allows for quantifying the phase offset introduced by the amplifier only. It is therefore an independent method to validate the results of DFT-based routine carried out on the re-acquired waveforms.

METAS Implementation

As implemented in the METAS laboratories, the different components of the measurement setup are listed here below.

- Time reference: A Meinberg LANTIME M600Time Server (Meinberg Funkuhren, Bad Pyrmont, Germany) that includes a GPS-disciplined 10-MHz clock, whose time accuracy and frequency stability are in the order of 50 ns and 0.5 nHz/Hz over an averaging time of 1800 s, respectively [17].
- Calibrator: An NI PXIe 1062 chassis (National Instruments, Austin, TX, USA) that hosts three boards: the NI PXIe 8880 controller, the NI PXI 6683 timing and synchronization module, and the NI PXI 4461 dynamic signal acquisition module. The NI PXIe 8880 is an Intel Xeon embedded controller (2.3 GHz Eight-Core) with two 10/100/1000BASE-TX (Gigabit) Ethernet ports. The NI PXI 6683 can generate events and clock signals at specified synchronized future times and timestamp input events with the synchronized system time. The NI PXI 4461 is a 2-input/2-output DAQ with a nominal resolution of 24 bits. The sampling rate and the vertical range are set equal to 192 kHz and ± 2 V, respectively, for both DAC and ADC channels.

- Transconductance amplifier: A Clarke-Hess 8100 (Clarke-Hess, Medford, NY, USA) characterized by a 50 ppm short-term stability, a maximum compliance voltage of 7 V, a total harmonic distortion lower than -60 dB up to 10 kHz, and six available output range from 200 mA to 100 A.
- Shunt: A set of Fluke A40B Precision DC and AC Current Shunts (Fluke, Norwich, UK) with a worst-case uncertainty of 55 ppm up to 1 kHz signal frequency. In particular, we adapted the shunt input range to the generated current level: a 500-mA shunt for $50 \leq I_N < 500$ mA, and a 10-A shunt for $1 \leq I_N \leq 10$ A. In this sense, a further improvement might be represented by the adoption of magnetoresistance sensors. Nevertheless, it should be noticed that the shunts are periodically calibrated and thus the non-ideal conversion ratio is suitably compensated, and its effect is negligible if compared to the calibrator and amplifier ones.
- Digital Voltmeters: A pair of Keysight 3458A Multimeters (Keysight Technologies, Santa Rosa, CA, USA) with a resolution of 8.5 digits and an accuracy of 100 ppm in synchronous mode. The DVMs are used as digitizers for a sine fitting technique with a sampling rate of 2.5 kHz and an aperture time of 920 μ s. The two digitizers operate in a master–slave configuration: the one connected to the calibrator output triggers also the acquisition of the one connected to the amplifier output. In this way, they mimic a two-channel ADC operating in synchronous sampling mode, as further discussed in [23]. It is also worth noticing that the sine fitting procedure returns the value of the amplitude, frequency and initial phase of the fundamental component. Therefore, by comparing the phase at the two channels, we can retrieve the phase contribution of the amplifier only with a standard uncertainty of 0.03 μ rad [24].
- Measuring bridge (DUT): A ZERA WM3000I (ZERA, Königswinter, Germany) with a current input range from 1 mA to 15 A. In non-conventional mode, the bridge guarantees a ratio and phase uncertainty not larger than 300 ppm and 1.5 min, respectively.

4. Uncertainty Contributions

In this section, we analyse the main uncertainty sources inherent in the proposed measurement setup and we derive a complete uncertainty budget based on technical specifications and statistical analysis (In case of statistical analysis, a coverage factor $k = 2$ (i.e., 95%) has been applied to the standard deviations).

In this context, four main contributions can be identified. Three descend from the measurement chain for the generation and re-acquisition of the analog test waveform, i.e., from the DAQ module, the transconductance amplifier and (marginally) the current shunt. One contribution, instead, is directly related to the definition of the SV data packets, i.e., to the vertical resolution loss due to quantization effects.

As regards the DAQ module, two synchronization aspects have to be taken into account: the sampling rate and the phase offset introduced by an improper triggering of the DAC and ADC boards. The sampling rate is derived from the internal time base that is disciplined to the external time reference. In our setup, we are able to retrieve the coerced sampling rate on both boards and the discrepancy between nominal and actual sampling rate is equal to 0.3 ppb [18]. Therefore, it is reasonable to say that the sampling rate has a negligible effect on the amplitude, frequency and phase of the generated and re-acquired waveform.

By connecting the channels ao1 and ai0, we were able to quantify also the distortion level introduced by the DAQ module. In the considered configuration, we evaluated a worst-case Signal-to-Noise Ratio (SNR) and a Total Harmonic Distortion of 92 and -96 dB, respectively. As a consequence, the effective number of bits is equal to 17 bits. In this respect, it should be noticed that such levels of accuracy require a precise control of temperature and power supply stability. In our case, the measurement campaign has been carried out in METAS laboratories with a controlled temperature of 23 Celsius degrees and adopting a power supply at 60 Hz for all the instrumentation, i.e., calibrator, amplifier and DUT, to avoid beating effects or interferences. For this analysis, we considered a dataset of 100

independent acquisition with a sample length of 4 s. Moreover, we quantified the purity of the test waveform by means of a nonlinear fit against a single-tone sinusoidal model that produced a Goodness-of-Fit index not lower than 99.7% [18]. Based on these considerations, we quantified the DAQ contribution to the estimation of the current amplitude in terms of the noise variation range. By also taking into account the integral nonlinearity of the ADC board, as characterized in [25], for a test waveform amplitude of 1 V, the uncertainty is 25.12 ppm.

A further validation of this result is provided by the RMS measurements carried out by the first channel of the DVM system. For this analysis, we considered a dataset of 200 measurements and evaluated the mean and standard deviation. As shown in Figure 3, the distribution is well approximated by a Gaussian distribution and the uncertainty can be quantified in the worst-case in 23.71 ppm. The peculiar non-monotonic uncertainty trend depends also on the adoption of two current shunts with different input range (namely, 500 mA and 10 A), as previously introduced.

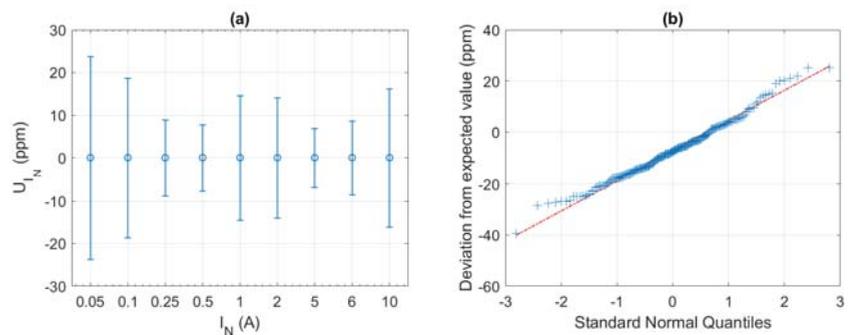


Figure 3. Uncertainty of the current amplitude at the output of the DAQ module as function of the selected current level (a). Quantile-quantile plot relative to a current level of 5 A (b).

The second synchronization aspect regards the triggering mechanism of DAC and ADC boards, with respect to the PPS signal output by the synchronization board (and the time-stamp of the digital data stream). In this regard, it should be noticed that the DAQ module relies on a Sigma-Delta technology: the phase offset introduced by the analog front end of DAC and ADC boards is dependent on the sampling rate. Nevertheless, in a calibration context, it is possible to characterize such contribution and minimize its systematic component by properly shifting the initial phase of the generated waveform. To this end, it is necessary to quantify precisely the phase offset introduced by DAC and ADC boards separately. In recent years, this problem has been widely investigated by several metrological institutes [26–28]. In our measurement setup, we adopted a DFT-based routine for the precise characterization of the phase offset of the signal supplied to the transconductance amplifier. The algorithmic details are beyond the scope of this analysis but can be found in [18]. At 50 Hz, the phase offset has been proven to exhibit a normal distribution with mean and standard deviation equal to 4.186 mrad and 0.004 mrad, respectively. The first one can be seen as a systematic contribution and thus compensated, whereas the second one is a random variable and is related to the phase uncertainty introduced by the DAQ module.

In this context, another aspect that should not be neglected is the proper alignment of the software triggers with the PPS used to synchronize the measuring bridge. With respect to the external time-reference, we quantified the delay introduced by the calibrator in the software triggers and in the PPS output of the synchronization module. As regards the first ones, the technical specifications guarantee the rising edge to occur within 5 ns of the selected time-stamp. Moreover, it should be noticed that the synchronization module guarantees the distribution of software triggers to neighbouring modules (as the DAQ and the controller) with a maximum delay of 2 ns. As regards the PPS output, we employed a

high-precision digital oscilloscope with a sampling rate of 2 GHz and we compared the PPS output against the external time reference. Over an observation interval of nearly 10 min, the PPS showed an average delay of 10 ns with a jitter on the order of few ps. These contributions sum up to 11.36 ns that corresponds to a phase uncertainty of 3.568 μrad .

Once output by the calibrator, the transconductance amplifier converts the voltage test waveform in the corresponding current waveform. In order to characterize the amplitude and phase contributions of this stage, this signal is re-acquired through a high-accuracy current shunt. It is worth noticing that the shunts (periodically calibrated) contribute to the overall uncertainty by at most 0.90 ppm for the amplitude, and 1.50 μrad for the phase [29,30]. On the other hand, the transconductance amplifier has a much more significant impact on the overall uncertainty. As per the calibrator output, we analysed the amplifier output via the DVM. In this case, we were also able to compute the phase displacement between the two channels, i.e., the phase displacement introduced by the transconductance amplifier (and the shunt). For each considered current level, we carried out 200 independent measurements and computed the corresponding statistical distributions: the mean value is taken as a systematic contribution and thus compensated, the standard deviation is used for the uncertainty computation. Figure 4 presents the uncertainty associated with amplitude and phase as function of the current level. In the worst-case, the former is equal to 160.75 ppm, whereas the latter is 200 μrad . In this regard, it is worth noticing how the uncertainty rapidly increases when the current levels fall below 500 mA. Indeed, the selected amplifier is designed for high current output and exhibits a poor accuracy at lower current levels. At the nominal value of 5 A, the uncertainty for amplitude and phase are just 8.36 ppm and 8 μrad , respectively.

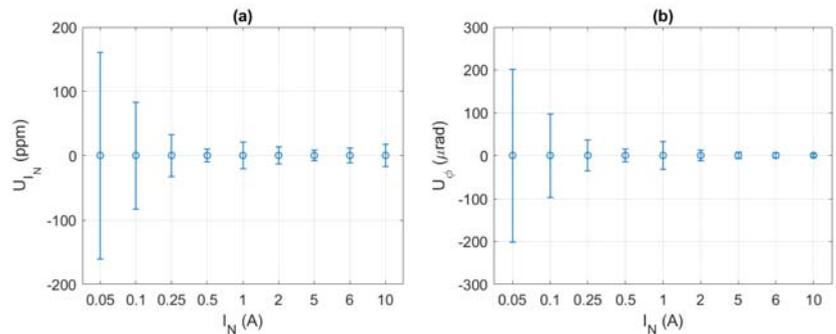


Figure 4. Uncertainty of the current amplitude (a) and phase (b) at the output of the transconductance amplifier as function of the selected current level.

Finally, the contribution of the digital output has been also assessed. The SV communication protocol provides a resolution of 32 bits for the analog converted quantities. In terms of quantization error, this corresponds to an amplitude uncertainty of 67 ppm. In terms of phase uncertainty, it is reasonable to set it equal to zero, as the calibrator outputs simulated packets, and thus no conversion error is possible (differently from the digital stream output by a NCIT or a SAMU where measurement errors might occur). On the other hand, it is difficult to merge such uncertainty contributions with the ones related to the analog measurement chain without knowing the algorithm employed by the measuring bridge for the definition of amplitude and phase on the X-channel. If a DFT-based approach is adopted (as in many other SV-based estimators), the recent literature has proven how the amplitude and phase uncertainty due to quantization errors decreases significantly as the resolution of the quantized samples exceed 14 bits [31]. Therefore, for the purpose of this paper, this contribution can be reasonably considered negligible.

As a summary, in Table 1, we report the overall uncertainty budget for the calibration infrastructure. By applying a conservative approach, the combined uncertainty has been computed under the assumption of independent and uncorrelated contributions. Consis-

tently with the common practice in current transformer calibration, the phase uncertainty has been expressed in minutes. In total, the amplitude and phase uncertainty are lower than 200 ppm and 0.7' for the entire range of considered test conditions.

Table 1. Uncertainty budget for amplitude and phase accuracy (coverage factor $k = 2$).

	Amplitude Uncertainty U_{I_N} (ppm)	Phase Uncertainty U_φ (')
Synchronization module	—	0.012
DAQ module	25.12	0.007
Transconductance amplifier	160.75	0.687
Combined Uncertainty	162.70	0.687

5. Experimental Validation

In this section, we present the results of an experimental validation carried out on the selected DUT, i.e., the ZERA WM3000I. This is intended as an experimental validation of the proposed calibration method as well as of its main uncertainty contributions in controlled laboratory testing conditions, as typical of metrological institute activities. In the following tables, the reported uncertainty takes into account both the calibration setup contribution and the statistical dispersion of the measurements of the device under test. For the sake of readability, a ceiling to the last significant digit has been applied.

For this analysis, we set the nominal frequency, current range and transformer ratio equal to 50 Hz, 5 A and 1:1. Otherwise differently stated, the signals at N - and X -channel consist of single-tone sinusoids, whose frequency, amplitude and initial phase are set equal to 50 Hz, 5 A, and 0 rad, respectively. In the following tests, such parameter values are suitably modified in order to reproduce different configurations of excitation, ratio and phase error, and thus span the entire operating range of the measuring bridge.

To this end, a total of 27 different configurations are taken into account. Each test has a duration of 5 min, including 1 min of settling time to allow for the proper stabilization of the current output of the transconductance amplifier. For each of the monitored quantities, 11 consecutive measurements are taken and their average and standard deviation values are employed to determine the corresponding measurement errors and uncertainties (In the presence of outliers, single measurements could be neglected. In this sense, the outlier detection criterion is based on the assumption that the measurements are normally distributed. Given a set of 11 measurements, if a single measurements differs from the average value by more than three standard deviations, its value is discarded from the computation of measurement errors and uncertainties). In more detail, the reported uncertainties for the excitation and ratio error, and for the phase displacement are obtained by merging the contributions of the calibration setup with the Type A uncertainty of the measuring bridge results.

For the sake of comparison, Table 2 reports the WM3000I specifications in terms of accuracy for the current measurement on the N -channel, the ratio error and the phase displacement. As previously observed, the excitation error in non-conventional mode descends directly from the accuracy of the measured I_N amplitude, as the I_X amplitude depends only on quantization and numerical errors whose impact on the overall uncertainty can be considered as negligible.

Table 2. Specifications of the measuring bridge under test.

Measurement	Accuracy	Current Range
I_N , Normal current RMS	100 ppm	$0.05 < I_s \leq 15$ A
	200 ppm	$0.05 \leq I_s < 0.005$ A
$\Delta\epsilon$, Ratio error	100 ppm	$0.05 < I_s \leq 15$ A
	200 ppm	$0.05 \leq I_s < 0.005$ A
$\Delta\varphi$, Phase displacement	1.1'	$0.05 < I_s \leq 15$ A
	1.5'	$0.05 \leq I_s < 0.005$ A

In this context, Table 3 reports the measurement results in the presence of ratio and phase errors. For this analysis, the ratio error is varied within $\pm 5\%$ and the phase displacement is set in such a way to consider small offsets (e.g., $\pm 10'$), large offsets (e.g., $\pm 5400'$), and nearly phase opposition conditions (e.g., $10,794'$) (Such variations have been obtained by modifying the content of the SV data packets, as the digital channel is characterized by lower uncertainty contributions. Nevertheless, similar results could be obtained by keeping unaltered the SV data packets and suitably modifying the current source flowing through the standard transformer).

The two phenomena are investigated both independently and simultaneously. In this way, it is possible to evaluate whether the measuring bridge is affected by any of the error source or by their combination.

As the excitation is kept equal to 100%, it is worth noticing as the measuring bridge exhibits an excitation error perfectly in line with its specifications and the uncertainty does not exceed 200 ppm. Similar considerations apply for ratio error and phase displacement. In this case, it is interesting to notice how $\Delta\epsilon$ and $\Delta\varphi$ do not exhibit any dependence on the test setting.

Table 3. Characterization of the measuring bridge performance in the presence of ratio errors and phase displacements (coverage factor $k = 2$).

Settings					Measurements			Uncertainty		
I_N (A)	I_X (A)	E (%)	ϵ (%)	φ (')	ΔE (%)	$\Delta\epsilon$ (%)	$\Delta\varphi$ (')	$U_{\Delta E}$ (%)	$U_{\Delta\epsilon}$ (%)	$U_{\Delta\varphi}$ (')
5	5	100	-5.000	0.00	0.00	-0.003	-0.9	0.02	0.012	1.0
5	5	100	-3.000	0.00	0.00	-0.004	-0.9	0.02	0.012	1.0
5	5	100	-0.200	0.00	0.00	-0.004	-0.9	0.02	0.012	1.0
5	5	100	-0.200	-10.00	0.01	-0.006	-0.9	0.02	0.012	1.0
5	5	100	-0.200	10.00	0.01	-0.006	-0.9	0.02	0.012	1.0
5	5	100	0.000	-5400.00	0.01	-0.005	-0.9	0.02	0.012	1.0
5	5	100	0.000	-180.00	0.01	-0.004	-0.9	0.02	0.012	1.0
5	5	100	0.000	-1.00	0.01	-0.005	-0.9	0.02	0.012	1.0
5	5	100	0.000	0.00	0.01	-0.005	-0.9	0.02	0.012	1.0
5	5	100	0.000	1.00	0.01	-0.006	-0.9	0.02	0.012	1.0
5	5	100	0.000	180.00	0.01	-0.005	-0.9	0.02	0.012	1.0
5	5	100	0.000	5400.00	0.01	-0.005	-0.9	0.02	0.012	1.0
5	5	100	0.000	10,794.00	0.01	-0.005	-0.9	0.02	0.012	1.0
5	5	100	0.200	0.00	0.01	-0.005	-0.9	0.02	0.012	1.0
5	5	100	0.200	-10.00	0.01	-0.005	-0.9	0.02	0.012	1.0
5	5	100	0.200	10.00	0.01	-0.006	-0.9	0.02	0.012	1.0
5	5	100	3.000	0.00	0.01	-0.004	-0.9	0.02	0.012	1.0
5	5	100	5.000	0.00	0.01	-0.005	-0.8	0.02	0.012	1.0

In Table 4, we report the measuring bridge errors and uncertainties in the presence of different excitation levels. For this analysis, we modified the N -channel current in

such a way that the measuring bridge senses an excitation between 1 and 200%. Once more, the excitation error and the corresponding uncertainty are in line with the previous considerations. As regards ratio error and phase displacement, it is worth noticing how both measurements and uncertainties present a rapid increase as the excitation falls below 5%. Nevertheless, it is reasonable to expect that, in the presence of lower current levels, the accuracy of the internal sensors as well as the SNR decrease and the corresponding computations are affected by larger errors and uncertainties. Similar considerations hold also for the digital counterpart. The SV data format has a fixed range and number of bits: as a consequence, when transmitting low-amplitude signals, there is an inefficient exploitation of the 32 bits and the resulting estimates are likely to be affected by higher relative uncertainty.

Table 4. Characterization of the measuring bridge performance in the presence of different excitation levels (coverage factor $k = 2$).

		Settings			Measurements			Uncertainty		
I_N	I_X	E	ε	φ	ΔE	$\Delta\varepsilon$	$\Delta\varphi$	$U_{\Delta E}$	$U_{\Delta\varepsilon}$	$U_{\Delta\varphi}$
(A)	(A)	(%)	(%)	($^\circ$)	(%)	(%)	($^\circ$)	(%)	(%)	($^\circ$)
5	5	1	0.000	0.00	0.00	−0.017	−1.1	0.02	0.028	2.0
5	5	2	0.000	0.00	0.00	0.013	−1.0	0.02	0.017	1.5
5	5	5	0.000	0.00	0.00	−0.009	−1.0	0.02	0.013	1.0
5	5	10	0.000	0.00	0.00	−0.007	−1.1	0.02	0.012	1.0
5	5	20	0.000	0.00	0.00	−0.009	−1.0	0.02	0.012	1.0
5	5	50	0.000	0.00	0.00	−0.006	−0.9	0.02	0.012	1.0
5	5	100	0.000	0.00	0.01	−0.004	−0.9	0.02	0.012	1.0
5	5	120	0.000	0.00	0.01	−0.004	−0.9	0.02	0.012	1.0
5	5	200	0.000	0.00	0.01	−0.001	−1.0	0.02	0.012	1.0

The specific device under test provides a useful extra feature, i.e., a representation of the current flowing in the N - and X -channel as rotating vectors characterized in terms of RMS amplitude, phase, and frequency. The estimation accuracy of the first two parameters has been already investigated in the previous tables, but the frequency (particularly, the one of the N -channel) requires a separate investigation. To this end, we characterized the frequency measurements in the presence of different excitation levels and phase displacements. For the sake of consistency, the variation ranges of E and φ correspond to the ones applied in Tables 3 and 4.

In this context, Table 5 reports the measurement results and the associated uncertainty. It is worth noticing how the frequency error Δf is quite stable around -0.54 mHz with a worst-case uncertainty of 0.06 mHz (when the excitation is set to its minimum value, i.e., 1%). In this case, the instrument specifications do not provide a performance target. Nevertheless, the obtained measurement accuracy is sufficient for the typical application of a measuring bridge for instrument transformers.

Table 5. Characterization of the measuring bridge frequency estimation accuracy in the presence of different excitation levels and phase displacements (coverage factor $k = 2$).

I_N (A)	I_X (A)	Settings				f (Hz)	Measurements Δf (mHz)	Uncertainty $U_{\Delta f}$ (mHz)
		E (%)	ε (%)	φ (°)				
5	5	1	0.000	0.00	50	−0.511	0.056	
5	5	2	0.000	0.00	50	−0.512	0.038	
5	5	5	0.000	0.00	50	−0.556	0.028	
5	5	10	0.000	0.00	50	−0.519	0.030	
5	5	20	0.000	0.00	50	−0.537	0.031	
5	5	50	0.000	0.00	50	−0.537	0.032	
5	5	100	0.000	0.00	50	−0.509	0.028	
5	5	100	0.000	−5400.00	50	−0.529	0.032	
5	5	100	0.000	−180.00	50	−0.557	0.028	
5	5	100	0.000	−1.00	50	−0.521	0.030	
5	5	100	0.000	0.00	50	−0.530	0.032	
5	5	100	0.000	1.00	50	−0.530	0.032	
5	5	100	0.000	180.00	50	−0.548	0.030	
5	5	100	0.000	5400.00	50	−0.548	0.032	
5	5	100	0.000	10,794.00	50	−0.548	0.030	
5	5	120	0.000	0.00	50	−0.528	0.032	
5	5	200	0.000	0.00	50	−0.511	0.028	

6. Conclusions

In this paper, we presented the measurement setup for the calibration of measurement bridges for non-conventional instrument transformers. We have discussed the main implementation challenges and characterized the most significant uncertainty contributions. Based on technical specifications and statistical analysis, we have performed a comprehensive uncertainty budget of the calibration setup that has been further validated by an experimental measurement campaign carried out at METAS laboratories.

The proposed analysis allows for identifying the main challenges of a calibration process that requires a synchronous generation of both analog and digital quantities. The research project will now focus on the minimization of the uncertainty contributions (with specific attention to the analog measurement chain) and on the extension of the proposed infrastructure to non-stationary signals, as the ones that a plausible instrument transformer might deal with in field applications.

Author Contributions: Conceptualization, G.F.; Formal analysis, G.F.; Investigation, G.F.; Methodology, G.F.; Supervision, M.A.; Writing—original draft, G.F. All authors have read and agreed to the published version of the manuscript.

Funding: This project (17IND06 FutureGrid II) has received funding from the EMPIR programme co-financed by the Participating States and from the European Union’s Horizon 2020 research and innovation programme.



**Co-financed by the Connecting Europe
Facility of the European Union**

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Terzija, V.; Valverde, G.; Cai, D.; Regulski, P.; Madani, V.; Fitch, J.; Skok, S.; Begovic, M.M.; Phadke, A. Wide-Area Monitoring, Protection, and Control of Future Electric Power Networks. *Proc. IEEE* **2011**, *99*, 80–93. [[CrossRef](#)]
2. Blaabjerg, F.; Yang, Y.; Yang, D.; Wang, X. Distributed Power-Generation Systems and Protection. *Proc. IEEE* **2017**, *105*, 1311–1331. [[CrossRef](#)]
3. Paolone, M.; Gaunt, T.; Guillaud, X.; Liserre, M.; Meliopoulos, S.; Monti, A.; Van Cutsem, T.; Vittal, V.; Vournas, C. Fundamentals of power systems modelling in the presence of converter-interfaced generation. *Electr. Power Syst. Res.* **2020**, *189*, 106811. [[CrossRef](#)]
4. Frigo, G.; Pegoraro, P.; Toscani, S. Low-Latency, Three-Phase PMU Algorithms: Review and Performance Comparison. *Appl. Sci.* **2021**, *11*, 2261. [[CrossRef](#)]
5. Frigo, G.; Derviškić, A.; Zuo, Y.; Paolone, M. PMU-Based ROCOF Measurements: Uncertainty Limits and Metrological Significance in Power System Applications. *IEEE Trans. Instrum. Meas.* **2019**, *68*, 3810–3822. [[CrossRef](#)]
6. Karpilow, A.C.; Derviskadic, A.; Frigo, G.; Paolone, M. Characterization of Non-Stationary Signals in Electric Grids: A Functional Dictionary Approach. *IEEE Trans. Power Syst.* **2021**, *37*, 1–13. [[CrossRef](#)]
7. Rietveld, G.; Wright, P.S.; Roscoe, A.J. Reliable Rate-of-Change-of-Frequency Measurements: Use Cases and Test Conditions. *IEEE Trans. Instrum. Meas.* **2020**, *69*, 6657–6666. [[CrossRef](#)]
8. Ingram, D.M.E.; Schaub, P.; Campbell, D.A. Use of Precision Time Protocol to Synchronize Sampled-Value Process Buses. *IEEE Trans. Instrum. Meas.* **2012**, *61*, 1173–1180. [[CrossRef](#)]
9. IEEE/IEC International Standard-Instrument transformers-Part 9: Digital interface for instrument transformers. In *IEC/IEEE 61869-9:2016*; IEC: Geneva, Switzerland, 2016; pp. 1–127. [[CrossRef](#)]
10. Mohns, E.; Mortara, A.; Cayci, H.; Houtzager, E.; Fricke, S.; Agustoni, M.; Ayhan, B. Calibration of Commercial Test Sets for Non-Conventional Instrument Transformers. In Proceedings of the 2017 IEEE International Workshop on Applied Measurements for Power Systems (AMPS), Liverpool, UK, 20–22 September 2017; pp. 1–6.
11. IEC/IEEE International Standard-Communication networks and systems for power utility automation—Part 9-2: Specific Communication Service Mapping (SCSM) Sampled values over ISO/IEC 8802-3. In *IEC 61850-9-2:2011+AMD1:2020 CSV Consolidated Version 2020-02*; IEC: Geneva, Switzerland, 2011; pp. 1–18. [[CrossRef](#)]
12. Castello, P.; Ferrari, P.; Flammini, A.; Muscas, C.; Rinaldi, S. An IEC 61850-Compliant distributed PMU for electrical substations. In Proceedings of the 2012 IEEE International Workshop on Applied Measurements for Power Systems (AMPS) Proceedings, Aachen, Germany, 26–28 September 2012; pp. 1–6.
13. Giustina, D.D.; Dedè, A.; Invernizzi, G.; Valle, D.P.; Franzoni, F.; Pegoiani, A.; Cremaschini, L. Smart Grid Automation Based on IEC 61850: An Experimental Characterization. *IEEE Trans. Instrum. Meas.* **2015**, *64*, 2055–2063. [[CrossRef](#)]
14. Blair, S.M.; Roscoe, A.J.; Irvine, J. Real-time compression of IEC 61869-9 sampled value data. In Proceedings of the 2016 IEEE International Workshop on Applied Measurements for Power Systems (AMPS), Aachen, Germany, 28–30 September 2016; pp. 1–6. [[CrossRef](#)]
15. Frigo, G.; Agustoni, M. Phasor Measurement Unit and Sampled Values: Measurement and Implementation Challenges. In Proceedings of the 2021 IEEE International Workshop on Applied Measurements for Power Systems (AMPS), Cagliari, Italy, 29 September–1 October 2021; pp. 1–6.
16. Kariyawasam, S.; Rajapakse, A.D.; Perera, N. Investigation of Using IEC 61850-Sampled Values for Implementing a Transient-Based Protection Scheme for Series-Compensated Transmission Lines. *IEEE Trans. Power Deliv.* **2018**, *33*, 93–101. [[CrossRef](#)]
17. Agustoni, M.; Mortara, A. A Calibration Setup for IEC 61850-9-2 Devices. *IEEE Trans. Instrum. Meas.* **2017**, *66*, 1124–1130. [[CrossRef](#)]
18. Agustoni, M.; Frigo, G. Characterization of DAC Phase Offset in IEC 61850-9-2 Calibration Systems. *IEEE Trans. Instrum. Meas.* **2021**, *70*, 1–10. [[CrossRef](#)]
19. Djokić, B.; Parks, H. A Synchronized Current-Comparator Bridge for the Calibration of Analog Merging Units. *IEEE Trans. Instrum. Meas.* **2019**, *68*, 1955–1960. [[CrossRef](#)]
20. Roccatò, P.; Sardi, A.; Pasini, G.; Peretto, L.; Tinarelli, R. Traceability of MV low power instrument transformer LPIT. In Proceedings of the 2013 IEEE International Workshop on Applied Measurements for Power Systems (AMPS), Aachen, Germany, 25–27 September 2013; pp. 13–18. [[CrossRef](#)]
21. Thomas, R.; Vujanic, A.; Xu, D.Z.; Sjödin, J.E.; Salazar, H.R.M.; Yang, M.; Powers, N. Non-conventional instrument transformers enabling digital substations for future grid. In Proceedings of the 2016 IEEE/PES Transmission and Distribution Conference and Exposition (T D), Dallas, TX, USA, 3–5 May 2016; pp. 1–5. [[CrossRef](#)]
22. Souders, T.M. A Wide Range Current Comparator System for Calibrating Current Transformers. *IEEE Trans. Power Appar. Syst.* **1971**, *PAS-90*, 318–324. [[CrossRef](#)]
23. Mester, C. Timestamping Type 3458A Multimeter Samples. In Proceedings of the 2018 Conference on Precision Electromagnetic Measurements (CPEM 2018), Paris, France, 8–13 July 2018; pp. 1–2. [[CrossRef](#)]
24. Mester, C. Sampling Primary Power Standard from DC up to 9 kHz Using Commercial Off-The-Shelf Components. *Energies* **2021**, *14*, 2203. [[CrossRef](#)]
25. Overney, F.; Rufenacht, A.; Braun, J.P.; Jeanneret, B.; Wright, P.S. Characterization of Metrological Grade Analog-to-Digital Converters Using a Programmable Josephson Voltage Standard. *IEEE Trans. Instrum. Meas.* **2011**, *60*, 2172–2177. [[CrossRef](#)]

26. Braun, J.P. Measure of the Absolute Phase Angle of a Power Frequency Sinewave with Respect to UTC. In Proceedings of the 2018 Conference on Precision Electromagnetic Measurements (CPEM 2018), Paris, France, 8–13 July 2018; pp. 1–2. [[CrossRef](#)]
27. Crotti, G.; Femine, A.D.; Gallo, D.; Giordano, D.; Landi, C.; Luiso, M. Measurement of Absolute Phase Error of Digitizers. In Proceedings of the 2018 Conference on Precision Electromagnetic Measurements (CPEM 2018), Paris, France, 8–13 July 2018; pp. 1–2. [[CrossRef](#)]
28. Goldstein, A.; Frigo, G. *Determination of Digitizer Absolute Phase Using Equivalent Time Sampling*; IEEE: Piscataway, NJ, USA, 2021.
29. Budovsky, I. Measurement of Phase Angle Errors of Precision Current Shunts in the Frequency Range From 40 Hz to 200 kHz. *IEEE Trans. Instrum. Meas.* **2007**, *56*, 284–288. [[CrossRef](#)]
30. Bosco, G.C.; Garcocz, M.; Lind, K.; Pogliano, U.; Rietveld, G.; Tarasso, V.; Voljc, B.; Zachovalová, V.N. Phase Comparison of High-Current Shunts up to 100 kHz. *IEEE Trans. Instrum. Meas.* **2011**, *60*, 2359–2365. [[CrossRef](#)]
31. Betta, G.; Liguori, C.; Pietrosanto, A. Propagation of uncertainty in a discrete Fourier transform algorithm. *Measurement* **2000**, *27*, 231–239. [[CrossRef](#)]

Article

Digital Representation of Measurement Uncertainty: A Case Study Linking an RMO Key Comparison with a CIPM Key Comparison

Blair D. Hall * and Annette Koo

Measurement Standards Laboratory of New Zealand, Lower Hutt 5010, New Zealand;
annette.koo@measurement.govt.nz

* Correspondence: blair.hall@measurement.govt.nz

Abstract: This paper considers a future scenario in which digital reporting of measurement results is ubiquitous and digital calibration certificates (DCCs) contain information about the components of uncertainty in a measurement result. The task of linking international measurement comparisons is used as a case study to look at the benefits of digitalization. Comparison linking provides a context in which correlations are important, so the benefit of passing a digital record of contributions to uncertainty along a traceability chain can be examined. The International Committee for Weights and Measures (CIPM) uses a program of international “key comparisons” to establish the extent to which measurements of a particular quantity may be considered equivalent when made in different economies. To obtain good international coverage, the results of the comparisons may be linked together: a number of regional metrology organization (RMO) key comparisons can be linked back to an initial CIPM key comparison. Specific information about systematic effects in participants’ results must be available during linking to allow correct treatment of the correlations. However, the conventional calibration certificate formats used today do not provide this: participants must submit additional data, and the report of an initial comparison must anticipate the requirements for future linking. Special handling of additional data can be laborious and prone to error. An uncertain-number digital reporting format was considered in this case study, which caters to all the information required and would simplify the comparison analysis, reporting, and linking; the format would also enable a more informative presentation of comparison results. The uncertain-number format would be useful more generally, in measurement scenarios where correlations arise, so its incorporation into DCCs should be considered. A full dataset supported by open-source software is available.

Citation: 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>

Academic Editor: Simona Salicone

Received: 16 September 2021

Accepted: 1 December 2021

Published: 6 December 2021

Keywords: digitalization; measurement uncertainty; metrological traceability; key comparison; digital calibration certificate; uncertain number

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The national and international infrastructures that disseminate critical measurement information throughout society are due for renovation. Designed to be operated and supervised by skilled people, there is now a call to digitalize these essentially paper-based systems. The best way to proceed is by no means clear. However, a coordinated international effort will be needed to reap real benefits from digitalization. A recent paper reviewed work performed so far to develop a common digital format for reporting measurement data, which is generically referred to as a “digital calibration certificate” (DCC) [1]. The DCC will be a fundamental component of digital measurement infrastructures, but many decisions still need to be made about its structure. Among these is how best to represent measurement uncertainty.

The International Committee for Weights and Measures (CIPM), which directs metrological activities carried out by parties to the Metre Convention [2], recognized the need to

coordinate digitalization of the international measurement system and has established a CIPM Task Group on the Digital SI (CIPM-TG-DSI) [3], supported by a team of experts. In early 2021, this team made a public request for use cases to identify situations where the digitalization of existing metrological infrastructure might improve outcomes or address difficulties. This paper presents a preliminary analysis of one of those use cases.

The study relates to the analysis and linking of international measurement comparisons. Although comparison analysis is a specialized topic, the limitations of current reporting in calibration certificates become clear in this context, and so, the benefits of digitalization are easily recognized. Comparison linking is an interesting case, because correlations in the data can significantly affect the results. Handling the correlations complicates data analysis, especially the evaluation of uncertainty. Our study shows that digitalization can help; it can enhance the information available in the final results while hiding and largely automating the more laborious aspects of processing data. The challenge posed by correlated data arises in many other measurement scenarios as well, so the representation of measurement uncertainty described in this work would offer advantages in other digital systems.

1.1. Measurement Comparisons in the CIPM MRA

The CIPM Mutual Recognition Arrangement (MRA) [4] is a framework used to establish the equivalence of measurement standards in different economies. Specific calibration and measurement capability (CMC) claims are approved by expert Consultative Committees of the CIPM and then published in a database by the International Bureau of Weights and Measures (BIPM). To maintain or extend CMC entries, national metrology institutes (NMIs) must provide evidence in support of their claims. This evidence is often obtained by participating in international measurement comparisons [5].

Our case study involves two kinds of comparison: a CIPM key comparison and a subsequent RMO key comparison (organized by a regional metrology organization). In a CIPM comparison, a group of NMIs submit measurements of a particular quantity associated with an artifact. The data are used to determine a comparison reference value, and then, for each participant, a degree of equivalence (DoE) is calculated, which characterizes the difference between the participant's result and the comparison reference value.

After an initial CIPM comparison has been completed, a number of other RMO comparisons may be carried out. This provides a way to assess the equivalence of NMIs that did not participate in the initial comparison. The results of an RMO comparison must be linked to those of the initial CIPM comparison, which means that several participants from the initial comparison must participate again in the RMO comparison.

A DoE is considered to reflect the level of consistency of one participant's measurement standard with those of other participants. An uncertainty is evaluated for each DoE, which allows the significance of each result to be assessed: if the magnitude of a DoE is greater than its expanded uncertainty (typically at a 95% level of confidence), then the evidence for equivalence is considered weak. DoEs evaluated during an RMO comparison have equal standing to DoEs obtained from the initial CIPM comparison.

1.2. CCPR Comparison Analysis

Measurement comparisons in the CIPM MRA follow strict rules and are guided by the policies of the Consultative Committee responsible for a particular technical area [5]. Our case study deals with the photometric quantity regular spectral transmittance, which falls under the Consultative Committee for Photometry and Radiometry (CCPR). A detailed description of the analysis recommended for CCPR comparisons was given in [6], where expressions for the uncertainty in DoE values were obtained according to the law of the propagation of uncertainty (LPU) [7]. In practice, these expressions have many terms, which can make data processing quite daunting.

1.3. A Case Study of Comparison Linking

The intention of this case study is to look for possible benefits from digitalization of the reporting and analysis of data. A future scenario is envisaged, where participants submit results in a digital format that contains more information than today's calibration certificates. This allows data processing to be handled more directly and in a straightforward and intuitive manner. The scenario offers a glimpse into the future where, once the digitalization of the international measurement system is complete, digital reporting (DCCs) will be ubiquitous.

The study applies the methodology prescribed in [6], but a digital format called an “uncertain number” is used to represent the data [8]. An uncertain number is a data type that encapsulates information about the measured value of a quantity and the components of uncertainty in that value. Software supporting uncertain numbers greatly simplifies data processing, because calculations simultaneously evaluate the value and the components of uncertainty. Mathematical operations are expressed in terms of a value calculation, but the results include a complete uncertainty budget. Furthermore, uncertain-number results are transferable, which is extremely important in this work. (The *Guide to the expression of uncertainty in measurement* (GUM) identifies transferability and internal consistency as the desirable properties of an ideal method of expressing uncertainty ([7], §0.4). Although we refer here only to transferability for simplicity, uncertain numbers provide both transferability and internal consistency.) An uncertain number obtained as the result of some calculation may be used immediately as an argument in further calculations (exactly as one can do with numerical results). When this happens, the components of uncertainty are rigorously propagated, from one intermediate result to the next, according to the LPU. Transferability in this case study allows the CIPM and RMO comparisons, with linking, to be processed as a single, staged, measurement (the importance of adequately linking the stages of a metrological traceability chain was discussed in [9,10]). It is this aspect of digitalization that delivers the benefits we describe below.

A software tool called the GUM Tree Calculator (GTC) that implements the uncertain-number approach was used. GTC is an open-source Python package [11,12]. A recent publication described GTC and its design in some detail [13]. A dataset containing the data and code used in the current study is available [14]. The snippets of code shown below are extracts from this dataset.

1.4. Digital Records

To create digital records for the participant and pilot measurements in this study, a small subset of data was taken from a CIPM comparison of transmittance and a subsequent RMO comparison [15,16]. In Sections 2 and 3, we describe the structure of these comparisons. Participants were required to submit an uncertainty budget for each measurement and to identify the systematic and random influence factors in that budget. The systematic factors are considered constant. For NMIs that participated in both CIPM and RMO comparisons, the systematic factors do not change. They are characterized as components of uncertainty, because the actual values of residual error are not known. The random factors are considered to be unpredictable effects that arise independently in each measurement. The nature of the components of uncertainty—systematic or random—must be known in order to account for correlations in the data.

We used the uncertainty budgets reported by participants to construct digital records for this study. In doing so, some assumptions were made about the data and some of the data were changed to resolve minor inconsistencies, so we do not identify actual participants with these records. The intention here is to present a future scenario where DCC formats have been widely adopted. The assumption was made that these formats are self-contained, with more detailed information than is available in today's calibration certificates, so there is no longer a need to request additional data for the comparison analysis. Were such a future to become reality, the processes leading to the production of

DCCs would not resemble the steps taken here to artificially create the scenario. Therefore, the detail of how digital records were assembled for this study is not discussed.

1.5. Mathematical Notation

Mathematical expressions use the notation adopted in [6]. Most details are explained when the notation first appears in the text. However, the reader should note that we distinguish between quantities and estimates of quantities with upper and lower case symbols, respectively. For instance, the uncertainty in a value y , obtained by measuring a quantity Y , will be expressed as $u(y)$ —the standard uncertainty of y as an estimate of Y . When GTC code is used to implement mathematical expressions, uncertain-numbers are associated with quantity terms (upper case terms). The corresponding estimates and uncertainties are the properties of these uncertain-number objects.

2. A CIPM Key Comparison

In the initial CIPM key comparison, there were eleven participants (identified here by the letters A, B, . . . , K) and a pilot laboratory (Q). Each participant measured a particular artifact, while the pilot measured all eleven artifacts. The comparison was carried out in five stages: first, the pilot measured the artifacts; second, each participant reported a measurement; third, the pilot measured the artifacts again; fourth, each participant made a second measurement; and fifth, the pilot made one last measurement of all the artifacts (Two participants submitted only the second measurement, so these data were processed with pilot results from only Stages 1 and 3).

Listing 1 displays information about the first measurement by Participant A (Stage 2). The measured transmittance appears at the top, with the combined standard uncertainty in parentheses. Two uncertainty budgets follow: first, the individual components of uncertainty; second, the net systematic and random effects components. Component labeling uses a capital letter to identify the participant (A, B, etc.). If a component of uncertainty contributed only to a specific stage, then a stage number (1, 2, 3, 4, or 5) is appended in parentheses. A colon then precedes the participant's name for the influence quantity, and finally, the component is classified as random or systematic (*rnd* or *sys*). For example, there are both random and systematic contributions to uncertainty in the wavelength, so Listing 1 includes two terms: `A:WaveLength (sys)` is a systematic component that contributes to uncertainty at every stage, and `A(2):WaveLength (ran)` is a random component that contributes at Stage 2 (another independent component `A(4):WaveLength (ran)` appears in the budget at Stage 4). It is important to understand that the information shown in Listing 1 was all obtained from a single entity representing the measurement result—a single uncertain number. In the scenario we considered, this was submitted by the participant in a digital record.

When mathematical operations are applied to uncertain numbers, the components of uncertainty are handled according to the LPU. To illustrate this, we compared the results submitted by Participant A at Stages 2 and 4 by subtracting the corresponding uncertain numbers. With `Y_A_2` and `Y_A_4` for the results, we display the uncertain-number difference in Listing 2. Notice that the only non-zero terms in the uncertainty budget are now associated with random effects at each stage. The systematic terms from the budget of Listing 1 (the non-linearity, wavelength, stray light, and the beam size and position) contribute nothing to the combined uncertainty in the difference. This is to be expected, because each systematic term contributes a fixed (albeit unknown) amount to the combined measurement error. The influence of these constant terms on the difference is zero. Uncertain-number calculations arrive at the correct result by strictly implementing the LPU. In order to do that, information about all uncertainty components must be encapsulated in the uncertain-number data.

Listing 1. Data from Participant A for the Stage-2 measurement. The measured value is shown at the top, with the combined standard uncertainty in parentheses. Two uncertainty budgets follow. The first shows the individual components of uncertainty reported by the participant. The second shows total systematic and random components.

```
A(2):Transmittance = 0.919644(0.000296)

Uncertainty budget:
  A:Beam Size & Position (sys): 0.00019516
  A(2):Beam Size & Position (ran): 0.00019516
    A(2):Type-A (ran): 0.00007800
  A(2):Instability (ran): 0.00006800
  A:Non-linearity (sys): 0.00003000
  A:Wavelength (sys): 0.00000354
  A(2):Wavelength (ran): 0.00000354
  A:Stray Light (sys): 0.00000300

Systematic and random:
  A(2) (ran): 0.00022093
  A (sys): 0.00019751
```

Listing 2. The difference between Participant A’s results at Stage 2 and Stage 4. The difference is shown at the top, with the combined standard uncertainty in parentheses. The uncertainty budget follows. Note that all systematic components are now zero.

```
Difference:Y_A_4 - Y_A_2 = 0.000449(0.000489)

Uncertainty budget:
  A(4):Type-A (ran): 0.00038400
  A(2):Beam Size & Position (ran): 0.00019516
  A(4):Beam Size & Position (ran): 0.00019516
    A(2):Type-A (ran): 0.00007800
  A(2):Instability (ran): 0.00006800
  A(4):Instability (ran): 0.00006800
  A(2):Wavelength (ran): 0.00000354
  A(4):Wavelength (ran): 0.00000354
  A:Non-linearity (sys): 0.00000000
  A:Wavelength (sys): 0.00000000
  A:Stray Light (sys): 0.00000000
  A:Beam Size & Position (sys): 0.00000000
```

2.1. Evaluating DoEs

The calculation of DoEs can be expressed succinctly (in the notation adopted in [6]). For the pilot, identified by the letter “Q” (and with a superscript “*” to indicate a CIPM comparison), the DoE, D_Q^* , is a weighted sum over all participants ([6], Equation (18)) (the weighting factors w_j are explained in Appendix A and the notation $\langle \cdot \rangle_{A_j}$ is the mean of measurements of the artifact associated with participant j):

$$D_Q^* = - \sum_j w_j \langle \bar{Y}_j^* - \bar{Y}_Q^* \rangle_{A_j} . \tag{1}$$

For any other participant i , the DoE is ([6], Equation (19)):

$$D_i^* = \langle \bar{Y}_i^* - \bar{Y}_Q^* \rangle_{A_i} + D_Q^* . \tag{2}$$

The bar in these expressions indicates the simple weighted mean of a series of measurements for one artifact ($Y_j^*[1], Y_j^*[2], \dots$), obtained at different stages:

$$\bar{Y}_j^* = \sum_{n=1}^N \eta_n Y_j^*[n],$$

where:

$$\eta_n = \frac{(u(y_j^*[n]))^{-2}}{\sum_{k=1}^N (u(y_j^*[k]))^{-2}}$$

and $u(y_j^*[n])$ is the standard uncertainty in the value of the n th result, $y_j^*[n]$, from participant j . There is only one artifact per participant in this scenario, so $\langle \bar{Y}_j^* - \bar{Y}_Q^* \rangle_{A_j}$ indicates the difference between the mean of participant j 's measurements and the mean of the pilot's measurements of the same artifact.

Equation (1) was implemented in the GTC software, as shown below. This code obtains an uncertain number representing D_Q^* :

```
d_Q = -sum(
    w[l_j] * ( mean(r_j.lab) - mean(r_j.pilot) )
    for l_j,r_j in kc_results.items()
```

The function `mean()` evaluates the mean of a sequence of uncertain numbers; `r_j.lab` and `r_j.pilot` contain, respectively, a sequence of results from participant j and the corresponding sequence of pilot measurements for the same artifact; `kc_results` is a container of objects as `r_j` for all participants; `w[l_j]` represents the weighting factors. Following Equation (2), a DoE is evaluated for each of the other comparison participants:

```
d_i = mean(r_i.lab) - mean(r_i.pilot) + d_Q
```

The results, with associated standard uncertainties, may be displayed as:

```
DoE[A] : 0.000249 (0.000278)
DoE[B] : 0.000305 (0.001225)
DoE[C] : 0.000008 (0.000097)
DoE[D] : 0.003274 (0.000886)
DoE[E] : 0.000487 (0.000480)
DoE[F] : 0.000012 (0.000266)
DoE[G] : -0.001776 (0.004078)
DoE[H] : -0.000219 (0.000138)
DoE[I] : -0.000011 (0.000110)
DoE[J] : 0.000192 (0.000285)
DoE[K] : -0.000097 (0.000750)
DoE[Q] : -0.000129 (0.000056)
```

2.2. DoE Uncertainty Budgets

The DoEs are influenced by factors in the participants' measurements, with each factor giving rise to one component of uncertainty. Because Equations (1) and (2) combine results from all participants, there is a large number (278) of components in the budget of each DoE in this scenario. Listing 3 shows the DoE for Participant A and an abridged uncertainty budget, in which the more significant components of uncertainty are shown—those with magnitudes greater than 10 % of the largest component. These factors can be identified as influences from A's own measurements and from those of the pilot on the same artifact.

Listing 3. The DoE for Participant A is shown at the top, with the combined standard uncertainty in parentheses. An abridged uncertainty budget follows. Only the components with a magnitude greater than trim times the largest component are shown. The components are listed in decreasing order of magnitude.

```
D_kc[A] = 0.000249(0.000278)

Uncertainty budget (trim=0.1):
  A:Beam Size & Position (sys): 0.00018602
  A(2):Beam Size & Position (ran): 0.00013449
      A(4):Type-A (ran): 0.00010139
      A(2):Type-A (ran): 0.00005375
  A(4):Beam Size & Position (ran): 0.00005153
      A(2):Instability (ran): 0.00004686
      A:Non-linearity (sys): 0.00002859
  Q(A):Non-Parallel Surfaces (sys): 0.00002217
  Q_A(1):Drift & Instability (ran): 0.00002101
  Q_A(3):Drift & Instability (ran): 0.00002101
  Q_A(5):Drift & Instability (ran): 0.00002101
```

There is quite a diversity of structure in the uncertainty budgets among the different participants. Listing 4 shows the DoEs obtained for Participants B and C. Participant B's result has a much larger combined standard uncertainty than Participant C, and the uncertainty budget is dominated by factors associated with B's own measurements. In contrast, the DoE for Participant C has the lowest uncertainty of all participants, and the corresponding uncertainty budget has many more significant influence factors. The largest of these are from C's own measurements and the corresponding pilot measurements. However, we also see components associated with measurements by Participants A, E, F, H, I, J, and K. These participants were weighted more heavily than B, D, and G during the DoE calculation (see Appendix A).

Listing 4. The DoEs for Participants B and C. See the caption to Listing 3 for further details.

```
-----
D_kc[B] = 0.000305(0.001225)

Uncertainty budget (trim=0.1):
  B:Inter-reflection (sys): 0.00109561
  B(2):Source Drift & Fluctuation (ran): 0.00036611
  B(4):Source Drift & Fluctuation (ran): 0.00031325
      B:Beam Size & Position (sys): 0.00018227
      B:Bandwidth (sys): 0.00015936
-----

D_kc[C] = 0.000008(0.000097)

Uncertainty budget (trim=0.1):
  C:Stray light (sys): 0.00004288
      C(2):SFK (ran): 0.00002973
  C:Non-linearity (sys): 0.00002751
      C(4):Type-A (ran): 0.00002727
  C:Inter-reflexion (sys): 0.00002185
  Q(C):Non-Parallel Surfaces (sys): 0.00002099
  Q_C(1):Drift & Instability (ran): 0.00001784
  Q_C(3):Drift & Instability (ran): 0.00001784
  Q_C(5):Drift & Instability (ran): 0.00001784
      H:Stray Light (sys): 0.00001754
      H:Inter-reflection (sys): 0.00001754
      J:Inter-reflection (sys): 0.00001523
      I(2):Type-A (ran): 0.00001441
```

```

I:Inter-reflection (sys): 0.00001426
J:Prismatic effect (sys): 0.00001362
  Q_C(1):Type-A (ran): 0.00001133
  Q_C(3):Type-A (ran): 0.00001133
  Q_C(5):Type-A (ran): 0.00001133
  F:Non-linearity (sys): 0.00001065
A:Beam Size & Position (sys): 0.00000914
  C(4):SFK (ran): 0.00000863
  E:Inter-reflection (sys): 0.00000767
  H:Polarization (sys): 0.00000760
E:Detector reproducibility (sys): 0.00000738
  C(2):Type-A (ran): 0.00000690
  F:Inter-reflection (sys): 0.00000666
  F:Polarization (sys): 0.00000666
  H(4):Filter instability (ran): 0.00000666
  K:Polarization (sys): 0.00000662
A(2):Beam Size & Position (ran): 0.00000661
  H(2):Filter instability (ran): 0.00000634
Q_I(1):Drift & Instability (ran): 0.00000631
Q_I(3):Drift & Instability (ran): 0.00000631
  F(2):Filter Stability (ran): 0.00000614
  K:Beam Size & Position (sys): 0.00000597
H(4):System reproducibility (ran): 0.00000586
  F(4):Filter Stability (ran): 0.00000585
  D(2):Beam Size & Position (ran): 0.00000568
H(2):System reproducibility (ran): 0.00000559
  A(4):Type-A (ran): 0.00000498
  F(4):Type-A (ran): 0.00000455
  D(2):Type-A (ran): 0.00000443
  B:Inter-reflection (sys): 0.00000439

```

The detail about individual influence factors shown in the listings above is more than the minimum required to analyze and link comparisons. Only the net systematic and random components are needed for that purpose. This is what is used at present, and the reduction in complexity makes the analysis tractable without digitalization. However, the physical origins of influence factors are obscured. For example, Listing 5 shows the budgets of Participants B and C in terms of systematic and random components. Compared to the information shown Listing 4, this offers little insight into the origins beyond participant and stage.

Listing 5. The DoEs for Participants B and C showing the total systematic and random effects as components of uncertainty. These budgets are equivalent to those in Listing 4; however, only the net random and systematic contributions at each stage are shown.

```
-----
D_kc[B] = 0.000305(0.001225)
```

```
Uncertainty budget (trim=0.1):
```

```

  B (sys): 0.00061183
  B(2) (ran): 0.00036717
  B(4) (ran): 0.00031562

```

```
-----
D_kc[C] = 0.000008(0.000097)
```

```
Uncertainty budget (trim=0.1):
```

```

  C(2) (ran): 0.00003052
  C(4) (ran): 0.00002860
  Q_C(1) (ran): 0.00002142
  Q_C(3) (ran): 0.00002142

```

```

Q_C(5) (ran): 0.00002142
  C (sys): 0.00001907
  I (sys): 0.00001461
I(2) (ran): 0.00001441
  H (sys): 0.00001344
  J (sys): 0.00001043
H(2) (ran): 0.00000914
H(4) (ran): 0.00000890
F(2) (ran): 0.00000816
F(4) (ran): 0.00000812
D(2) (ran): 0.00000790
A(2) (ran): 0.00000748
Q_I(1) (ran): 0.00000748
Q_I(3) (ran): 0.00000748
  Q_C (sys): 0.00000700
  F (sys): 0.00000696
A(4) (ran): 0.00000566
  E (sys): 0.00000540
Q_H(1) (ran): 0.00000499
Q_H(3) (ran): 0.00000499
Q_H(5) (ran): 0.00000499
  J(2) (ran): 0.00000483
  J(4) (ran): 0.00000460
  K (sys): 0.00000448
E(2) (ran): 0.00000384
E(4) (ran): 0.00000384

```

3. The RMO Key Comparison

Seven NMIs participated in the subsequent RMO key comparison (identified by the letters T, U, V ..., Z) and a pilot laboratory (P). The pilot and Participant Z had both taken part in the initial CIPM comparison, so their results were used to link the two comparisons. The participants each measured a different artifact, and the pilot measured all seven artifacts. The comparison was carried out in three stages: first, the pilot measured the artifacts; second, each participant reported a measurement; third, the pilot measured all the artifacts again.

3.1. Evaluating DoEs

The possibility of slight shifts in the scales of the linking participants since the initial CIPM comparison must be accounted for when linking. Therefore, linking participants provide information on the stability of their scales as part of their report during the RMO comparison. Formally, in the analysis, a quantity that includes a term $E_{D,i}$ representing scale movement is used for the DoE of each linking participant:

$$D'_i = D_i^* + E_{D,i} . \tag{3}$$

$E_{D,i}$ can be thought of as a residual error in the scale that contributes to uncertainty in the DoE. To provide a link to the RMO comparison, we then evaluate ([6], Equation (46)):

$$\begin{aligned}
 D_P &= - \sum_I v_I (\langle \bar{Y}_I - \bar{Y}_P \rangle_{A_I} - D'_I) \\
 &= -v_Z (\langle \bar{Y}_Z - \bar{Y}_P \rangle_{A_Z} - D'_Z) + v_P D'_P ,
 \end{aligned}
 \tag{4}$$

where v_i are the weight factors for linking participants (see Appendix B). Finally, the DoEs of non-linking participants are ([6], Equation (45)):

$$D_i = \langle \bar{Y}_i - \bar{Y}_P \rangle_{A_i} + D_P . \tag{5}$$

Our data processing uses a Python dictionary to hold the uncertain numbers for each DoE evaluated according to Equation (5):

```
D_rc = dict()
for l_i in lab_IDs[:-1]:
    r_i = rc_results[l_i]
    D_rc[l_i] = mean(r_i.lab) - mean(r_i.pilot) + D_P
```

A link to the initial comparison is obtained, following Equation (4), from:

$$D_P = -(\nu_Z * M_Z + \nu_P * M_P)$$

where ν_Z and ν_P correspond to v_Z and v_P , respectively, and correspond to Equations (A3) and (A4),

$$M_Z = \text{mean}(rc_results['Z'].lab) - \text{mean}(rc_results['Z'].pilot) - rc_link_doe['Z']$$

$$M_P = -rc_link_doe['P']$$

where $rc_results['Z'].lab$ is the sequence of measurements submitted by Participant Z and $rc_results['Z'].pilot$ are the corresponding pilot measurements. Following Equation (3), the uncertain numbers $rc_doe['Z']$ and $rc_doe['P']$ were calculated by adding an uncertain number for the participant's scale stability to the participant's DoE obtained in the CIPM comparison (see the dataset for further details [14]).

The resulting DoEs, with standard uncertainties in parentheses, are:

```
DoE[T] = 0.00136 (0.00203)
DoE[U] = -0.00137 (0.00076)
DoE[V] = 0.00182 (0.00094)
DoE[W] = -0.00138 (0.00095)
DoE[X] = 0.00032 (0.00042)
DoE[Y] = 0.00297 (0.00314)
```

3.2. DoE Uncertainty Budgets

In the linked RMO comparison, the DoEs are each influenced by 302 factors (these factors were identified by participants when submitting their results and, as explained above, the influences from all participants to contribute to the uncertainty). Again, there is diversity in the uncertainty budgets of different participants. For example, the uncertainty budget in Listing 6 shows that the most important components of uncertainty for Participant Y, the participant with the largest DoE uncertainty, are all related to Y's own measurement.

Listing 6. The DoE for participant Y with an abridged uncertainty budget.

```
D[Y] = 0.002972(0.003139)

Uncertainty budget (trim=0.1):
    Y(2):Scale bias (ran): 0.002910
    Y:Beam Size & Position (sys): 0.000691
    Y:Non-linearity (sys): 0.000600
    Y(2):Source Drift & Fluctuation (ran): 0.000536
    Y(2):Type-A (ran): 0.000346
```

In contrast, Listing 7 shows the budget for X, the participant with the least DoE uncertainty, which is influenced most by measurements performed by others: the pilot's measurements of the artifacts used by X and the other linking Participant Z. This budget also includes important components from some factors in the initial comparison.

Listing 7. The DoE for Participant X with an abridged uncertainty budget. Note that Participant Z in the RMO comparison was I in the initial comparison. The component of uncertainty labeled Z-I:Scale Instability accounts for the stability of Z's measurement scale.

```
D[X] = 0.000319(0.000420)

Uncertainty budget (trim=0.1):
    P_X(3):Inter-reflection (ran): 0.000155
    P_X(1):Inter-reflection (ran): 0.000153
    P_Z(3):Inter-reflection (ran): 0.000144
    P_Z(1):Inter-reflection (ran): 0.000142
        X(2):Type-A (ran): 0.000132
        X:Stray Light (sys): 0.000130
        Z(2):Type-A (ran): 0.000127
    Z-I:Scale Instability (sys): 0.000106
    X:Beam Size & Position (sys): 0.000076
    P-E:Scale Instability (sys): 0.000074
    X(2):Source Drift & Fluctuation (ran): 0.000058
        I(2):Type-A (ran): 0.000056
        P_X(1):Type-A (ran): 0.000033
        P_Z(1):Type-A (ran): 0.000030
    X:Non-linearity (sys): 0.000025
    Q_I(1):Drift & Instability (ran): 0.000024
    Q_I(3):Drift & Instability (ran): 0.000024
        X:Obliquity (sys): 0.000023
    X(2):Obliquity (ran): 0.000023
    H:Stray Light (sys): 0.000018
    H:Inter-reflection (sys): 0.000018
        Q_I(1):Type-A (ran): 0.000016
        Q_I(3):Type-A (ran): 0.000016
```

4. Discussion

This study looked at the use of uncertain numbers as a digital format for reporting measurement data. The context of the study is a specialized area, but the underlying concern is a more general problem: the presence of fixed (systematic) influence factors at different stages of a traceability chain give rise to correlations in data that affect the uncertainty at the end of a chain, but are difficult to account for. Uncertain numbers address this issue and support a simple and intuitive form of data processing. The method is fully compliant with the recommendations in the GUM [7].

Comparison analysis can be a rather laborious and error-prone task at present, because there is a large amount of data to be manipulated. The task could be greatly simplified if something such as the uncertain-number format were adopted. Digital records could then include information about any common factors that lead to correlations. Algorithms would use that information to streamline the data processing and produce more informative results. This is an interesting possibility in the context of the CIPM MRA, because comparison results are used to support NMI claims of competency (CMCs), which are a matter of considerable importance. Greater transparency in the composition of uncertainty budgets for degrees of equivalence would surely be welcome. For example, the largest components of uncertainty in Listing 7 are associated with influence factors for the pilot measurements, not those of Participant X. This shows that the weight of evidence provided by a DoE and its uncertainty to support a CMC claim may be limited by the performance of the pilot and/or linking participants in the comparisons.

Situations where common factors may give rise to correlations in measurement data are not infrequent, but conventional calibration certificate formats do not allow an accurate evaluation of uncertainty in such cases (as is possible in comparison analysis and comparison linking) [9,10]. Our study therefore draws attention to some informal decision, made decades ago, not to report information about influence factors—the uncertainty

budget. This decision was surely made for pragmatic reasons, because additional effort would be required to curate uncertainty budget data in paper-based systems. However, the policy should be reviewed now as part of the digital transformation process. The currently favored DCC formats only report uncertainty intervals or expanded uncertainties [17]. If these formats are ultimately adopted, DCCs will not contain enough information about common influences upstream to handle correlations in downstream data: the scenario envisaged in this article would not be realized.

4.1. Metrological Traceability

Metrological traceability is realized by forming a chain of calibrations that link primary realizations of the SI units to an end user of measurement data. At each stage, influence factors cause the result obtained to differ slightly from the actual quantity of interest. At present, this is usually accounted for by a single uncertainty statement in the stage report. However, influence factors at one stage can give rise to correlations at later stages, and accounting for these effects requires influences to be tracked along the chain [9,10]. That is why the notion of transferability, referred to in the Introduction and the GUM, is important, and comparison analysis highlights this by requiring substantial additional information from all participants, which complements the information in calibration certificates. The study showed that transferability, and hence better support for traceability, is provided by the uncertain number format. In more general terms, the approach to digitalization can keep track of influences and hence identify the *provenance* of contributions to uncertainty in a result.

4.2. Measurement Comparisons in Other Fields

This case study considered CCPR comparisons, but in other areas, comparisons may have different general characteristics. For instance, degrees of freedom are usually high in CCPR comparisons and can be ignored; however, in some other areas, they need to be taken into consideration. For instance, in a CIPM key comparison of polychlorinated biphenyl (PCB) congeners in sediment [18], some participants reported very low degrees of freedom (as low as two). The uncertain number format used in GTC also handles degrees of freedom.

The CCPR requires participants in CIPM comparisons to realize their scales independently. Therefore, the results of participants in the initial comparison were not correlated. However, in other areas, the assumption of independence at the CIPM level may not hold. In comparisons involving mass, for example, the lack of independence among participants' scales has to be considered [19]. In the future, if digital reporting were to adopt something equivalent to the uncertain number format, the necessary information about shared influences would be accessible in DCCs. This would once again simplify data analysis and deliver more informative results.

4.3. Unique Identifiers and Digital Records

Details about the digital storage format used by GTC are outlined in [13]. The role of unique identifiers associated with influence quantities is of interest.

Uncertain-number algorithms must keep track of the identity of all the influence factors, which is analogous to the need for adequate notation in mathematical expressions. In the GUM, a general measurement function is represented as:

$$Y = f(X_1, X_2, \dots), \quad (6)$$

where Y is the quantity intended to be measured and X_1, X_2, \dots are the quantities that influence the measurement. In the GUM notation for a component of uncertainty, $u_i(y)$ is understood to be the component of uncertainty in y , as an estimate of Y , due to the uncertainty in an estimate, x_i , of the influence quantity X_i . Therefore, the subscript i may take the value of any of the X 's subscripts in Equation (6). Uncertain-number software and digital records must somehow keep track of all the i 's as well. This is more complicated

than it first appears, because measurements are carried out in stages that occur in different locations and at different times.

GTC uses a standard algorithm to produce universally unique 128 bit integers, which it uses to form unique digital identifiers [13]. The format is simple, and the identifier reveals nothing about influence. Would a more sophisticated type of identifier be appropriate [20]? For the purposes of data processing alone, there is no need to complicate matters: the only requirement is uniqueness. However, GTC does allow text labels to be associated with identifiers (used as influence quantity labels in uncertainty budgets), and a planned enhancement to GTC will allow unique identifiers to index a manifest of information about each influence quantity. A manifest could accompany the digital record of uncertain-number data, which would address needs for metadata about influence quantities without the burden of minting and configuring digital objects that give access to information on the Internet.

4.4. Comparison Analysis by Generalized Least Squares

The analysis equations used in this work take a fairly straightforward mathematical form. However, we have alluded to the complexity, due to the large number of terms, in handling the associated uncertainty calculations, and in [6], we suggested that a more practical analysis tool is generalized least squares (GLS). GLS is a more opaque “black box” method, but software packages are available to perform the linear algebra once the required matrices have been prepared (see [6], §4).

It is interesting to note that, in order to link an RMO comparison, GLS has been used to simultaneously process CIPM comparison data and RMO comparison data [19]. This is another example of comparison analysis compensating for the lack of transferability in standard reporting formats.

A GLS algorithm could also be applied to uncertain-number data, in which case results such as those described here would be obtained. In the formulation of the GLS calculation (Equation (69) of [6], §4, repeated here; note that bold Roman type is used to represent matrices):

$$\begin{bmatrix} \mathbf{m} \\ \mathbf{d} \end{bmatrix} = (\mathbf{x}^T \mathbf{u}^{-1} \mathbf{x})^{-1} \mathbf{x}^T \mathbf{u}^{-1} \begin{bmatrix} \mathbf{y} \\ \mathbf{d}_I \end{bmatrix}, \tag{7}$$

the elements of the design matrix, \mathbf{x} , are pure numbers, as are the elements of the covariance matrix, \mathbf{u} , so conventional numerical routines can be used to evaluate the matrix:

$$\mathbf{g} = (\mathbf{x}^T \mathbf{u}^{-1} \mathbf{x})^{-1} \mathbf{x}^T \mathbf{u}^{-1}. \tag{8}$$

Then, with uncertain-number elements in the vector of participant results, \mathbf{y} , and in the vector of linking participant DoEs, \mathbf{d}_I , the final calculation of degrees of equivalence:

$$\begin{bmatrix} \mathbf{m} \\ \mathbf{d} \end{bmatrix} = \mathbf{g} \begin{bmatrix} \mathbf{y} \\ \mathbf{d}_I \end{bmatrix} \tag{9}$$

obtains the vector of linking participant DoEs, \mathbf{d} , as a linear combination of uncertain numbers. The results would, as in this work, reflect the influence of all terms contributing to participants’ measurements. This would be much more informative than the information available from the covariance matrix usually obtained as an additional calculation in GLS analysis.

5. Conclusions

This case study of comparison analysis and linking has identified benefits in a particular approach to digitalization using a digital format called an uncertain number. Because comparison participants must provide more information than is available in standard calibration certificates, the context of the study highlights deficiencies in current reporting formats. These deficiencies can be summarized as a lack of support for transferability and internal consistency in the expression of uncertainty. However, if the uncertain-number

format were widely adopted, as was assumed in this case study, transferability and internal consistency would be achieved.

The study shows that more rigorous uncertainty calculations are enabled by uncertain numbers. Algorithms for data processing can be expressed in a more intuitive and streamlined manner, and it is no longer necessary to formulate separate calculations for measurement uncertainty. Because the approach keeps track of all influences, it can deliver more accurate uncertainty statements. Uncertain numbers would be advantageous to a wider range of measurement problems than just international comparisons. Adopting the format for DCCs could therefore enhance the quality of new digital infrastructures.

Author Contributions: Conceptualization, B.D.H. and A.K.; methodology, B.D.H.; software, B.D.H.; formal analysis, A.K.; data curation, A.K.; writing—original draft preparation, B.D.H.; writing—review and editing, A.K.; All authors have read and agreed to the published version of the manuscript.

Funding: This work was funded by the New Zealand government.

Data Availability Statement: A dataset for the case study in this article is available [14]. It consists of digital records (JSON files) of the participant results, as well as JSON files containing the uncertain-number DoE results obtained from the data processing. Python modules that display the contents of these files are provided, as are modules to carry out the data processing. The Python software package GUM Tree Calculator (GTC) is required (Version 1.3.6 or above) [11].

Acknowledgments: The authors are grateful to Peter Saunders for carefully reviewing this work. B.D.H. is a member of the team supporting the CIPM-TG-DSI, and A.K. is a member of the CCPR. The opinions expressed by the authors are not necessarily shared by other members of those groups.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

BIPM	International Bureau of Weights and Measures
CCPR	Consultative Committee for Photometry and Radiometry
CIPM	International Committee for Weights and Measures
CIPM-TG-DSI	CIPM Task Group on the Digital SI
CMC	Calibration and measurement capability
DoE	Degree of equivalence
LPU	Law of the propagation of uncertainty
MRA	CIPM Mutual Recognition Arrangement
NMI	National Metrology Institute
RMO	Regional metrology organization

Appendix A. Weights for Evaluating DoEs in a CIPM Key Comparison

The CCPR uses weights in the determination of DoEs that depend on the uncertainties reported by the participants ([6], Appendix A). These weights are calculated from the arithmetic mean of the combined standard uncertainties reported by each participant. For participant i , with a mean combined standard uncertainty $\bar{u}(y_i^*)$,

$$w_i = \frac{(\bar{u}(y_i^*))^{-2}}{\sum_j (\bar{u}(y_j^*))^{-2}}, \quad (\text{A1})$$

where the sum is over all participants and $\bar{u}(y_j^*)$ is the arithmetic mean of the combined standard uncertainties reported by participant j .

CCPR guidelines recommend a cut-off to limit the influence of any measured values with very low uncertainty. The minimum uncertainty value is taken to be the mean of

those $\bar{u}(y_i^*)$ values that are less than or equal to the median of the uncertainties reported by all participants:

$$u_{\text{cut}} = \text{mean}(\bar{u}(y_i^*)) \quad \text{for all } \bar{u}(y_i^*) \leq \text{median}(\{u(y_i^*)\}). \quad (\text{A2})$$

The weighting factor is evaluated by Equation (A1) while taking the greater of $\bar{u}(y_i^*)$ and u_{cut} as the mean uncertainty for participant i . In this work, $u_{\text{cut}} = 1.92 \times 10^{-4}$ and the weights obtained were:

$$\begin{aligned} w_A &= 0.047, & w_E &= 0.025, & w_I &= 0.191, \\ w_B &= 0.004, & w_F &= 0.067, & w_J &= 0.073, \\ w_C &= 0.191, & w_G &= 0.000, & w_K &= 0.012, \\ w_D &= 0.009, & w_H &= 0.191, & w_Q &= 0.191. \end{aligned}$$

A cut off was imposed on participants C, H, I and Q. The results from those laboratories were weighted most heavily, followed by those from participants J and F.

Appendix B. Weights for Linking Participants in an RMO Key Comparison

The contributions from linking participants in an RMO comparison are weighted to take account of correlations. With two linking participants, we follow ([6], Appendix B) and evaluate the required weights as a function of the uncertainties of

$$M_Z = (\bar{Y}_Z - \bar{Y}_P)_{A_Z} - D_Z, \quad (\text{A3})$$

$$M_P = -D_P, \quad (\text{A4})$$

where P is the pilot (E in the initial CIPM comparison) and Z (I in the initial comparison) is the other linking participant. Writing the standard uncertainties in these quantities as $u(m_Z)$ and $u(m_P)$, and their covariance as $u(m_Z, m_P)$, the weights are ([6], Equation (76))

$$v_i = \frac{\frac{1}{u^2(m_i)} - \frac{u(m_Z, m_P)}{u^2(m_Z)u^2(m_P)}}{\frac{1}{u^2(m_Z)} + \frac{1}{u^2(m_P)} - 2\frac{u(m_Z, m_P)}{u^2(m_Z)u^2(m_P)}}, \quad (\text{A5})$$

for $i = Z$ and P . In this work, $v_Z = 0.931$ and $v_P = 0.069$ were obtained.

References

- Gadelrab, M.S.; Abouhoggail, R.A. Towards a new generation of digital calibration certificate: Analysis and survey. *Measurement* **2021**, *181*, 109611. [CrossRef]
- Quinn, T. The Metre Convention and world-wide comparability of measurement results. *Accred. Qual. Assur.* **2004**, *9*, 533–538. [CrossRef]
- CIPM. CIPM Task Group on the Digital SI. Available online: <https://www.bipm.org/en/committees/ci/cipm/wg/cipm-tg-dsi> (accessed on 29 August 2021).
- CIPM. Technical Supplement revised in October 2003. In *Mutual Recognition of National Measurement Standards and of Calibration and Measurement Certificates Issued by National Metrology Institutes*; BIPM: Sèvres, France, 1999; pp. 38–41.
- CIPM. *Measurement Comparisons in the CIPM MRA*; Technical Report; Version 1.0; Bureau International des Poids et Mesures: Sèvres, France, 2021.
- 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. [CrossRef]
- 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, France, 2008.
- Hall, B.D. Computing with Uncertain Numbers. *Metrologia* **2006**, *43*, L56–L61. [CrossRef]
- Hall, B.D.; White, D.R. Digital representation of measurement uncertainty for metrological traceability. In *Advanced Mathematical and Computational Tools for Metrology XII*; World Scientific Publishing: Singapore, 2021; Volume 90, pp. 262–272.
- Hall, B.D.; White, D.R. Digital representation of measurement uncertainty. *Meas. Sens.* **2021**, *18*, 100074. [CrossRef]
- Hall, B.D.; Borbely, J.S. GUM Tree Calculator. 2021. Available online: <https://github.com/MSLNZ/GTC> (accessed on 29 August 2021).

12. Hall, B.D.; Borbely, J.S. GTC Documentation. 2021. Available online: <https://gtc.readthedocs.io/en/stable/> (accessed on 29 August 2021).
13. Hall, B.D. The GUM Tree Calculator: A Python package for measurement modelling and data processing with automatic evaluation of uncertainty. *Metrology* **2021**, submitted.
14. Hall, B.D. [Dataset] Digital Representation of Measurement Uncertainty: A Case Study Linking an RMO Key Comparison with a CIPM Key Comparison. [[CrossRef](#)]
15. Koo, A. Report on the consultative committee for photometry and radiometry key comparison of regular spectral transmittance 2010 (CCPR-K6.2010). *Metrologia* **2017**, *54*, 02001. [[CrossRef](#)]
16. Cooksey, C.C.; Toman, B. Report on the SIM photometry and radiometry Key comparison of spectral regular transmittance. *Metrologia* **2021**, *58*, 02004. [[CrossRef](#)]
17. Hackel, S.; Härtig, F.; Hornig, J.; Wiedenhöfer, T. The digital calibration certificate. *Metrol. Digit. Wirtsch. Ges.* **2017**, *127*, 75–81.
18. Schantz, M.; Wise, S. CCQM-K25: Determination of PCB congeners in sediment. *Metrologia* **2004**, *41*, 08001. Available online: http://www.bipm.org/utls/common/pdf/final_reports/QM/K25/CCQM-K25.pdf (accessed on 29 August 2021).
19. Sutton, C.M. Analysis and linking of international measurement comparisons. *Metrologia* **2004**, *41*, 272–277. [[CrossRef](#)]
20. Klump, J.; Huber, R. 20 Years of Persistent Identifiers—Which Systems are Here to Stay? *Data Sci. J.* **2017**, *16*, 9. [[CrossRef](#)]

Article

Systematic Distortion Factor and Unrecognized Source of Uncertainties in Nuclear Data Measurements and Evaluations

Nikolay V. Kornilov ^{1,*†}, Vladimir G. Pronyaev ^{2,†} and Steven M. Grimes ^{1,†}¹ Department of Physics and Astronomy, Ohio University, Athens, OH 45701, USA; grimes@ohio.edu² Private Institution Atomstandart, 45701 Moscow, Russia; vgpronyaev@yandex.ru

* Correspondence: kornilov@ohio.edu

† retired.

Abstract: Each experiment provides new information about the value of some physical quantity. However, not only measured values but also the uncertainties assigned to them are an important part of the results. The metrological guides provide recommendations for the presentation of the uncertainties of the measurement results: statistics and systematic components of the uncertainties should be explained, estimated, and presented separately as the results of the measurements. The experimental set-ups, the models of experiments for the derivation of physical values from primary measured quantities, are the product of human activity, making it a rather subjective field. The Systematic Distortion Factor (SDF) may exist in any experiment. It leads to the bias of the measured value from an unknown “true” value. The SDF appears as a real physical effect if it is not removed with additional measurements or analysis. For a set of measured data with the best evaluated true value, their differences beyond their uncertainties can be explained by the presence of Unrecognized Source of Uncertainties (USU) in these data. We can link the presence of USU in the data with the presence of SDF in the results of measurements. The paper demonstrates the existence of SDF in Prompt Fission Neutron Spectra (PFNS) measurements, measurements of fission cross sections, and measurements of Maxwellian spectrum averaged neutron capture cross sections for astrophysical applications. The paper discusses introducing and accounting for the USU in the data evaluation in cases when SDF cannot be eliminated. As an example, the model case of $^{238}\text{U}(n,f)/^{235}\text{U}(n,f)$ cross section ratio evaluation is demonstrated.

Keywords: metrology; nuclear data; data evaluation; systematic distortion factor; unrecognized source of uncertainties

Citation: Kornilov, N.V.; Pronyaev, V.G.; Grimes, S.M. Systematic Distortion Factor and Unrecognized Source of Uncertainties in Nuclear Data Measurements and Evaluations. *Metrology* **2022**, *2*, 1–18. <https://doi.org/10.3390/metrology2010001>

Academic Editor: Simona Salicone

Received: 31 August 2021

Accepted: 8 December 2021

Published: 24 December 2021

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

We examine the nature of the experimental investigations in this paper. Therefore, we should start with some definitions. Every experiment begins with conceptualization. The model of the experiment should be prepared. After this stage, we implement this model in the experimental set-up and in the relations for obtaining physical quantities from the primarily measured data. The important step is an evaluation of uncertainties of physical quantities applying the error propagation law to the uncertainties of primary measured quantities.

Each experiment can contain a Systematic Distortion Factor (SDF). The SDF [1] changes experimental results and should be treated as a real physical effect. Therefore, it is very important to analyze and to compare the results of all experiments. If we can understand the nature of SDF, one can calculate the corrections, and corrected data can be used in the evaluation. If the data set contains substantial SDF leading to extremely outlying data and they cannot be corrected, these data should be removed from the evaluation procedure. In some cases, the SDF can be estimated from the comparison of experimental results with the results of a Monte Carlo (MC) simulation of the experiment. MC simulation allows us to calculate or justify the introduced corrections.

However, the best approach for SDF identification is the comparison of experimental results obtained in different measurements. Some examples of the SDF existence in Prompt Fission Neutron Spectrum (PFNS) measurements, measurements of fission cross sections used as standards and in Maxwellian Average Cross Section (MACS) measurements for astrophysical applications are demonstrated in this paper.

The present approach to the nuclear data evaluation for neutron cross section standards, which a priori can contain the SDF, are the following [2]:

- The use of model-independent fits and fits with highly credible physical models (e.g., R-matrix model),
- Combined Generalized Least Square (GLSQ) evaluation of many multivariate experimental data sets for different reactions obtained by different methods of measurements, with different energy resolution and at different energies,
- The use of experimental and evaluated covariance matrices with one standard deviation for uncertainty presentation,
- The use of systematic uncertainties for fully or partially correlated components of total uncertainty according to the model of the experiment for measured observables (sample mass, detector efficiency, room return correction, etc.),
- The determination of outlying data and work with their uncertainties.

To obtain realistically evaluated uncertainties consistent with a spread of experimental data relative to the evaluated values, the component of the Unrecognized Source of Uncertainty (USU) could be added to the covariance matrix of evaluated data. We examine the nature of the experimental investigations in this paper also. The following notations are used in this text: E_0 is the neutron incident energy, and E is the energy of outgoing neutrons.

2. SDF in the $^{235}\text{U}(n,f)$ PFNS Measurements

2.1. Time Resolution and Bin Correction

The PFNS is usually measured with Time of Flight (TOF) in Direct Beam (DB) experiments or in Pulsed Beam (PB) experiments. The DB means that these experiments were realized with a direct neutron beam, and a “stop” signal was produced from fission fragments. The PB experiments used a pulsed neutron beam for TOF.

The primary experimental data (counts registered by neutron detector) are collected at the time interval between the “start” event of a neutron detector and the “stop” event of a Fission Fragment (FF) detector. The PFNS is obtained after the transformation of a number of events from the time scale to the energy scale.

Several parameters and the accuracy of their determination are important for this type of experiment. We should examine additional measurements of the flight path, the time channel width, the neutron detector efficiency, possible shift of the timing position as a function of the neutron energy, the efficiency of the FF detector. In the case of the DB or the spontaneous fission measurements, the measured spectra should be corrected for the random coincidence of FF and neutron registration events.

We compare two experiments for PFNS measurements at ^{235}U fission by thermal neutrons performed about 25 years apart. The most important parameters of the old DB experiment (1983) [3,4] and the new experiment (2008) [5,6] are presented in Table 1. The $^{252}\text{Cf}(sf)$ PFNS was used as the standard for the determination of neutron detector efficiency in both cases.

The better time resolution for ^{235}U in the experiment [5,6] is explained by a larger distance between the cathode and anode in the uranium section of the chamber (fission fragment detector) and, as a result, a higher amplitude of the FF pulse.

The uncertainties for channel width and flight path are similar, about 0.1%. In the 1983 experiment, the time resolution was better by a factor of about 2. Correction at finite time resolution and channel width (bin) was achieved using the Maxwellian shape of the spectrum and parameters from Table 1 (1983 data). The influence of these corrections at the PFNS was simulated with Monte Carlo calculations.

Table 1. Parameters of the different experiments. Full Width at Half Maximum (FWHM) is the time resolution.

Parameter	1983	2008
Time bin (channel width), ns	1.414	0.1178
FWHM, ns	4.2	1.7 (^{235}U), 2.1 (^{252}Cf (sf))
Flight path (L), m	6.11	3.00
Time resolution, FWHM/L, ns/m	0.69	0.57 (^{235}U), 0.7 (^{252}Cf (sf))

The PFNS for $^{252}\text{Cf}(S_{\text{Cf}}(E))$ and $^{235}\text{U}(S_{\text{U}}(E))$ were measured together in the same experiment. The influence of these corrections at the measured ratio of PFNS $R(E) = S_{\text{Cf}}(E)/S_{\text{U}}(E)$ is very small, and we can conclude that we do not have any SDF connected with this part of data reduction if we measure relative to the ^{252}Cf standard.

The analyses in [7] confirm this conclusion. Average energies of ^{235}U PFNS estimated for these two experiments are very close $\langle E \rangle = 1.976 \pm 0.002$ MeV [3,4], $\langle E \rangle = 1.982 \pm 0.004$ MeV [5,6]. The comparison of spectra shown in Figure 1 is given as ratios to the spectrum predicted by the Scale Method [7]. The small difference of 0.006 MeV may be related to the SDF in the neutron angular-energy distribution relative to FF, which will be discussed in the following sections.

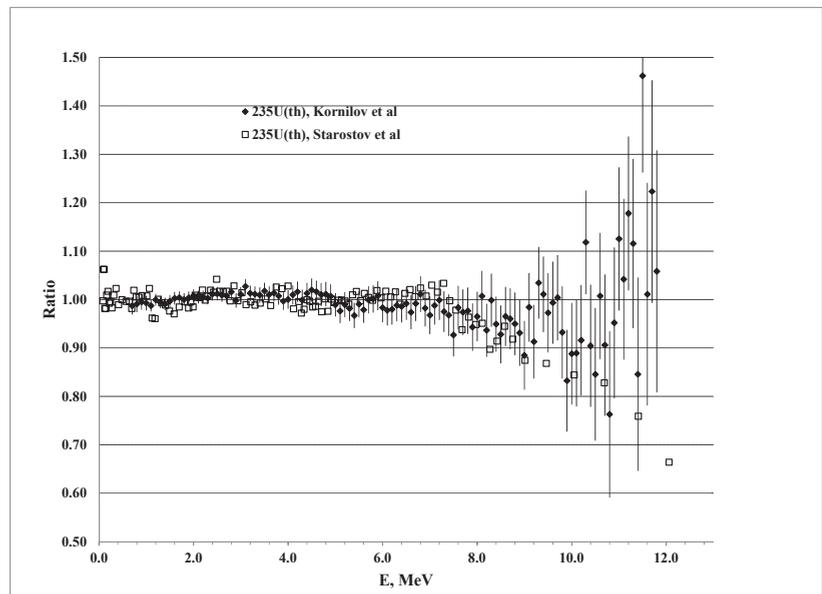


Figure 1. Ratio of the ^{235}U PFNS for thermal neutrons [3–5] to the fitted SM function [7].

2.2. Neutron Scattering in Fission Chamber

Three neutron detectors were applied in the experiment [5] (Figure 2). The ionization chamber used for FF counting is shown in Figure 3. Both fissile materials ^{252}Cf and ^{235}U were placed in the same chamber.

The MCNP code was used for the simulation of neutron scattering in the environment. Preliminary data analysis of the experimental results had shown that the experimental PFNS has very strong angular-energy dependence. At the same time, the MCNP simulation calculates an isotropic angular distribution. More careful analysis of the model used in the

calculations had shown that the simplification was achieved in the geometry; namely, ^{235}U and ^{252}Cf samples were placed in the same position in the center of the fission chamber.

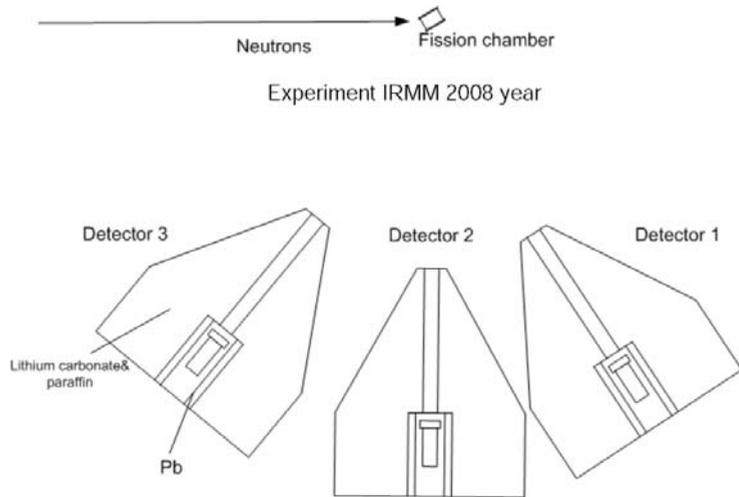


Figure 2. Experimental set-up for the IRMM-2008 experiment [5].

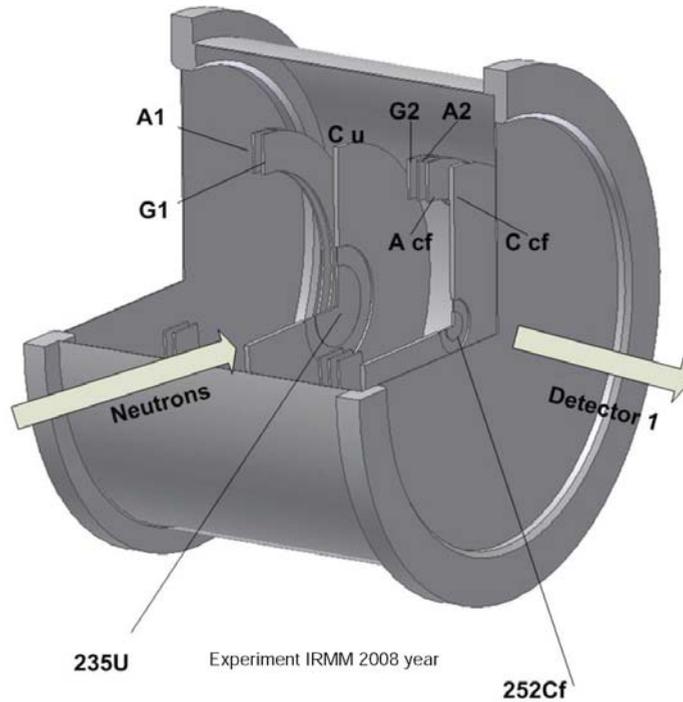


Figure 3. Drawing of the fission ionization chamber. C_u , C_{cf} —cathodes with U and Cf neutron sources, A_1 , A_2 , G_1 , G_2 —cathodes and grids for U part of the ionization chamber, A_{cf} —anode for Cf part [5].

After the correction of the MCNP input data, the angular effect appeared in the MCNP results (Figure 4) and disappeared in the corrected PFNS measured by detectors placed under different angles. The experimental spectra for different angles are in good agreement (inside experimental uncertainties (Figure 5)). This case demonstrates that the use of a simplified model of the experimental set-up (construction) may produce SDF. However, the accurate model of the experiment may remove this SDF.

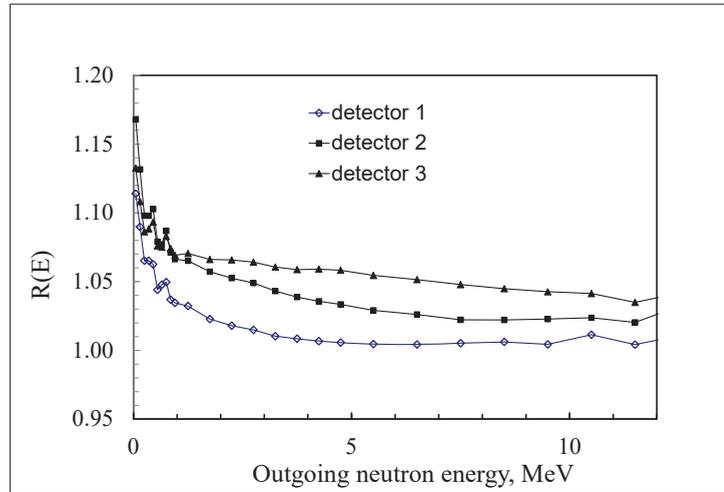


Figure 4. Ratio of multiple scattering corrections for Cf- and U-sources calculated in the MCNP simulations.

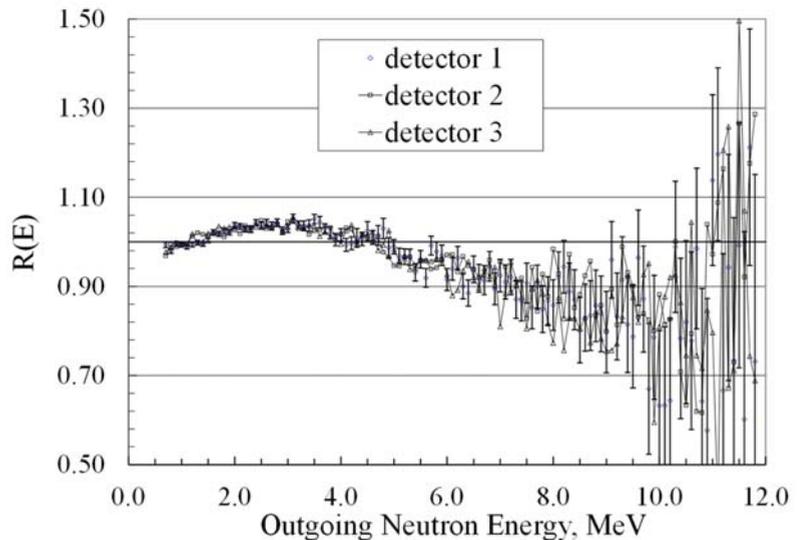


Figure 5. Results of ^{235}U PFNS measurements for three detectors after introducing the multiple scattering correction. They are shown as the ratios of the Maxwellian spectrum with the average energy $\langle E \rangle = 1.988$ MeV.

2.3. The Energy-Angular Distribution of Neutrons Relative to FF Flight Direction

Fission neutrons have a very strong energy-angular correlation in the Laboratory System (LS) of coordinate relative to the FF flight direction. This is a well-known experimental fact, which has a simple theoretical justification: if most neutrons are emitted after full acceleration of the FF by the Coulomb field, they should have high translational velocity of FF in the LS. However, evaluated data libraries and all practical applications are based on the trivial assumption, “during the fission, neutrons with the same spectrum as the angular integrated spectrum are emitted”. This is an incorrect simplification, which may stimulate strong SDF during the construction of the experimental set-up. This problem was discussed in several papers [8–10]. We would like to remind the reader of some results from [8].

The novel measurements of PFNS with registration of the angle between the flight direction of the fission fragment and neutron were performed at the PINP [11]. The registration of FF was performed in flat 2π geometry, with the azimuth angle fixed by “the belt” of FF counters and registration of neutrons by two neutron detectors located at the same plane.

This flat geometry was included in MC simulations. It was assumed that fission neutrons are emitted from fixed FF (single fragment) with Center of Mass (CM) energy E_V . The Maxwellian neutron spectrum for CM was used for calculation. Therefore, if we apply full integration (total 4π angle range), we should obtain the Watt distribution in the LS. The ratio of simulated spectra for flat geometry to the Watt distribution is shown in Figure 6. We see that the SDF for this geometry of measurements can be very strong. It increases the high energy part of the spectrum and average energy. In this simulation, we did not use the FF yields with their mass and kinetic energy distributions; therefore, this result cannot be used as the correction at the SDF. However, it demonstrates clearly that the limitation in the geometry of the measurements of neutron angular distributions relative to the FF flight direction may lead to the appearance of the SDF.

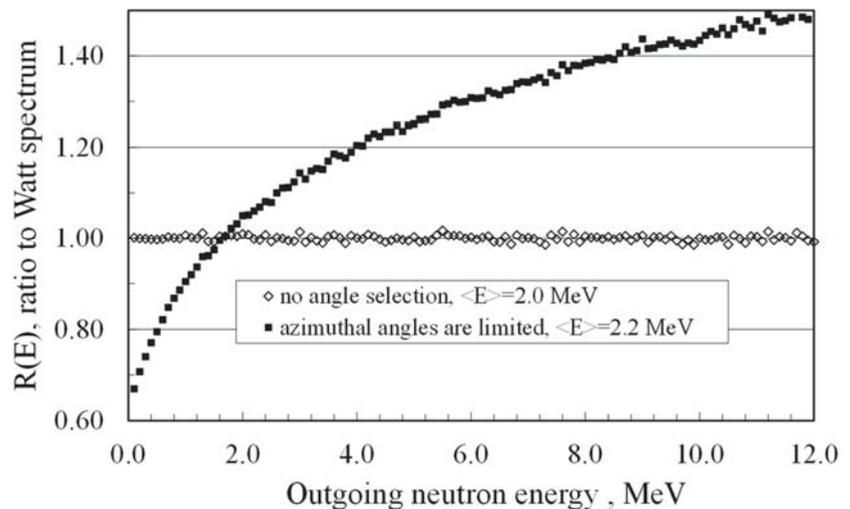


Figure 6. Distortion effect in the derived PFNS due to incomplete (angle selection) azimuth integration in [8].

2.4. SDF and Evaluation Procedure

The SDF can be created by the model used in the evaluation. The only way to find the SDF caused by the evaluation procedure is a comparison of the evaluations based on the same experimental data. The understanding of the influence of the selected data set is also important. The different sets of experimental data for $^{235}\text{U}(n_{\text{th}},f)$ PFNS in fission

induced by thermal neutrons were evaluated using the GMA (Gauss–Markov–Aitken approach) code. Because all measurements were undertaken relative to the $^{252}\text{Cf}(\text{sf})$ PFNS standard, the shape of the ratio $^{252}\text{Cf}(\text{sf})$ to $^{235}\text{U}(\text{n}_{\text{th}},\text{f})$ and Mannhart’s standard for absolute $^{252}\text{Cf}(\text{sf})$ spontaneous PFNS with covariances was used in the evaluation. Using the shape-of-ratio experimental data in the fit excludes large components of uncertainties contributing to absolute measurements. The normalization constraint was applied in the fit with an uncertainty close to the uncertainty of the evaluated average number of prompt neutrons for ^{235}U . The evaluated PFNS properly extrapolated to 0 and to 20 MeV on the energy of emitted neutrons was used for calculation of average neutron energy $\langle E \rangle$ of the spectrum. The data used in GMA evaluation were also used for $\langle E \rangle$ calculation with the scale method [7] fitting procedure. The 2% uncertainty at each point from GMA was applied in the SM analysis. It was applied for scaling of the PFNS shape. The results are presented in Table 2.

Table 2. The comparison of the $\langle E \rangle$ calculated with numerical integration of GMA evaluation in the energy range 0–20 MeV, and SM evaluation for the energy range 0.43–7.3 MeV.

Data Set	$\langle E \rangle$, MeV (GMA)	$\langle E \rangle$, MeV (SM)	χ^2/N (SM)
All data including [11]	1.997	1.992 ± 0.004	0.7
Without [11]	1.984	1.982 ± 0.004	0.4
No [11] but with [12]	1.981	1.978 ± 0.004	0.6

We can reach the same conclusion as before. The inclusion of the experimental data as they are given in [11] increases the $\langle E \rangle$ outside the estimated uncertainties. Therefore, there is no evidence that the procedure of the evaluation itself contains the SDF.

2.5. The Evidence of the SDF in LANL Double TOF Experiment

The final experimental data obtained at LANL are still not available. The values for analysis in this paper were obtained by digitalizing the data from figures presented in [13,14]. The uncertainty of this digitalization is small. The data were analyzed with the SM. The experimental PFNS as a ratio to the SM functions is shown in Figures 7 and 8. Data were analyzed in the energy range of the emitted neutrons $0.25 < E < 7$ MeV. The parameters of the SM fit were used for PFNS calculation in the energy range of emitted neutrons above 0.25 MeV. The data below 0.25 MeV were excluded from the analysis due to possible problems with neutron registration by ^6Li -glass detectors near 0.244 MeV ^6Li resonance. One should keep in mind that only three points are available in the range 7–10 MeV for the LANL experiment. The average energy for thermal point and LANL data are different. After the correction due to different $\langle E \rangle$ (ratio to SM fit, Figures 7 and 8), the shape of LANL PFNS is in reasonable agreement (inside uncertainties) with old data.

Parameters of the SM (the normalization and $\langle E \rangle$) are given in Table 3. The same parameters with numerical integration of data are given for comparison.

Table 3. Parameters for the LANL result of PFNS measurements evaluated with SM for different incident neutron energies (E_0) in comparison with the results of numerical integration.

E_0 , MeV	SM Parameters (0.25 < E < 7 MeV)			Numerical (0.1 < E < 10 MeV)	
	χ^2/N	Normalization	$\langle E \rangle$, MeV	Normalization	$\langle E \rangle$, MeV
1.0–1.5	1.19	1.133 ± 0.006	2.056 ± 0.009	1.136	2.036
1.5–2.0	0.73	1.127 ± 0.007	2.077 ± 0.012	1.130	2.069
2.0–3.0	0.72	1.132 ± 0.007	2.073 ± 0.011	1.134	2.048
3.0–4.0	0.60	1.133 ± 0.007	2.100 ± 0.012	1.127	2.075
4.0–5.0	0.86	1.123 ± 0.007	2.128 ± 0.013	1.126	2.093

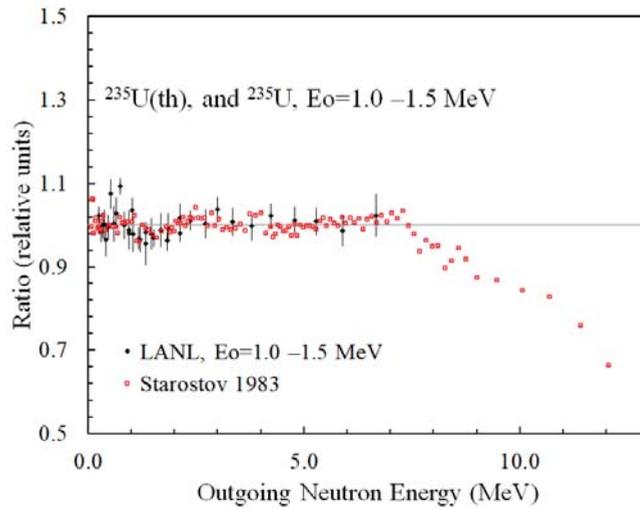


Figure 7. Ratio of experimental data for PFNS to the calculated SM functions. Starostov’s data for thermal neutrons are shown for comparison.

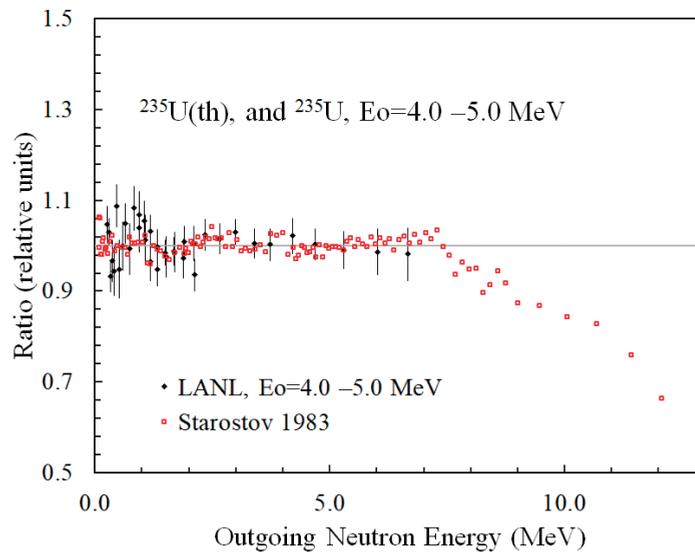


Figure 8. Ratio of experimental data for PFNS to the calculated SM functions. Starostov’s data for thermal neutrons are shown for comparison.

The analysis with SM demonstrated very interesting peculiarities:

1. The SM approach describes LANL data within the experimental uncertainties (χ^2/N of order 1). The numerical integration provides the same normalization, but $\langle E \rangle$ is lower. It can be related to the LANL spectra behavior below 0.25 MeV. The numerical integration with Maxwellian spectrum gave ~20 keV difference for $\langle E \rangle$ calculated in the 0–10 and 0–20 MeV ranges.
2. Experimental PFNS shown in the figures in [7,8] were not normalized to unity. SM and numerical integration are in reasonable agreement.

3. Average energies estimated from LANL data contradict the majority of previous measurement results (see Table 3 and Figure 9). The energy dependence obtained from the fit of the results of the old measurements is $\langle E \rangle = 1.978 + 0.020 \times E_0$, MeV [7]. The fit to LANL data (only) gives $\langle E \rangle = 2.032 + 0.020 \times E_0$, MeV. There is a clear shift in more than 30 keV for $\langle E \rangle$ between LANL and DB measurements.
4. The LANL results are supported only by the data from [11], which were not included in the fitting procedure and discussed above in this paper. The conclusion was that this experiment most probably contains the SDF. It is interesting that if [11] used multi-detector registration of fission fragments, LANL used multi-detector neutron registration.
5. The ratio of LANL data at 4.5 MeV to the Maxwellian spectrum with $kT = 1.379$ MeV was compared in Figure 10 with the ratio of IPPE data [15] at 5 MeV to the Maxwellian spectrum with $kT = 1.385$ MeV. The agreement is good. PFNS spectra demonstrate a rather broad “bump” in the energy range 2–6 MeV with the average ratios 1.045 (LANL) and 1.015 (IPPE). Uncertainties of the PFNS are large enough to make the conclusion clear about the SDF presence. However, the difference in the spectrum average energies of 3% between LANL and other data (Figure 9) evaluated with the SM with low uncertainties (0.5–1%) shows that LANL result may contain the SDF.

One may assume that the origin of the SDF in double TOF LANL results can be some unaccounted background neutrons. These neutrons reached the chamber at the same time interval due to a longer flight path. These background neutrons cause fission at higher neutron energy, and as a consequence, they contribute to a higher average neutron energy of PFNS.

SDF connected with the time structure of the proton beam or other causes of SDFs are also possible.

One may assume that LANL results consistent with the results in [11] are accurate and that other measurements performed in the years 1983 to 2018 contain the SDF. This seems a doubtful assumption, at least concerning the results [11]. As discussed in Section 2.3, the results of the measurements [11] contain rather strong SDF, and this should be accounted for in the PFNS evaluation procedure.

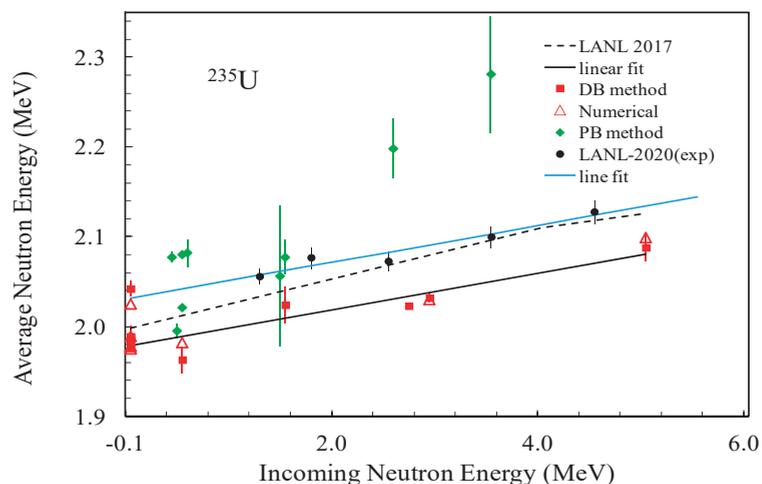


Figure 9. Comparison of ^{235}U PFNS average neutron energies estimated with the SM for new LANL-2020 double TOF measurements [13] (closed black circles, Table 3) with the results of DB and PB measurements. The open red triangles present the results of numerical integration when calculating the average energy. The thin blue line is a linear fit of LANL-2020 data, and the black line is a linear fit of the results of DB measurements. The data were taken from [7].

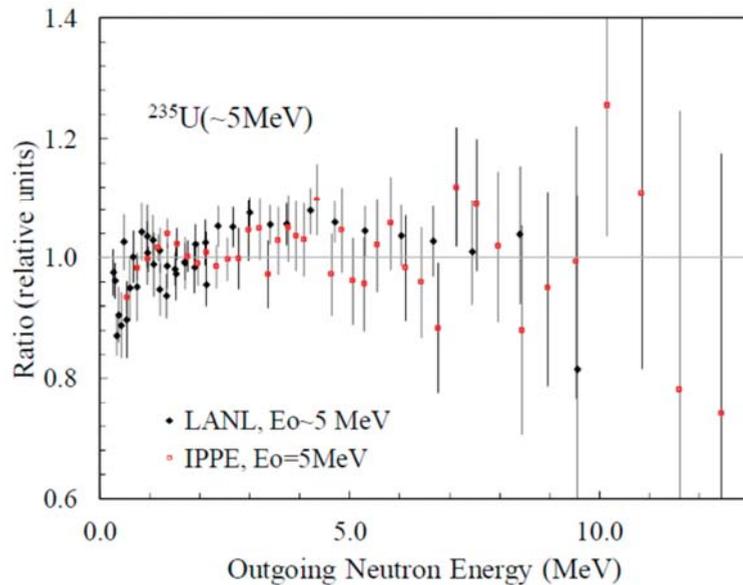


Figure 10. Ratios of the measured PFNS to the Maxwellian spectra (see details in the text).

3. SDF in Fission cross Section Ratio Measurements

A new experiment using the Time Projection Chamber (TPC) for registration of FF was implemented in Los Alamos National Laboratory (LANL) by a team from different universities and laboratories (NIFTE collaboration). The results of the absolute ratio of $^{239}\text{Pu}(n,f)$ to $^{235}\text{U}(n,f)$ cross section measurements are shown in Figure 11. The data were taken from [16].

The constant bias at about 2% is clearly visible in the energy range 0.2–15 MeV and has an even larger spread above 15 MeV. The authors of [16] provided a very detailed and deep analysis of the modeling of their experiment (experimental details, data reduction procedure, uncertainties of different parameters and so on) but at present could not explain the existing bias, which can be treated as SDF. From our point of view, it is premature to assign the SDF to this measurement, especially taking into account that the reaction rate ratios measured in clean benchmarks show better consistency with the NIFTE results.

The authors of [16] came to the conclusion that the difference in 2% absolute normalization with the ENDF/B-VIII.0 [20] evaluation based on its turn on neutron standards evaluation [2] is too large to be ignored. The large non-uniformity and mass value of the ^{239}Pu sample, which can degrade with time, are the largest concern. Although the TPC results with multi-parametric data allow an estimate of many sources of systematic uncertainties (or potential SDF by other words), the present decision is to repeat the measurements with a newly prepared ^{239}Pu sample. At the same time, we should admit that the results of the reaction rate measurements obtained as benchmarks with fast neutron spectrum and Mannhart's evaluations [21] of ^{252}Cf PFNS-averaged cross sections endorse the data obtained with the TPC.

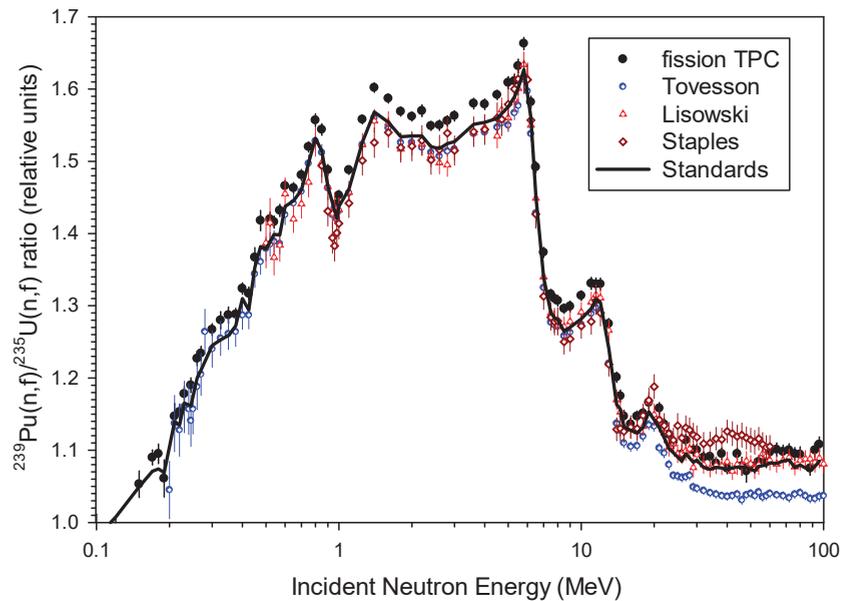


Figure 11. $^{239}\text{Pu}(n,f)/^{235}\text{U}(n,f)$ cross section ratio measured with a fission TPC [16] in comparison with the data [17–19] and the results of the GMA fit (solid line) with fission TPC data added to the standard (old) database.

4. SDF in Maxwellian Averaged cross Sections (MACS) for Astrophysical Application

MACS for neutron capture are used in astrophysics to model the stellar nucleosynthesis of elements. The range of needed neutron temperatures (kT) is varied from a few keV to 100 keV. The novel method of direct MACS measurement for $kT = 25\text{--}30$ keV was proposed by Beer and Kaeppler [22]. It is based on the kinematics of $^7\text{Li}(p,n)$ reaction at a proton energy of 1912 keV. As it was shown experimentally and through modeling, the neutrons in this case are emitted in the forward cone with the spectrum integrated on the angles close to the Maxwellian spectrum with kT between 25 and 30 keV.

Ratinsky and Kaeppler recorded the accurate activation MACS measurements [23] for $^{197}\text{Au}(n,\gamma)$ (582 ± 9 mb at $kT = 30$ keV, stellar definition), which were used as the standard for measurements of other nuclides by this method. Slightly renormalized results of Macklin's $^{197}\text{Au}(n,\gamma)$ microscopic cross section, which provided this MACS value, were used in calculations for extrapolation of calculated MACS to lower and higher kT . These values were inconsistent with MACS calculated for the evaluation of $^{197}\text{Au}(n,\gamma)$ standards [22] based on a combined fit of 62 measurements of captured cross sections and their ratios to other standard reactions. MACS calculated for standard cross section evaluation in the energy range 5 keV–2.8 MeV, supplemented by the ENDF/B-VII.0 evaluation below and above this energy range, was 614 mb at $kT = 30$ keV. The ENDF/B-VII.0 evaluation had missed resonances at the upper end of the resolved resonance range. The correction at the missed resonances increases the MACS value to 619 mb [24].

This controversy and new results of nTOF [25] and GELINA [26] measurements of microscopic capture cross sections, consistent with the standards evaluation, led to a new cycle of measurements and analysis of the MACS. To resolve the discrepancy, PINO [27] and SimLiT [28] Monte Carlo codes were developed for modeling the neutron source. GEANT4 code was used for modeling the neutron transport. There have been a number of publications, but the latest review of the Spectrum Averaged Cross Section (SACS) measurements and re-evaluation is published in [29]. Two major sources that may cause bias in the MACS values were discussed [29]: attenuation of the neutron flux at the copper

backing of the ^7Li target, and the difference between the measured neutron spectrum and “true” neutron spectrum incident at the gold sample following a reduction in the measured value to the MACS value at $kT = 30$ keV.

Simulation of the neutron source had shown [29] that backing of the ^7Li target with copper of 1 mm thickness requires the introduction of neutron scattering correction between 6.4 and 7.1% depending on the nuclear data library used in simulation calculations. The comparison with experimental data obtained for different thicknesses of backings and size of deposited ^7Li targets shows [29] that the large part of the discrepancy with MACS obtained with the standards evaluation and calculated with the latest experimental data [25,26] may be explained by an improper account of backing in the activation measurements. The authors of [29] were unable to introduce corrections in the results [23] at the base of their simulation because of the inconsistencies between experimental data [23] after introducing corrections based at the simulation.

The measured neutron spectrum induced by 1912 keV protons and integrated on the angles is similar to the 25.3 keV Maxwellian neutron spectrum [23] but shows a clear lack of neutrons above 80 keV. The neutron temperature, which should be best assigned to the measured MACS values, has rather large uncertainty. The comparison of the measured spectrum [23] with Maxwellian spectra at $kT = 25.3$ (best fitted to the experimental simulated spectrum [23]) and $kT = 28.5$ keV (obtained from calculated mean energy of the experimental simulated spectrum [24]) is shown in Figure 12. It was estimated [24] that, depending on the procedure of the reduction to the true Maxwellian spectrum, the difference between the measured value [23] at the best assigned temperature and the calculated value may reach 2.5%. A similar difference (1.7%) is shown in [29] between recommended values obtained from the latest measurements of SACS and SACS calculated for the true Maxwellian spectrum at the same temperatures.

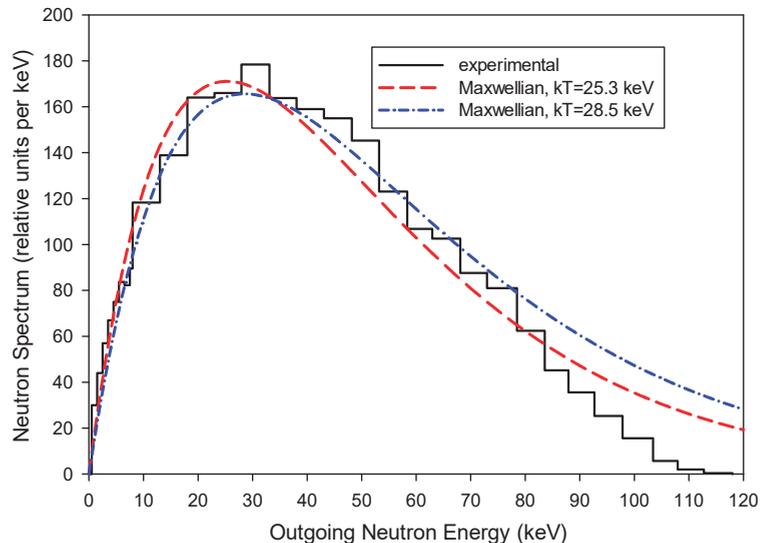


Figure 12. Comparison of experimental neutron spectrum with a true Maxwellian neutron spectrum for two temperatures $kT = 25.3$ and 28.5 keV. All spectra have free normalization.

A series of new measurements of MACS with a $^7\text{Li}(p,n)$ reaction for different thicknesses of lithium target and spectrum of protons incident on the target supported by modeling of the experiment had allowed obtaining the best simulation of the Maxwellian-like spectra for different temperatures [29]. A new version of the neutron cross section standards [2] took into account the results of the new measurements [29] of $^{197}\text{Au}(n,\gamma)$

cross sections. MACS at $kT = 30$ keV calculated for the standard evaluation in the energy range 3 keV–2.8 MeV embedded in the ENDF/B-VIII.0 evaluation provides the new recommended value of 611.4 ± 4.2 mb with the uncertainty increasing to 11.2 mb when USU is accounted for. This value has excellent consistency with the new value 612 ± 6 mb [29] recommended for use by the astrophysical community.

We may conclude that the modeling of the neutron source, including the energy angular correlation and attenuation of neutrons in lithium target backing excludes the bias (~5%) caused by SDF from the results of SACS measurements. The proof of this is the consistency between results derived from simulated SACS measurements and calculated from the evaluated cross sections obtained in the independent microscopic cross section measurements.

5. Unrecognized Sources of Uncertainty (USU) in the Data Evaluation

5.1. Small Uncertainties Problem in Neutron Cross Section Standards Evaluation and USU

In 1991, the Cross Section Evaluation Working Group (CSEWG) concluded that the uncertainties of the evaluated neutron cross section standards [30] are strongly underestimated. The standards were obtained in the combined model-independent statistical fit of about 400 data sets for 10 reactions and their combinations. The relative uncertainties obtained from variances of the covariance matrix of the evaluated standards were two to three times lower than the spread of the experimental data estimated for the same broad energy groups.

The spread of the experimental data can be best characterized by the variances of the evaluation obtained with the use of the sample statistics. For this, the model-independent least square fit of the data in the energy groups can be performed without consideration of the uncertainties assigned to the data. The uncertainty of integral data calculated with an account of the evaluated covariance matrices (such as SACS for $^{197}\text{Au}(n,\gamma)$ reaction discussed above) was also considered as too small. This can be partly explained by an incomplete budget of uncertainty sources for some measurements and absence or not a full account of cross-correlations between the same components of uncertainties in different measurements, which use the same sample or detector, or even method. This conclusion remains generally true with the revision of the outlying experimental data uncertainties.

The difference in the data values obtained in different experiments, which cannot be explained by uncertainties assigned to them, indicates the presence of the USU [31]. In the case when better consistency cannot be achieved through the revision or introduction of the corrections based on the Monte Carlo modeling of the experiments (SDF removing), the additional uncertainty can be introduced in the evaluated covariance matrix, making the evaluated uncertainties more realistic.

This approach can be applied to the neutron cross section standards evaluation [2]. The GMA code [32] for model-independent evaluation of the standards uses the GLSQ method for a simultaneous fit of the cross sections and integral parameters with an iterative approach. Starting from the second iteration, a posterior evaluation for data values with an uninformative covariance matrix is used as a new prior. Usually, three iterations are needed to obtain full convergence when the last posterior data evaluation is practically indistinguishable from the last prior evaluation. All experimental data are reduced in the model-independent fit to a common grid of energy nodes.

The covariance matrix for each experimental data set is constructed from statistical, fully correlated systematical and medium energy range correlated to systematical components of uncertainties. The correlations between the same components of the uncertainties in different measurements can be accounted for.

5.2. Sample Method for Determination of USU Covariance Matrices

A sample method can be adapted for construction of the USU component of the evaluated covariance matrix using the biases between the evaluated and experimental data with an account of recognized (known) uncertainties. A sample method for the uncertainty

evaluation in the measurements is usually formulated for a set of repeated measurements of multivariate (vector) data [33].

For independent random vectors X_i ($i = 1, \dots, I$) of dimension n and with a zero mean, the sample covariance matrix (without disclosure of the nodes indices) is:

$$CovX_I = \frac{1}{I} \sum_{i=1}^I X_i \otimes X_i \tag{1}$$

The closeness of the sample covariance matrix to the actual covariance matrix depending on vector dimension (n) and number of samples (I) was studied in [34]. Sample covariances for the USU component can be constructed in the framework of an ad hoc procedure based on the biases between evaluated and experimental data reduced by known (and accounted for in the evaluation) systematic uncertainties.

The sample vector δ_i for the USU component can be written as:

$$\begin{aligned} \delta_i &= (y_i - \mu) - u_i, \text{ if } y_i > \mu \text{ and } \delta_i > 0; \\ \delta_i &= (y_i - \mu) + u_i, \text{ if } y_i < \mu, \text{ and } \delta_i < 0; \\ \delta_i &= 0 \text{ in all other cases,} \end{aligned}$$

where y_i is a vector of i -the experimental data set, μ is a vector of evaluated data (best approximation to the true value), $(y_i - \mu)$ is a vector of biases between experimental and evaluated data, $u_i = \sqrt{\varepsilon_i^2 + \eta_i^2}$ is a vector of total uncertainty of experimental data, ε_i is a vector of the statistical component of the uncertainty, and η_i is a vector of the systematic component of the uncertainty, which consists of two components: assigned to the analysis of experimental uncertainty $\eta_{i,exp}$ and assigned to the outlying data $\eta_{i,out}$:

$$\eta_i = \sqrt{\eta_{i,exp}^2 + \eta_{i,out}^2}$$

There is an established procedure for obtaining the evaluated values with GMA. It includes:

- Analysis and correction of the experimental data, assigning of all components of uncertainties given or not given by the authors, introducing of the correlations between the components of the uncertainties in different experiments;
- Data reduction to the same nodes on neutron energy;
- GLSQ combined fit of all data;
- Adding the component of the uncertainty to the outlying data making them consistent with the evaluation;
- New fit to obtain new posterior evaluation, redetermination of the outlying data and repeating the fit up to the convergence when no redetermination of the outlying data is needed.

This work with outlying data allows avoiding big local discrepancies and reducing the general chi-square per degree of freedom for standard evaluation from the initial 3.4 to a value close to 1 without strong local discrepancies.

The main differences to the classic sampling method are the following:

- Data in some sets can be missed at some nodes and cover different energy ranges;
- The sample vector in the node may have not a random distribution relative evaluated value;
- The number of data in some nodes can be too small to be statistically significant;
- Some data are non-normalized shape type data, which will require the additional procedure of their normalization for determination of a sample vector;

Covariances for the USU component can be written with the discloser of the node indices m and n for vectors as

$$Cov\delta^{mn} = \frac{1}{K_{mn}} \sum_{i=1}^I \delta_i^m \delta_i^n, \tag{2}$$

In our case, not all data sets contribute at the energy node m or n , and K_{mn} is a number of non-zero terms in the sum. The total covariances of evaluated data ($Cov\delta_{tot}^{mn}$) can be written as a sum of covariances obtained in the GMA fit with the ad hoc increase in the uncertainties for the outlying data ($Cov\delta_{GMA}^{mn}$) and covariance estimated for the USU component ($Cov\delta_{USU}^{mn}$).

$$Cov\delta_{tot}^{mn} = Cov\delta_{GMA}^{mn} + Cov\delta_{USU}^{mn} \tag{3}$$

The covariance matrix for the USU component (2) may turn out to be semi-positive definite, and because of this, the total covariances (3) may lose its semi-positive definite. This introduction of USU covariances can be considered as a rather crude approach to the estimation of the realistic uncertainties of the evaluated data, but it is definitely a better approach than expert estimation [30]. If we increase the uncertainties of the outlying data using stricter ad hoc requirements to the data consistency, we will reduce the USU covariances, or even exclude them. Then we have a strong connection between the treatment of the outliers and USU uncertainties, with a clear distinction that uncertainties for outliers are introduced into the experimental data iteratively in the fit, and USU uncertainties as additional components to the evaluated data.

The work with the GMA database of experimental data has shown that outliers are often the “poor” data with large uncertainties. Increasing the uncertainty of these outliers with the procedure described above reduces the chi-square per degree of freedom, changes the evaluated (mean) values, and to a lower extent, changes the covariances. The “smallness” of the evaluated uncertainties is determined mainly by the “good” experimental data with small uncertainties.

5.3. Numerical Example of USU Covariance Matrix Construction

This approach for determination of USU was applied for testing of $^{238}\text{U}(n,f)$ to the $^{235}\text{U}(n,f)$ cross section ratio evaluated with the GMA for the model case of 11 cross section ratio measurements taken as absolute in 11 nodes (Figure 13). The outliers were determined, and their uncertainties were increased.

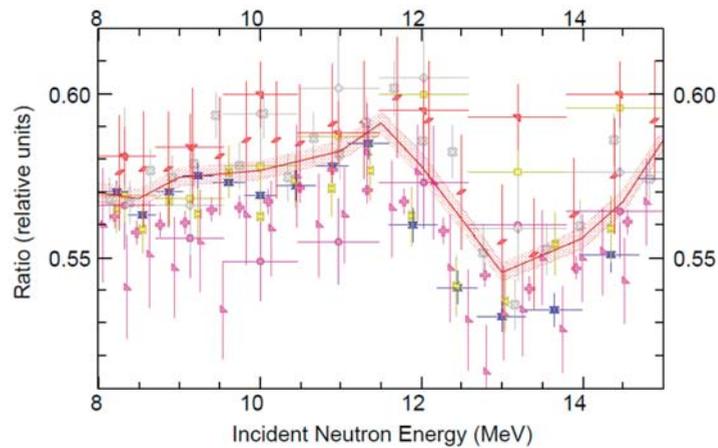


Figure 13. Experimental data used in the model case of the USU determination for the $^{238}\text{U}(n,f)/^{235}\text{U}(n,f)$ cross section ratio are shown by different symbols. GMA evaluation with an uncertainty band is shown by lines.

The results shown in Figure 14 demonstrate an increase in total percent uncertainty of up to 4–5 times in a few nodes where the spread of the data was large. The covariance matrix for the USU component has rather large positive correlations, which shows that most experimental data have normalization problems.

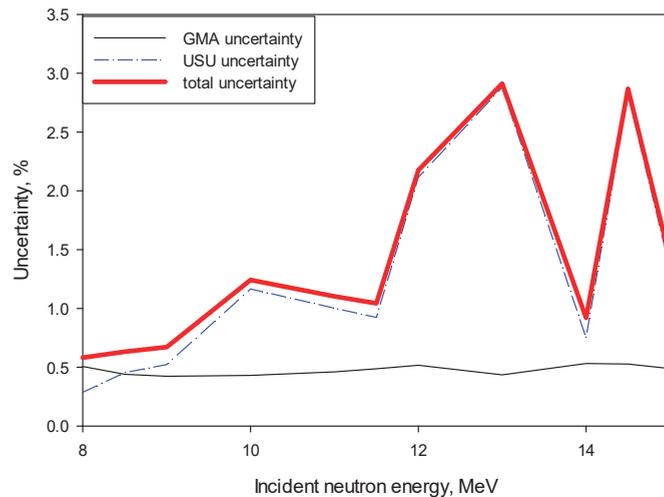


Figure 14. Contribution of the USU in the total uncertainty of evaluated data for the model case of the $^{238}\text{U}(n,f)/^{235}\text{U}(n,f)$ cross section ratio.

6. Conclusions

We analyzed several measurements and evaluations for demonstration of the SDF existence in the experimental data and their determination with the comparison of different experiments, as well as a simulation of the experimental set-up based on the models of the experiment. In cases when SDF cannot be removed from the results of measurements and considering that the spread of the experimental data has some statistical nature, the USU can be introduced in the evaluation procedure to obtain realistic uncertainties of the evaluated data. SDF can be removed through the corrections introduced in the experimental data, which lead to changes in the evaluated values. The account of the USU changes the uncertainties of the evaluated data but not the evaluated values. The introduction of the USU will not be needed if, after SDF removal, the experimental data are consistent. In cases when SDF exists in the experimental data and cannot be removed, these data, in the process of the evaluation, should be considered as outliers with their uncertainties increased up to the consistency of other experimental data or even excluded from the evaluation.

Author Contributions: Conceptualization, N.V.K., V.G.P. and S.M.G.; methodology, N.V.K., V.G.P.; software, N.V.K., V.G.P.; validation, N.V.K., V.G.P. and S.M.G.; formal analysis, N.V.K., V.G.P.; writing—original draft preparation, N.V.K., V.G.P.; writing—review and editing, N.V.K., V.G.P. and S.M.G.; visualization, N.V.K., V.G.P. and S.M.G. All authors have read and agreed to the published version of the manuscript.

Funding: This paper was prepared on the basis personal initiative by authors.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Kornilov, N.V.; Grimes, S.M. Did we solve the old contradiction between microscopic and macroscopic experimental PFNS data? *Ann. Nucl. Energy* **2021**, *162*, 10850. [CrossRef]
- Carlson, A.D.; Pronyaev, V.G.; Capote, R.; Hale, G.; Chen, Z.-P.; Duran, I.; Hamsch, F.-J.; Kunieda, S.; Mannhart, W.; Marcinkevicius, B.; et al. Evaluation of the Neutron Data Standards. *Nucl. Data Sheets* **2018**, *148*, 143. [CrossRef]
- Nefedov, V.N.; Starostov, B.I.; Boytsov, A.A. Precision Measurements of ^{252}Cf , ^{233}U , ^{235}U and ^{239}Pu Prompt Fission Neutron Spectra (PFNS) in the Energy Range 0.04–5 MeV. In Proceedings of the 6th All-Union Conference Neutron Physics, Kiev, Ukraine, 2–6 October 1983; Volume 2, p. 285.
- Starostov, B.I.; Nefedov, V.N.; Boytsov, A.A. Precision Measurements of ^{252}Cf , $^{233}\text{U}+\text{nth}$, $^{235}\text{U}+\text{nth}$ and $^{239}\text{Pu} + \text{nth}$ Prompt Fission Neutron Spectra (PFNS) in the Energy Range 2–11 MeV. In Proceedings of the 6th All-Union Conference Neutron Physics, Kiev, Ukraine, 2–6 October 1983; Volume 2, p. 290.
- Kornilov, N.V.; Hamsch, F.-J.; FABRY, I.; Oberstedt, S.; Belgya, T.; Kis, Z.; Szentmiklosi, L.; Simakov, S. The $^{235}\text{U}(\text{n},\text{f})$ Prompt Fission Neutron Spectrum at 100K Input Neutron Energy. *Nucl. Sci. Eng.* **2010**, *194*, 927–937. [CrossRef]
- Kornilov, N.V.; Hamsch, F.-J. Absolute ratio ^{252}Cf to ^{235}U Prompt Fission Neutron Spectra. *Nucl. Sci. Eng.* **2011**, *168*, 73–74.
- Kornilov, N.V. The Scale Method as an Approach for Prompt Fission Neutron Spectra Analysis and Evaluation. *Nucl. Data Sheets* **2019**, *155*, 75. [CrossRef]
- Kornilov, N.V. Open problems for neutron emission in fission. *Phys. Procedia* **2014**, *59*, 144–153. [CrossRef]
- Kornilov, N.V. Effect of Angular-Energy Selection due to Anisotropy of Fission Neutron Emission in Non-homogeneous Environment. *Int. J. Nucl. Energy Sci. Eng.* **2012**, *2*, 97–100.
- Kornilov, N.V. What is the real PFNS of ^{235}U at thermal point? *Nucl. Sci. Eng.* **2017**, *186*, 190. [CrossRef]
- Vorobyev, A.S.; Shcherbakov, O.A. *Total Prompt Fission Neutron Spectrum from Thermal-Neutron-Induced Fission of ^{235}U* ; Report INDC(CCP)-0455; IAEA: Vienna, Austria, 2013; p. 37.
- Goeeck, A.; Hamsch, F.-J.; Oberstedt, S.; Vidali, M. Prompt neutrons in correlation with fission fragments from $^{235}\text{U}(\text{n},\text{f})$. *Phys. Rev.* **2018**, *C98*, 044615.
- Kelly, K.J.; Devlin, M.J.; Taieb, J.; Marini, P.; Haight, R.C.; Neudecker, D.; Belier, G.; Laurent, B.; Morfouace, P.J.L.; White, M.C. *Summary of the NNSA/CEA Meeting on Prompt Fission Neutron Spectrum Results*; Los Alamos National Laboratory: Los Alamos, NM, USA, 2019.
- Devlin, M.; Gomez, J.A.; Kelly, K.J.; O'Donnell, J.M.; Haight, R.C.; Taddeucci, T.N.; Neudecker, D.; Wu, C.; Henderson, J.; White, M.C.; et al. Prompt Fission Neutron Spectra for Neutron-Induced Fission of ^{239}Pu and ^{235}U . *EPJ Web Conf.* **2020**, *239*, 01003. [CrossRef]
- Trufanov, A.M.; Lovchikova, G.N.; Smirenkin, G.N.; Polyakov, A.V.; Vinogradov, V.A. Measurements and Estimates of the Average Energy of Neutrons from the $\text{U}^{235}(\text{N},\text{F})$ Reaction. *Phys. At. Nuclei* **1994**, *57*, 572–578.
- Snyder, L.; Anastasioua, A.M.; Bowdena, N.S.; Bundgaard, J.; Casperson, R.J.; Cebra, D.A.; Classen, T.; Dongwi, D.H.; Fotiades, N.; Gearhart, J. Measurement of the $^{239}\text{Pu}(\text{n},\text{f})/^{235}\text{U}(\text{n},\text{f})$ Cross-Section Ratio with the NIFFTE fission Time Projection Chamber. *Nucl. Data Sheets* **2021**, *178*, 1–40. [CrossRef]
- Staples, P.; Morly, K. Neutron-induced fission cross-section ratios for ^{239}Pu , ^{240}Pu , ^{242}Pu , and ^{244}Pu relative to ^{235}U from 0.5 to 400 MeV. *Nucl. Sci. Eng.* **1998**, *129*, 149. [CrossRef]
- Lisowski, P.; Ullmann, J.; Balestrini, S.; Carlson, A.D.; Wasson, O.A.; Hill, N.W. Neutron induced fission cross section ratios for ^{232}Th , ^{235}U , ^{237}Np , and ^{239}Pu from 1 to 400 MeV. In Proceedings of the International Conference on Nuclear Data for Science and Technology, Mito, Japan, 30 May–3 June 1988; p. 97.
- Tovesson, F.; Hill, T.S. Cross sections for $\text{Pu-239}(\text{n},\text{f})$ and $\text{Pu-241}(\text{n},\text{f})$ in the range $E_n = 0.01 \text{ eV}$ to 200 MeV. *Nucl. Sci. Eng.* **2010**, *165*, 224. [CrossRef]
- Brown, D.A.; Chadwick, M.B.; Capote, R.; Kahler, A.C.; Trkov, A.; Herman, M.W.; Sonzogni, A.A.; Danon, Y.; Carlson, A.D.; Dunn, M. ENDF/B-VIII.0: The 8th Major Release of the Nuclear Reaction Data Library with CIELO-project Cross Sections, New Standards and Thermal Scattering Data. *Nucl. Data Sheets* **2018**, *148*, 1. [CrossRef]
- Mannhart, W. Status of the Evaluation of the Neutron Spectrum of $^{252}\text{Cf}(\text{sf})$. Report INDC(NDS)-0540. 2009, p. 29. Available online: <https://www-nds.iaea.org/standards-cm-oct-2008/6.PDF> (accessed on 6 August 2021).
- Beer, H.; Kaeppler, F. Neutron capture cross sections on ^{138}Ba , $^{140,142}\text{Ce}$, $^{175,176}\text{Lu}$, and ^{181}Ta at 30 keV. Prerequisite for investigation of the ^{176}Lu cosmic clock. *Phys. Rev.* **1980**, *C21*, 5341.
- Ratynski, W.; Kaeppler, F. Neutron capture cross section of ^{197}Au : A standard for stellar nucleosynthesis. *Phys. Rev.* **1988**, *C27*, 36.
- Pronyaev, V.G. $^{197}\text{Au}(\text{n},\gamma)$ Standard Cross Section and Experimental Data. Report INDC(NDS)-0540. 2009, p. 29. Available online: <https://www-nds.iaea.org/standards-cm-oct-2008/9.PDF> (accessed on 6 August 2021).
- Praena, J.; Igashira, M.; Quesada, J.; Pigni, M.T.; Jericha, E.; Plag, R.; Isaev, S.; Plompen, A.; Karadimos, D.; Cennini, P.; et al. $^{197}\text{Au}(\text{n},\gamma)$ cross section in the unresolved resonance region. *Phys. Rev.* **2011**, *C83*, 034608.
- Massini, C.; Becker, B.; Dupond, E.; Gunsing, F.; Kappeler, F.; Massimi, C.; Mengoni, A.; Wallner, A.; Abbondanno, U.; Aerts, G.; et al. Neutron capture cross section measurements for ^{197}Au from 3.5 to 84 keV at GELINA. *Eur. Phys. J.* **2014**, *A50*, 124. [CrossRef]
- Reifarth, R.; Heil, M.; Kaeppler, F.; Plag, R. PINO—A tool for simulating neutron spectra resulting from the $^7\text{Li}(\text{p},\text{n})$ reaction. *Nucl. Instrum. Methods Phys. Res.* **2009**, *A608*, 139. [CrossRef]

28. Friedman, M.; Cohen, D.; Paul, M.; Berkovits, D.; Eisen, Y.; Feinberg, G.; Giorginis, G.; Halfon, S.; Krása, A.; Plompen, A.J.M.; et al. Simulation of the neutron spectrum from the ${}^7\text{Li}(p,n)$ reaction with a liquid-lithium target at Soreq Applied Research Accelerator Facility. *Methods Phys. Res. Sect. A Accel. Spectrom. Detect. Assoc. Equip.* **2013**, *698*, 117. [[CrossRef](#)]
29. Reifarth, R.; Erbacher, P.; Fiebiger, S.; Göbel, K.; Heftrich, T.; Heil, M.; Käppeler, F.; Klapper, N.; Kurtulgil, D.; Langer, C.; et al. Neutron-induced cross sections: From raw data to astrophysical rates. *Eur. Phys. J. Plus* **2018**, *133*, 424. [[CrossRef](#)]
30. Carlson, A.D.; Poenitz, W.P.; Hale, G.M.; Peele, R.W.; Dodder, D.C.; Fu, C.; Mannhart, W. *ENDF/B-VI Neutron Cross Section Measurement Standards*, National Institute of Standards and Technology; National Institute of Standards and Technology: Gaithersburg, MD, USA, 1993.
31. Capote, R.; Badikov, S.; Carlson, A.D.; Duran, I.; Günsing, F.; Neudecker, D.; Pronyaev, V.G.; Schillebeeckx, P.; Schnabel, G.; Smith, D.L.; et al. Unrecognized Sources of Uncertainties (USU) in Experimental Nuclear Data. *Nucl. Data Sheets* **2020**, *163*, 191. [[CrossRef](#)]
32. Poenitz, W.P.; Aumeier, S.E. *The Simultaneous Evaluation of the Standards and Other Cross Sections of Importance for Technology*; Report ANL/NDM-139; Argonne National Lab.: Argonne, IL, USA, 1997.
33. ISO. *Evaluation of Measurement Data. Supplement 2 to the Guide to the Expression of Uncertainty in Measurement. Extension to Any Number of Output Quantities*; ISO/IEC GUIDE 98-3:2008/Suppl.2:2011(E); ISO: Geneva, Switzerland, 2011.
34. Vershynin, R. How close is the sample covariance matrix to the actual covariance matrix? *J. Theor. Probab.* **2012**, *25*, 655. [[CrossRef](#)]

Article

The Storage within Digital Calibration Certificates of Uncertainty Information Obtained Using a Monte Carlo Method

Ian Smith ^{1,*}, Yuhui Luo ¹ and Daniel Hutzschenreuter ²

¹ National Physical Laboratory, Teddington, London TW11 0LW, UK; yuhui.luo@npl.co.uk

² Physikalisch-Technische Bundesanstalt, 38116 Braunschweig, Germany; daniel.hutzschenreuter@ptb.de

* Correspondence: ian.smith@npl.co.uk

Abstract: Supplement 1 to the ‘Guide to the expression of uncertainty of measurement’ describes a Monte Carlo method as a general numerical approach to uncertainty evaluation. Application of the approach typically delivers a large number of values of the output quantity of interest from which summary information such as an estimate of the quantity, its associated standard uncertainty, and a coverage interval for the quantity can be obtained and reported. This paper considers the use of a Monte Carlo method for uncertainty evaluation in calibration, using two examples to demonstrate how so-called ‘digital calibration certificates’ can allow the complete set of results of a Monte Carlo calculation to be reported.

Keywords: digital calibration certificate; DCC; machine-readable; data communication; uncertainty; Monte Carlo method; MCM

Citation: Smith, I.; Luo, Y.; Hutzschenreuter, D. The Storage within Digital Calibration Certificates of Uncertainty Information Obtained Using a Monte Carlo Method. *Metrology* **2022**, *2*, 33–45. <https://doi.org/10.3390/metrology2010003>

Academic Editor: Simona Salicone

Received: 30 September 2021

Accepted: 10 January 2022

Published: 18 January 2022

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Technological advancements within the last few decades have served to digitalise many aspects of metrology. For example, instruments can be programmed to undertake time-consuming measurements with little or no need for human interaction, while the availability of greater computer processing power allows complex systems to be modelled increasingly accurately. There is one aspect of metrology, viz. the provision of calibration services, to which digital transformation has been applied in a much more modest way. Many calibration service providers continue to disseminate calibration information using paper-based certificates. Some organisations have moved to providing certificates in electronic form, for example, in archiveable Portable Document Format (PDF-A) [1]. While the provision of electronic certificates brings obvious benefits such as decreased use of paper and the potential for storage within dedicated document management systems, one undesirable property persists—a lack of machine-readability, i.e., information is not presented in a form that can be processed by computer. Currently, information on a paper-based or electronic calibration certificate can only be used if it is transcribed manually. Such a process is inevitably prone to error.

Recent initiatives have looked at how paper-based or electronic calibration certificates can be replaced by fully machine-readable certificates. The European Metrology Programme for Innovation and Research (EMPIR) [2] has funded the Joint Research Project ‘Communication and validation of smart data in IoT-networks’ (short name ‘SmartCom’) [3,4]. One objective of the SmartCom project has been to develop a framework for what are referred to as ‘digital calibration certificates’, abbreviated hereafter in this paper to ‘DCCs’. From the perspective of the SmartCom project, the critical property of DCCs is that they are fully machine-readable. It is noted that the term ‘digital calibration certificate’ has been and is used by other authors to refer to calibration certificates that take the form of electronic files but are not machine-readable.

When presenting the outcome of calibration, measurement data must be accompanied by associated uncertainty information. On a calibration certificate, it is common for a quantity value to be provided along with an associated expanded uncertainty (defined in the International Vocabulary of Metrology (VIM) [5], clause 2.35) and coverage factor (VIM [5], clause 2.38), or a coverage interval (VIM [5], clause 2.36), corresponding to a specified coverage probability (VIM [5], clause 2.37). A standard uncertainty may also be provided.

The focus of this paper is on the storage of uncertainty information obtained using the numerical approach described in the supporting document to the ‘Guide to the expression of uncertainty in measurement’ (GUM) [6] known as Supplement 1 to the GUM (GUMS1) [7]. The approach is a Monte Carlo method (MCM) for the propagation of probability distributions and is based on repeated random sampling. A key aspect underpinning the approach is the provision of a measurement model that describes mathematically how a quantity of interest (the measurand or output quantity) depends on other quantities (input quantities) to which probability distributions can be assigned. The output of an implementation of MCM (a ‘Monte Carlo calculation’) provides rich information in the form of (often hundreds of thousands of) sampled values of the measurand. The sampled values can be used to define an approximation to the probability distribution for the measurand. Summary information can be calculated using those sampled values. For example, the expectation and the standard deviation provide, respectively, an estimate of the measurand and its associated standard uncertainty, while a coverage interval for the measurand corresponding to a specified coverage probability can also be determined. The provision of additional summary information has been considered, e.g., in [8].

For calibration services where MCM is used to undertake uncertainty evaluation, it is common for only summary information to be provided on calibration certificates. The reasons for not including the sampled values of the measurand on the certificate are understandable, e.g., the number of pages could increase significantly, and the effort required to transcribe the sampled values would make it highly unlikely that they would ever be used in practice. It is possible for the sampled values to be made available in an electronic file. When doing so, consideration must be given to aspects including the provision of additional information such as units of measurement and appropriate metadata, while the electronic file must also be transmitted using a suitably secure means that ensures the file cannot be corrupted.

Should information about the measurand be required as input to a subsequent calculation, it is common, in the absence of any other information, for a Gaussian (normal) distribution, with expectation and standard deviation given, respectively, e.g., by the estimate and standard uncertainty quoted on the calibration certificate, to be assigned to the measurand. Such an assignment is often made even though the true probability distribution may be significantly different. The quality of the result of the subsequent uncertainty calculation may be significantly influenced by the assumption of normality. Were the sampled values generated by MCM available, one could instead implement MCM for the subsequent calculation by drawing randomly from those values. The SmartCom project has developed a data model that allows measurement data and associated uncertainty information to be stored in digital form. The model builds upon the International System of Units (SI) [9], the globally-agreed system of measurement units that has at its heart the seven base units of the kilogram, metre, second, ampere, kelvin, mole and candela. The data model, referred to as the ‘Digital SI’ (frequently shortened simply to ‘D-SI’) [10], allows the representation of quantities that are real or complex, and univariate or multivariate.

This paper focuses on how the D-SI allows uncertainty information, including the complete set of results of a Monte Carlo calculation, to be provided within a DCC. While the GUMS1 approach to uncertainty evaluation is well-established, DCCs are a much more recent development and the potential overlap between GUMS1 and DCCs has not previously been discussed. Consideration is given to the cases where the measurand is real and univariate, i.e., a single real quantity, and real and multivariate, i.e., comprises

more than one real quantity. Section 2 provides a brief summary of uncertainty evaluation undertaken using MCM for both cases. Section 3 introduces the main components of the DCC and outlines how measurement data and associated uncertainty information for real quantities can be encapsulated in the D-SI. Section 4 describes two examples, the first relating to the measurement of a univariate real quantity and based on an example in GUMS1, the second relating to the measurement of a multivariate real quantity and based on an example in Supplement 2 to the GUM (GUMS2) [11]. Concluding remarks are presented in Section 5.

Note that this paper does not discuss technical and legal aspects associated with the generation, delivery and use of DCCs. Such aspects are considered in, e.g., [12].

2. The Monte Carlo Method

Uncertainty evaluation is generally considered to comprise two stages, formulation and calculation. The stages are summarised below for the cases where the measurand is real and univariate, and real and multivariate.

2.1. Univariate Real Quantity

1. The formulation stage involves the following steps:

- Identification of the measurand Y and the input quantities $\mathbf{X} = (X_1, \dots, X_N)^\top$ on which the measurand depends.
- Assignment of the mathematical relationship between the measurand and the input quantities, e.g.,

$$Y = f(\mathbf{X}).$$

- Assignment of probability distributions for the input quantities. The quantities may all be independent, in which case each quantity is assigned a probability distribution, or there may be correlation between some of the quantities, meaning that a joint probability distribution is assigned to those quantities.

2. The calculation stage, when implementing MCM, involves the following steps:

- Assign a number M of trials.
- For $k = 1, \dots, M$, sample values $x_{1,k}, \dots, x_{N,k}$ from the probability distributions for the input quantities and evaluate

$$y_k = f(x_{1,k}, \dots, x_{N,k}).$$

- Calculate the estimate y of the measurand and its associated standard uncertainty $u(y)$ given, respectively, by the expectation and standard deviation of the values y_k , $k = 1, \dots, M$.
- Use the approximation to the distribution function for the measurand to determine a coverage interval corresponding to a specified coverage probability.

Note that, both for simplicity and to reflect the choice of examples in Section 4, this section considers only the case of a measurement model that can be classified as explicit, i.e., the measurand can be expressed as an explicit mathematical function of the input quantities. Variants of MCM are available for the case where the relationship between the measurand and input quantities cannot be expressed explicitly. The storage within DCCs of uncertainty information obtained using MCM is equally applicable to explicit and implicit measurement models. In addition, again for simplicity, this section considers a particular implementation of MCM where a fixed number M of trials is assigned in the first step. Alternative applications of MCM can, and possibly should, be implemented, e.g., an adaptive approach as described in clause 7.9 of GUMS1 [7] where an increasing number of trials are carried out until results are deemed to have stabilised sufficiently according to predetermined criteria.

2.2. Multivariate Real Quantity

1. The formulation stage involves the following steps:

- Identification of the measurand $\mathbf{Y} = (Y_1, \dots, Y_m)^\top$ and the input quantities $\mathbf{X} = (X_1, \dots, X_N)^\top$ on which the measurand depends.
- Assignment of the mathematical relationship between the measurand and the input quantities, e.g.,

$$\mathbf{Y} = \mathbf{f}(X_1, \dots, X_N) \equiv (f_1(\mathbf{X}), \dots, f_m(\mathbf{X}))^\top.$$

2. The calculation stage, when implementing MCM, involves the following steps:

- Assign a number M of trials.
- For $k = 1, \dots, M$, sample values $x_{1,k}, \dots, x_{N,k}$ from the probability distributions for the input quantities and evaluate

$$\mathbf{y}_k \equiv (y_{1,k}, \dots, y_{m,k})^\top = \mathbf{f}(x_{1,k}, \dots, x_{N,k}).$$

- From the values $\mathbf{y}_k, k = 1, \dots, M$, calculate an estimate $\mathbf{y} = (y_1, \dots, y_m)^\top$ of the measurand and its associated covariance matrix

$$V_{\mathbf{y}} = \begin{bmatrix} u^2(y_1) & u(y_1, y_2) & \dots & u(y_1, y_{m-1}) & u(y_1, y_m) \\ u(y_2, y_1) & u^2(y_2) & \dots & u(y_2, y_{m-1}) & u(y_2, y_m) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ u(y_{m-1}, y_1) & u(y_{m-1}, y_2) & \dots & u^2(y_{m-1}) & u(y_{m-1}, y_m) \\ u(y_m, y_1) & u(y_m, y_2) & \dots & u(y_m, y_{m-1}) & u^2(y_m) \end{bmatrix},$$

where $u(y_i)$ is the standard uncertainty associated with y_i and $u(y_i, y_j) \equiv u(y_j, y_i)$ is the covariance associated with y_i and y_j .

- Use the approximation to the distribution function for the measurand to determine a coverage region corresponding to a specified coverage probability.

3. Digital Calibration Certificates

3.1. Overview

A DCC [13] provides all information relating to a calibration in machine-readable form. Its structure mirrors the information that is required by ISO/IEC 17025 [14] for reporting the results of calibration. Consequently, a DCC is divided into four main sections:

- Administrative data (compulsory, regulated)—this section contains information that is typically displayed on the front page of a paper-based certificate. For example, identification of the calibration laboratory, the calibration object and the calibration service customer.
- Measurement results (compulsory, partially regulated)—this section allows measurement results, including uncertainty information, from different metrology domains and of different types to be presented. Currently, only measurement results that rely on the International System of Units (SI) can be provided in this section.
- Comments (optional, not regulated)—this section contains non-regulated information that is specifically intended for humans, e.g., proprietary data such as calibration-specific data sheets, formatting information, etc., and that cannot be used by computer without the need for human interpretation. The section may include graphical, video or audio information.
- Document (optional)—this section allows a human-readable version of the calibration certificate to be stored and allows users to view an electronic version of the certificate more akin to the traditional paper-based certificate.

The focus of this paper is on uncertainty-related aspects of the measurement results section of a DCC.

3.2. Measurement Results Section

A number of internationally recognised documents provide the foundation for the representation of measurement data within the D-SI: the BIPM SI brochure [9], the GUM [6], the VIM [5] and ISO 80000-1 [15]. The D-SI allows several types of measurement data—real, complex, univariate, multivariate, etc.—to be represented.

3.2.1. Univariate Real Quantity

Table 1 lists the components of the D-SI for a univariate real quantity. The information marked in bold shows the minimal information required and comprises the numerical value of the quantity and the unit of measurement. Additional information may optionally be provided: a label providing descriptive information for the quantity, e.g., to provide metadata related to the measurement, and the date and time of the measurement.

Table 1. Component structure for a univariate real quantity. Components in bold are mandatory. The notation ‘>’ indicates that the component on the right is a subcomponent of the component on the left.

real	>	value
	>	unit
	>	label
	>	dateTime

In practice, a measurement result is generally considered to be incomplete if it is not accompanied by information regarding its uncertainty. The D-SI allows for uncertainty information by extending the basic concept in Table 1 to allow the provision of an expanded uncertainty or a probabilistically symmetric coverage interval.

Tables 2 and 3 list the components of the D-SI that can be used to represent a univariate real quantity with uncertainty information provided in the form of an expanded uncertainty and a coverage interval, respectively. For each case, uncertainty information is provided using an additional, optional, component which itself comprises a number of mandatory and optional components. Therefore, for example, **uncertainty** is a subcomponent of **expandedUnc** which is itself a subcomponent of **real**. Note that the unit of measurement is not explicitly provided for the uncertainty information but is implicitly inherited from the unit component.

Table 2. Component structure for a univariate real quantity with expanded uncertainty.

real	>	value	
	>	unit	
	>	expandedUnc	>
			>
			>
			>
	>	label	
	>	dateTime	

For expanded uncertainty, in addition to the numerical value and the unit of measurement, the mandatory information comprises the coverage probability and the expanded uncertainty and coverage factor corresponding to that coverage probability. The expanded uncertainty and coverage factor allow the standard uncertainty to be determined if required. Information about the probability distribution, e.g., distribution type, parameter values, may optionally be provided.

For a coverage interval, the additional mandatory information comprises the standard uncertainty, the coverage probability and the lower and upper limits of the coverage

interval corresponding to the coverage probability. Again, information about the probability distribution may optionally be provided.

Table 3. Component structure for a univariate real quantity with coverage interval.

real	>	value		
	>	unit		
	>	coverageInterval	>	standardUnc
			>	intervalMin
			>	intervalMax
			>	coverageProbability
			>	distribution
	>	label		
	>	dateTime		

3.2.2. Multivariate Real Quantity

Frequently in metrology, it is necessary to consider a multivariate real quantity, i.e., a vector of real quantities. The D-SI allows multivariate quantities to be treated by employing a ‘list’ structure. In its most general form, the list structure allows a multivariate real quantity to be represented as shown in Table 4, where each element is of a type specified in Tables 1–3.

Table 4. Component structure for a multivariate quantity comprising a series of real quantities.

list	>	real
	>	real
		⋮
	>	real

GUMS2 [11] describes how both hyper-ellipsoidal and hyper-rectangular coverage regions can be defined for multivariate quantities. Tables 5 and 6 list the components of the D-SI that can be used to represent a multivariate real quantity with uncertainty information provided in the form of a hyper-ellipsoidal coverage region and hyper-rectangular coverage region, respectively. For each case, uncertainty and covariance information is provided in the form of an additional component which itself comprises a number of components.

Table 5. Component structure for a multivariate real quantity with a hyper-ellipsoidal coverage region.

list	>	real		
	>	real		
		⋮		
	>	real		
	>	ellipsoidalRegion	>	covarianceMatrix
			>	coverageFactor
			>	coverageProbability
			>	distribution

Table 6. Component structure for a multivariate real quantity with hyper-rectangular coverage region.

list	>	real		
	>	real		
		⋮		
	>	real		
	>	rectangularRegion	>	covarianceMatrix
			>	coverageFactor
			>	coverageProbability
			>	distribution

Consider the covariance matrix

$$V_y = \begin{bmatrix} u^2(y_1) & u(y_1, y_2) \\ u(y_2, y_1) & u^2(y_2) \end{bmatrix}$$

of size 2×2 . Table 7 lists the components of the D-SI that can be used to represent V_y . Information is provided one column at a time, starting at column one, and within each column information is presented one row at a time, starting at row one. Therefore, in Table 7, information is presented in the order $u(y_1, y_1) \equiv u^2(y_1)$, $u(y_2, y_1)$, $u(y_1, y_2)$ and $u(y_2, y_2) \equiv u^2(y_2)$. The approach generalises straightforwardly for covariance matrices of larger size.

Table 7. Component structure for a covariance matrix of size 2×2 .

covarianceMatrix	>	column	>	covariance	>	value
					>	unit
			>	covariance	>	value
					>	unit
	>	column	>	covariance	>	value
					>	unit
			>	covariance	>	value
					>	unit

A multivariate quantity may consist of multiple measurements of quantities of the same type, e.g., measurements of temperature at a particular location taken at regular time intervals, or quantities of different types, e.g., measurements of different environmental factors within a laboratory. In the former case, if all quantities have the same unit of measurement, the list structure as presented leads to unnecessary repetition of information. The D-SI has been adapted to allow for more efficient representation in such cases. Table 8 shows how the same unit of measurement may be assigned to all individual quantities in a vector of real quantities. (For ease of reading, the optional label and dateTime components have been omitted.) When the **listUnit** component is used, there is no longer the mandatory requirement to provide a unit of measurement for each quantity (c.f. Table 1).

Table 8. Component structure for a multivariate quantity comprising a series of real quantities with the same unit of measurement.

list	>	listUnit		
	>	real	>	value
	>	real	>	value
		\vdots		
	>	real	>	value
	>	real	>	value

The D-SI also allows for the same expanded uncertainty or coverage interval to be associated with all the real quantities of a multivariate quantity but this functionality is not discussed further in this paper. A complex quantity is treated as a special case of a multivariate quantity and is considered in [16].

3.2.3. Matrices and Tensors

The D-SI allows data of dimension higher than that of vectors to be treated using a 'list of lists' structure. This structure allows, e.g., a matrix to be represented as shown in Table 9. Depending on the application and the preference of the user, a matrix may be represented as a list of rows or a list of columns.

Table 9. Component structure for a matrix comprising a list of lists.

list	>	list
	>	list
		:
		:
	>	list

3.3. *Implementation of the Data Model*

The D-SI may be implemented in the language of choice of the user. Within the SmartCom project, Extensible Markup Language (XML) [17] has been used, but the use of alternative languages such as JavaScript Object Notation (JSON) [18] is equally valid. The following features have been imposed:

- The expression of numerical values is compatible with decimal floating-point numbers in the ANSI/IEEE 754 double precision format [19].
- Date and time information is presented relative to Universal Coordinated Time (UTC) and complies with the format described in ISO 8601 [20] for legal local date and time with a difference to UTC.
- The bases for the expression of units of measurement are the BIPM SI brochure [9], the siunitx package for LaTeX [21] and IEC TS 62720 [22].
- Standard Unicode Transfer Format 8-bit (UFT-8) is to be used for all character strings including those that indicate numerical values.

Within the XML implementation of the D-SI [23], to reflect the dependence on the SI, all structural elements have the prefix ‘si’, e.g., si:real.

4. **Examples**

4.1. *Univariate Real Quantity*

To illustrate the ability of DCCs to capture uncertainty information obtained from MCM, consider the example of gauge block calibration from clause 9.5 of GUMS1 [7]. In this example, the length of a gauge block of nominal length 50 mm is determined by comparing it with a known reference standard that has the same nominal length. The measurand is the deviation from the nominal length.

The measurand is expressed as an explicit function of nine input quantities. The probability distributions assigned to the input quantities comprise scaled and shifted t-distributions, a rectangular distribution, a normal distribution, an arc sine distribution and rectangular distributions with inexactly prescribed limits.

For this paper, the calculation stage has been implemented using a fixed number $M = 10^6$ of trials (c.f. the implementation in GUMS1 where an adaptive approach was used). Using the component structure of Table 3, summary information from the calculation stage can be encapsulated as follows:

```

<!-- MCM, 1e6 samples - Summary information -->
<si:real>
  <si:label>Deviation from nominal length</si:label>
  <si:value>838</si:value>
  <si:unit>\nano\metre</si:unit>
  <si:coverageInterval>
    <si:standardUnc>36</si:standardUnc>
    <si:intervalMin>745</si:intervalMin>
    <si:intervalMax>932</si:intervalMax>
    <si:coverageProbability>0.99</si:coverageProbability>
  </si:coverageInterval>
</si:real>

```

Using the component structure of Table 8, the full set of values y_k , $k = 1, \dots, M$, of the output quantity returned by the Monte Carlo calculation can be encapsulated as follows, showing only the first three ($k = 1, 2, 3$) and final three ($k = M - 2, M - 1, M$) values:

```
<!-- MCM, 1e6 samples - Output quantity values -->
<si:list>
  <si:listUnit>\nano\metre</si:listUnit>
  <si:real>
    <si:value>829.5221</si:value>
  </si:real>
  <si:real>
    <si:value>873.3864</si:value>
  </si:real>
  <si:real>
    <si:value>822.9225</si:value>
  </si:real>
  ...
  <si:real>
    <si:value>825.8857</si:value>
  </si:real>
  <si:real>
    <si:value>862.1964</si:value>
  </si:real>
  <si:real>
    <si:value>798.6789</si:value>
  </si:real>
</si:list>
```

As discussed in Section 1, it would be impractical to generate a paper certificate that contains such a large number of numerical values. When written to file (without any spaces or indentation), the full set of values in the format above takes up approximately 48.6 MB (and requires $3M + 3$ lines). For comparison, were the representation that allows the same unit of measurement to be assigned to all quantities not available, using the component structure of Table 4 would lead to a file of approximate size 78.2 MB ($4M + 2$ lines).

4.2. Multivariate Real Quantity

To illustrate the ability of DCCs to present information for a multivariate real quantity, consider the example of simultaneous measurement of resistance and reactance from clause 9.4 of GUMS2 [11]. In this example, the resistance and reactance of a circuit element are determined by measuring the amplitude of a sinusoidally-alternating potential difference across its terminals, the amplitude of the alternating current passing through it, and the phase angle of the alternating potential difference relative to the alternating current. There are three output quantities: the resistance, the reactance and the impedance.

For this paper, the calculation stage has been implemented using a fixed number $M = 10^6$ of trials. Using the component structure of Table 8, summary information from the calculation stage can be encapsulated as follows:

```
<!-- MCM, 1e6 samples - Summary information -->
<si:list>
  <si:listUnit>\ohm</si:listUnit>
  <si:real>
    <si:value>127.732</si:value>
  </si:real>
  <si:real>
    <si:value>219.847</si:value>
  </si:real>
```

```

<si:real>
  <si:value>254.260</si:value>
</si:real>
<si:ellipsoidalRegion>
  <si:covarianceMatrix>
    <si:column>
      <si:value>0.003364</si:value>
      <si:unit>\ohm\ohm</si:unit>
      <si:value>-0.04090216</si:value>
      <si:unit>\ohm\ohm</si:unit>
      <si:value>-0.02688114</si:value>
      <si:unit>\ohm\ohm</si:unit>
    </si:column>
    <si:column>
      <si:value>-0.04090216</si:value>
      <si:unit>\ohm\ohm</si:unit>
      <si:value>0.058081</si:value>
      <si:unit>\ohm\ohm</si:unit>
      <si:value>0.2281734312</si:value>
      <si:unit>\ohm\ohm</si:unit>
    </si:column>
    <si:column>
      <si:value>-0.02688114</si:value>
      <si:unit>\ohm\ohm</si:unit>
      <si:value>0.2281734312</si:value>
      <si:unit>\ohm\ohm</si:unit>
      <si:value>0.037249</si:value>
      <si:unit>\ohm\ohm</si:unit>
    </si:column>
  </si:covarianceMatrix>
  <si:coverageFactor>2.80</si:coverageFactor>
  <si:coverageProbability>0.95</si:coverageProbability>
</si:ellipsoidalRegion>
</si:list>

```

Using the component structure of Table 9, the full set of values y_k , $k = 1, \dots, M$, of the output quantity returned by the Monte Carlo calculation can be encapsulated as follows, showing only the first ($k = 1$) and final ($k = M$) values:

```

<!-- MCM, 1e6 samples - Output quantity values -->
<si:list>
  <si:list>
    <si:listUnit>\ohm</si:listUnit>
    <si:real>
      <si:value>127.763</si:value>
    </si:real>
    <si:real>
      <si:value>219.303</si:value>
    </si:real>
    <si:real>
      <si:value>254.397</si:value>
    </si:real>
  </si:list>
  ...
</si:list>

```

```

    <si:listUnit>\ohm</si:listUnit>
    <si:real>
      <si:value>127.838</si:value>
    </si:real>
    <si:real>
      <si:value>220.055</si:value>
    </si:real>
    <si:real>
      <si:value>253.699</si:value>
    </si:real>
  </si:list>
</si:list>

```

When written to file (without any spaces or indentation), the full set of values in the format above takes up approximately 194 MB (and requires 12M + 2 lines).

5. Discussion

The GUM recommends that uncertainty information be presented in such a way that permits it to be used in a subsequent uncertainty calculation. On a calibration certificate, information about a quantity is frequently provided in the form of an estimate of the quantity and either an associated standard uncertainty or expanded uncertainty corresponding to a specified coverage probability. If the quantity is to be used in another calculation, it is common, in the absence of any additional information, to assign a Gaussian (or normal) probability distribution to the quantity. This assignment may be made even if the true probability distribution for the quantity is significantly different. Consequently, the results of the subsequent calculation may be unreliable.

Even if uncertainty evaluation has been undertaken using a Monte Carlo method, it is common for only the summary information mentioned above to be presented on the calibration certificate. The reasons are understandable, e.g., limiting the number of pages in the certificate to a reasonable value, the difficulty in putting to practical use the full set of values returned by the Monte Carlo method.

Digital calibration certificates (DCCs) provide two key benefits that immediately aid the reporting and use of a complete set of Monte Carlo results. First, the presentation of information in a fully machine-readable form. Second, through the application of the Digital SI (D-SI) data model, the potential to include much greater volumes of data than is currently practical in a paper-based or electronic (e.g., PDF-A) certificate.

For a subsequent uncertainty calculation, a second Monte Carlo approach can be implemented. The sample values for the first quantity can be read from the DCC and samples from (an approximation to) the probability distribution for the first quantity can be obtained by randomly sampling from that set of values. The storage and use in calculations of a large number (e.g., 10^6) of sample values is facilitated in modern personal computers and mathematical software packages.

DCCs therefore provide the means to transfer uncertainty information that is encapsulated in a set of Monte Carlo samples. However, one should be aware of the circumstances under which the use of a Monte Carlo approach does not support transferability of results, e.g., as discussed in [24].

Although the D-SI was not designed with a Monte Carlo approach to uncertainty evaluation explicitly in mind, the data components in the D-SI are readily suitable for reporting the full set of results from a Monte Carlo calculation. As its use becomes more widespread, it is anticipated that the D-SI will be subject to updates, e.g., to take account of feedback from users from various metrology domains. One potential update could be the development of a specific component in the D-SI that allows all information from a Monte Carlo calculation to be encapsulated. That information could include details of the approach (standard or adaptive) and the number of Monte Carlo trials implemented. It

may also be useful to provide all relevant information on the input quantities and their probability distributions as well as the measurement model (in an appropriate format).

Author Contributions: Conceptualization, I.S.; methodology, I.S. and D.H.; software, I.S.; validation, D.H. and Y.L.; formal analysis, I.S.; investigation, I.S.; resources, I.S.; data curation, I.S.; writing—original draft preparation, I.S.; writing—review and editing, D.H. and Y.L.; visualization, I.S.; supervision, I.S.; project administration, I.S. All authors have read and agreed to the published version of the manuscript.

Funding: This work has been carried out as part of the European Metrology Programme for Innovation and Research (EMPIR) project 17IND02. The EMPIR initiative is co-funded by the European Union’s Horizon 2020 research and innovation programme and the EMPIR Participating States.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

PDF-A	Archiveable Portable Document Format
EMPIR	European Metrology Programme for Innovation and Research
DCC	Digital calibration certificate
VIM	International Vocabulary of Metrology
GUM	Guide to the expression of uncertainty in measurement
GUMS1	Supplement 1 to the GUM
GUMS2	Supplement 2 to the GUM
SI	International System of Units
D-SI	Digital SI
MCM	(The) Monte Carlo method
XML	Extensible Markup Language
JSON	JavaScript Object Notation
UTC	Universal Coordinated Time
UTF-8	Unicode Transfer Format 8-bit

References

1. ISO 19005-4:2020; Document Management—Electronic Document File Format for Long-Term Preservation—Part 4: Use of ISO 32000-2 (PDF/A-4). International Organization for Standardization: Geneva, Switzerland, 2020.
2. EMPIR—EURAMET. Available online: <https://www.euramet.org/research-innovation/research-empir/> (accessed on 14 September 2021).
3. SmartCom—Project. Available online: <https://www.ptb.de/empir2018/smartcom/home/> (accessed on 14 September 2021).
4. Bojan, A.; Weber, H.; Hutzschenreuter, D.; Smith, I. Communication and validation of metrological smart data in IoT-networks. *Adv. Prod. Eng. Manag.* **2020**, *15*, 107–117. [CrossRef]
5. JCGM 200:2012. *International Vocabulary of Metrology—Basic and General Concepts and Associated Terms (VIM 2008 with Minor Corrections)*; Joint Committee for Guides in Metrology: Sèvres, France, 2012. Available online: https://www.bipm.org/documents/20126/2071204/JCGM_200_2012.pdf (accessed on 9 January 2022).
6. JCGM 100:2008. *Evaluation of Measurement Data—Guide to the Expression of Uncertainty in Measurement (GUM 1995 with Minor Corrections)*; Joint Committee for Guides in Metrology: Sèvres, France, 2008. Available online: https://www.bipm.org/documents/20126/2071204/JCGM_100_2008_E.pdf (accessed on 9 January 2022).
7. JCGM 101:2008. *Evaluation of Measurement Data—Supplement 1 to the “Guide to the Expression of Uncertainty in Measurement”—Propagation of Distributions Using a Monte Carlo Method*; Joint Committee for Guides in Metrology: Sèvres, France, 2008. Available online: https://www.bipm.org/documents/20126/2071204/JCGM_101_2008_E.pdf (accessed on 9 January 2022).
8. Harris, P.M.; Matthews, C.E.; Cox, M.G.; Forbes, A.B. Summarizing the output of a Monte Carlo method for uncertainty evaluation. *Metrologia* **2014**, *51*, 243–252. [CrossRef]
9. BIPM. *The International System of Units (SI)*, 9th ed.; Bureau International des Poids et Mesures: Sèvres, France, 2019. Available online: <https://www.bipm.org/documents/20126/41483022/SI-Brochure-9-EN.pdf> (accessed on 9 January 2022).

10. Hutzschenreuter, D.; Härtig, F.; Heeren, W.; Wiedenhöfer, T.; Hackel, S.G.; Scheibner, A.; Brown, C.; Forbes, A.; Smith, I.; Rhodes, S.; et al. SmartCom. Digital System of Units (D-SI). Guide for the Use of the Metadata-Format Used in Metrology for the Easy-to-Use, Safe, Harmonised and Unambiguous Digital Transfer of Metrological Data, Version 1.3. 2019. Available online: https://zenodo.org/record/3522631#.YeF_Rv7P0dU (accessed on 9 January 2022).
11. JCGM 102:2011. *Evaluation of Measurement Data—Supplement 2 to the “Guide to the Expression of Uncertainty in Measurement” — Extension to Any Number of Output Quantities*; Joint Committee for Guides in Metrology: Sèvres, France, 2011. Available online: https://www.bipm.org/documents/20126/2071204/JCGM_102_2011_E.pdf (accessed on 9 January 2022).
12. Nikander, P.; Elo, T.; Mustapää, T.; Kuosmanen, P.; Hovhannisyanyan, K.; Maennel, O.; Brown, C.; Dawkins, J.; Rhodes, S.; Smith, I.; et al. SmartCom. Document Specifying Rules for the Secure Use of DCC Covering Legal Aspects of Metrology, Version 1.0. 2020. Available online: https://zenodo.org/record/3664211#.YeF_yv7P0dW (accessed on 9 January 2022).
13. Hackel, S.G.; Härtig, F.; Hornig, J.; Wiedenhöfer, T. The Digital Calibration Certificate. *PTB-Mitteilungen* **2017**, *127*, 75–81. [CrossRef]
14. *ISO/IEC 17025:2017*; General Requirements for the Competence of Testing and Calibration Laboratories. International Organization for Standardization and International Electrotechnical Commission: Geneva, Switzerland, 2017.
15. *ISO 80000-1:2009*; Quantities and Units—Part 1: General. International Organization for Standardization: Geneva, Switzerland, 2009.
16. Paciello, V.; De Santis, L.; Hutzschenreuter, D.; Smith, I. A universal metadata model for metrological complex quantities. In Proceedings of the 2020 IEEE International Workshop on Metrology for Industry 4.0 & IoT, Rome, Italy, 3–5 June 2020. [CrossRef]
17. Extensible Markup Language (XML) 1.0 (Fifth Edition). Available online: <https://www.w3.org/TR/xml/> (accessed on 14 September 2021)
18. Pezoa, F.; Reutter, J.L.; Suarez, F.; Ugarte, M.; Vrgoč, D. Foundations of JSON schema. In Proceedings of the 25th International Conference on World Wide Web, Montréal, QC, Canada, 11–15 May 2016.
19. *IEEE 754:2008*; IEEE Standard for Floating-Point Arithmetic. Institute of Electrical and Electronics Engineers: Piscataway, NJ, USA, 2008.
20. *ISO 8601:2004*; Data Elements and Interchange Formats—Information Interchange—Representation of Dates and Times. International Organization for Standardization: Geneva, Switzerland, 2004.
21. Wright, J. CTAN: Package Siunitx. Available online: <https://ctan.org/pkg/siunitx?lang=en> (accessed on 14 September 2021)
22. *IEC TS 62720:2017*; Identification of Units of Measurement for Computer-Based Processing. International Electrotechnical Commission: Geneva, Switzerland, 2017.
23. Hutzschenreuter, D.; Lin, S.; Loewe, J.H.; Scheibner, A.; Klobucar, R.; Acko, B.; Müller, B.; Heindorf, L. SmartCom Digital-SI (D-SI) XML Exchange Format for Metrological Data, Version 2.0.0. 2021. Available online: <https://zenodo.org/record/4709001> (accessed on 9 January 2022).
24. Wübbeler, G.; Elster, C. On the transferability of the GUM-S1 type A uncertainty. *Metrologia* **2020**, *57*, 015005. [CrossRef]

Article

The GUM Tree Calculator: A Python Package for Measurement Modelling and Data Processing with Automatic Evaluation of Uncertainty

Blair D. Hall

Measurement Standards Laboratory of New Zealand, Lower Hutt 5010, New Zealand;
blair.hall@measurement.govt.nz

Abstract: There is currently interest in the digitalisation of metrology because technologies that can measure, analyse, and make critical decisions autonomously are beginning to emerge. The notions of metrological traceability and measurement uncertainty should be supported, following the recommendations in the *Guide to the Expression of Uncertainty in Measurement* (GUM). However, GUM offers no specific guidance. Here, we report on a Python package that implements algorithmic data processing using ‘uncertain numbers’, which satisfy the general criteria in GUM for an ideal format to express uncertainty. An uncertain number can represent a physical quantity that has not been determined exactly. Using uncertain numbers, measurement models can be expressed clearly and succinctly in terms of the quantities involved. The algorithms and simple data structures we use provide an example of how metrological traceability can be supported in digital systems. In particular, uncertain numbers provide a format to capture and propagate detailed information about quantities that influence a measurement along the various stages of a traceability chain. More detailed information about influence quantities can be exploited to extract more value from results for users at the end of a traceability chain.

Keywords: measurement uncertainty; guide to the expression of uncertainty in measurement; measurement modelling; uncertainty propagation; metrological traceability; uncertain number

Citation: 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. <https://doi.org/10.3390/metrology2010009>

Academic Editors: Simona Salicone and Richard Leach

Received: 12 November 2021

Accepted: 9 March 2022

Published: 15 March 2022

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The worldwide dissemination of *Système International* (SI) units is a person-oriented paper-based process that is carefully managed by national bodies and coordinated by international organisations; however, that is about to change. The emergence of technologies that can measure and make critical decisions autonomously requires more of our measurement infrastructure to be implemented by digital systems. A growing number of initiatives are replacing paper-based, expert-oriented processes with automated digital ones (e.g., machine-readable formats for calibration reports [1] and a secure cloud-based platform for the legal metrology infrastructure in Europe [2]). Dissemination of SI provides what the metrology community calls traceability. Metrological traceability ensures that measurements are accompanied by information that can be used to determine the fitness-for-purpose of results in different situations. Traceability may be thought of as support for interoperability with measurement data but at present the expertise of skilled individuals is needed to interpret data and supporting information correctly. One outcome of digitalisation will be an ability to produce traceable measurement results in machine-actionable formats.

Measurement accuracy is fundamental to traceability. Traceable measurements must report information about the likely magnitude of the difference (error) between a measured value and the quantity intended to be measured. Metrologists refer to this as measurement uncertainty. During the 1980s, considerable effort went into harmonising the manner that measurement uncertainty is evaluated and communicated and this resulted in the

publication of the *Guide to the Expression of Uncertainty in Measurement* (GUM) [3], which remains today the primary reference for dealing with measurement uncertainty (the GUM was produced by a group of experts representing eight international scientific and technical organisations: the BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP, and OIML).

Digitalisation of metrological infrastructure will inevitably need algorithmic implementations for GUM methods. However, the GUM itself offers no specific guidance, as it was written long before this need could have been anticipated. One approach, called uncertain numbers, provides an abstract representation for physical quantities that satisfies general criteria in the GUM for an ideal format to express uncertainty [4]. Quantities in a problem, such as specific lengths, masses, etc., are always considered to have definite values that can only be estimated with limited accuracy by measurement. Therefore, an uncertain number is designed to encapsulate information about the measured value and measurement accuracy. Using this abstraction allows data processing to be expressed algorithmically in terms of the quantities involved, leaving the associated uncertainty calculations to be handled automatically in the background. The uncertain-number approach can capture the effects of influence quantities at different stages of a traceable measurement and propagate this information along the chain to an end user. Detailed information about the uncertainty budget can sometimes be used to enhance the value of results. This report describes a Python package, called the *GUM Tree Calculator* (GTC), that uses uncertain numbers [5]. GTC is a very flexible tool that has been used in two quite different applications, which have been reported on recently [6,7]. We discuss the tool's design and comment on aspects that support traceability in digital systems.

The next section provides an overview of GUM uncertainty calculations, metrological traceability, and GTC software. GTC is tested on a variety of platforms and different versions of Python 2 and 3 (refer to the github repository [5]). All code snippets use GTC version 1.3.6. The GUM presents its approach to evaluating measurement uncertainty in mathematical terms; thus, Section 2.1 summarises the key equations. However, an alternative formulation of these equations is more practical. This is described in Section 2.2, which motivates the development of the uncertain-number data type. The notion of metrological traceability is discussed in Section 2.3, and we explain why the uncertain-number format is useful to support traceability. Section 2.4 presents an example of GTC data processing applied to an electrical circuit. Section 3 looks in more detail at aspects of GTC design. The method used to automate uncertainty calculation is presented in Section 3.1. The data structures that support uncertainty propagation in uncertain-number objects are described in Sections 3.2–3.5; then, Section 3.6 describes how uncertain numbers can be saved and restored. More general considerations are discussed in Section 4. Appendix A briefly describes support for complex quantities, extensions for handling degrees of freedom, and some additional implementation details.

Notation

Upper and lower case letters are used to distinguish between quantities, which will never be determined exactly, and values that will be known (such as a numerical indication on a measuring instrument), respectively. For example, a measurement result y is written in lower-case because a result always has a definite numerical value: The value y is an estimate of Y , which is the quantity intended to be measured. Upper-case is used for Y to indicate that the quantity cannot be known exactly. It is helpful to make this distinction because uncertain-number objects are used to represent quantities; the associated numerical estimates and other known values, such as uncertainty, will appear as attributes of uncertain numbers.

2. Overview

2.1. GUM Method

Measurements are always influenced by unpredictable factors, so a result can only ever approximate the quantity of interest. Influence factors can, however, be identified and

described in probabilistic terms. In this manner, a measurement process can be represented by a mathematical model. This is the approach taken in the GUM.

The first step is to identify a function called the measurement model that *contains every quantity, including all corrections and correction factors that can contribute a significant component of uncertainty to the measurement result* [3] [Section 4.1.2]. In the GUM, this is expressed as an explicit function:

$$Y = f(X_1, X_2, \dots, X_l), \tag{1}$$

where Y is the quantity intended to be measured (the *measurand*) and the input arguments X_1, X_2, \dots, X_l are quantities that influence the measurement outcome. All arguments of $f(\dots)$ are treated in the same way when evaluating uncertainty. Some of the input terms may represent other measured quantities. For example, electrical resistance could be measured by first measuring potential difference V and current I and then evaluating ratio $R = V/I$. Other input terms may represent nuisance factors that perturb the measurement, such as Johnson noise in a resistor. The measurement model can be thought of as a recipe for evaluating the measurand: if X_1, X_2, \dots, X_l were all known exactly, then Y could be determined. However, with only approximate values available for input quantities, an approximate value for Y will be obtained.

$$y = f(x_1, x_2, \dots, x_l). \tag{2}$$

GUM uses a specific term, *standard uncertainty*, in relation to the unpredictability of measurement outcomes (i.e., the fact that y and Y differ by an unpredictable amount). A standard uncertainty is an estimate of the standard deviation of a probability distribution for the difference (error) between y and Y . There is a formula in the GUM to propagate standard uncertainties through a measurement model and obtain the standard uncertainty in a result. For a model in the form of Equation (1), the standard uncertainty of y , as an estimate of Y , is calculated as follows:

$$u(y) = \left[\sum_{i=1}^l \sum_{j=1}^l u_i(y) r(x_i, x_j) u_j(y) \right]^{1/2}, \tag{3}$$

where $u_i(y)$ and $u_j(y)$ are *components of uncertainty* that relate small changes in input values to corresponding changes in y .

$$u_i(y) = \left. \frac{\partial Y}{\partial X_i} \right|_{X_i=x_i} u(x_i) \quad \text{and} \quad u_j(y) = \left. \frac{\partial Y}{\partial X_j} \right|_{X_j=x_j} u(x_j). \tag{4}$$

The terms $u(x_i)$ and $u(x_j)$ are the standard uncertainties in the input values x_i and x_j , respectively. The correlation coefficient attributed to a pair of input estimates is $r(x_i, x_j)$, where $r(x_i, x_i) = 1$ when $i = j$.

Standard uncertainties are associated with a number called the *degrees of freedom*, usually denoted ν . The interpretation given to x , $u(x)$, and ν in the GUM is analogous to familiar sample statistics: the sample mean, the standard error in the sample mean, and the degrees of freedom [8] [Chapter 9]. However, degrees of freedom are interpreted more broadly in the GUM, because uncertainty evaluation is not always based on a sample of data. When a standard uncertainty is considered to be known very accurately, the degrees of freedom is large (up to infinity), but a small number of degrees of freedom (as low as unity) signifies a very rough estimate of the underlying standard deviation. Again, the GUM provides an equation for propagating degrees of freedom, called the Welch-Satterthwaite formula. The number of degrees of freedom associated with a standard uncertainty $u(y)$ is as follows.

$$\nu_y = \frac{u^4(y)}{\sum_{i=1}^l \frac{u_i^4(y)}{\nu_i}}. \tag{5}$$

However, there is an important restriction on the use of this equation. It is not valid when input estimates that have finite degrees of freedom are correlated with each other. This is not a rare occurrence. For example, estimates obtained from linear least-squares regressions are often correlated and have finite degrees of freedom. Fortunately, an extended form of (5) can be used in some important special cases (see Appendix A.1) [9].

Equations (1)–(5) describe a methodology for evaluating measurement uncertainty that any GUM-compliant data processing should adhere to. However, in many situations, it is inconvenient, if not impossible, to formulate a single, complete, measurement model such as Equation (1). Usually, a traceable measurement is perceived as a staged process, and it is difficult to describe when approached at more than one stage at a time. However, there is a mathematically equivalent formulation of these calculations that allows staged models to be handled. This formulation leads to a new abstract data type called an *uncertain number*, which can represent inexactly known quantities [4]. The uncertain-number format satisfies the requirements for information exchange identified in the GUM [3] [Section 0.4].

2.2. The Uncertain-Number Methodology

A mathematical expression may often be decomposed into stages and evaluated algorithmically as a sequence of basic operations. For instance, the following is the case:

$$V = v(1 - E_{\text{rel}}) - E_{\text{off}} - E_{\text{rnd}}$$

and can be broken into four stages (also shown in Figure 1):

$$\begin{aligned} y_1 &= 1 - E_{\text{rel}} \\ y_2 &= v \times y_1 \\ y_3 &= y_2 - E_{\text{off}} \\ V &= y_4 = y_3 - E_{\text{rnd}} \end{aligned}$$

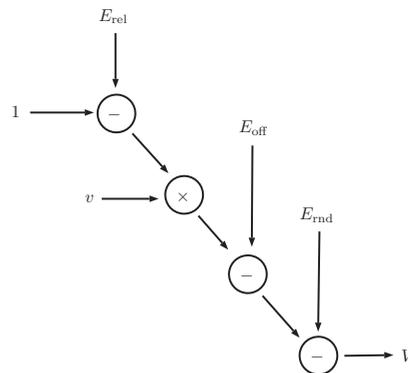


Figure 1. Decomposition of $V = v(1 - E_{\text{rel}}) - E_{\text{off}} - E_{\text{rnd}}$ into a sequence of arithmetic operations.

This approach can be applied to measurement models. The evaluation of some arbitrary function $f(x_1, \dots, x_l)$ can be decomposed into $h = 1, \dots, m$ stages, each producing an intermediate result:

$$y_h = f_h(\Lambda_h), \tag{6}$$

with the final stage yielding the result $y = y_m$. The set of inputs to the h th stage function, denoted here as Λ_h , may include previous stage results y_1, \dots, y_{h-1} and model inputs x_1, \dots, x_l . Using the chain rule for partial differentiation, the components of uncertainty

defined in Equation (4) can be evaluated at each stage. The component of uncertainty in y_h due to uncertainty in the j th model input is as follows.

$$u_j(y_h) = \sum_{z_k \in \Lambda_h} \frac{\partial f_h}{\partial z_k} u_j(z_k). \quad (7)$$

Thus, the set of components of uncertainty $\{u_1(y_h), u_2(y_h), \dots, u_l(y_h)\}$, corresponding to $\{x_1, x_2, \dots, x_l\}$, can be evaluated stage-by-stage to finally obtain the set of components of uncertainty in the result, $\{u_1(y), u_2(y), \dots, u_l(y)\}$ (when $j = k$, the notation $u_j(z_j)$ may be simplified to $u(z_j)$, which is the standard uncertainty of model input x_j).

In GTC, an uncertain number is used to encapsulate results at each stage (y_h), and the associated components of uncertainty, $\{u_1(y_h), u_2(y_h), \dots\}$. Uncertain numbers provide a convenient and succinct representation for quantities. Their algebraic properties essentially match those of ordinary number types. Thus, data processing algorithms can be expressed with familiar mathematical operations applied to uncertain-number terms representing quantities in a model. There is no need to derive the expressions for components of uncertainty; this is handled algorithmically.

The results of uncertain-number calculations are also transferable: the result of one calculation may be used as an argument in further calculations (as is performed routinely in numerical calculation). This is an open-ended process that can, in principle, continue indefinitely. The transferability of results is needed to support metrological traceability in staged measurement models. This will be discussed further in the next section and in Section 3.6. The open-ended nature of uncertain-number computations is also illustrated in the example shown in Section 2.4.

2.3. Traceability Chains and Uncertainty

Traceability provides accurate and reliable information about physical quantities that can be used to inform decisions. Because a quantity of interest can never be determined exactly, a decision based on the information available may not be correct; there will be some uncertainty—in a colloquial sense—about the correctness of a decision informed by data subject to measurement error. However, the risks associated with poor decision outcomes can be managed if the unpredictability of measurement results can be described in probabilistic terms (i.e., if the accuracy can be quantified). In this sense, the metrologist's use of the term measurement uncertainty is associated with a likely magnitude of measurement error. To address the need for results that can be relied upon, metrological traceability requires the careful evaluation of measurement uncertainty.

Traceable measurement can be thought of as a collaborative process that is carried out in stages. Ultimately, a traceable measurement is of benefit to a nominal 'end user' at the last stage of a traceability chain, who needs information about a quantity to inform a decision (e.g., measuring the weights of shipping containers to inform the loading distribution of a container ship). The accuracy of a final result depends on all the stages; thus, the sources of uncertainty must be traced as far back as the units of measurement realised at the beginning of the process. This ensures that the result is meaningful and comparable with other traceable measurements of the same quantity.

While the GUM's expression of a measurement model takes the form of a single all-encompassing Equation (1), the staged formulation in Section 2.2 handles the fact that people involved at one stage generally do not have detailed knowledge about processes carried out at other stages. For example, Figure 2 shows a traceable measurement in four parts (e.g., stages 1 and 2 correspond to the realisation of reference standards, stage 3 to the calibration of a measuring instrument using those standards, and stage 4 to an end-user measurement using the calibrated instrument). The staged model is described as follows;

$$\begin{aligned}
 Y_1 &= f_1(\dots) \\
 Y_2 &= f_2(\dots) \\
 Y_3 &= f_3(Y_1, Y_2, \dots) \\
 Y_4 &= f_4(Y_3, \dots)
 \end{aligned}$$

where unspecified arguments ‘...’ represent some subset of the influence quantities X_1, X_2, \dots, X_l . The end user can probably only formulate a model for stage 4, $f_4(Y_3, \dots)$; thus, information about earlier stages must be summarised and reported down the chain in a suitable format. If the model was expressed as a single function, the composition of the stages would provide the following.

$$f(X_1, X_2, \dots, X_l) = f_4(f_3(f_1(\dots), f_2(\dots), \dots), \dots).$$

Now, the outcome of data processing should not be affected by the expression of the model as a single function or a series of functions. This has a bearing on how information should be communicated along a traceability chain [10]. By reporting uncertain numbers between stages, final results can be obtained that are the same as would be found for a single model. Uncertain numbers realise the GUM’s ideal method for evaluating and expressing the uncertainty of a result [3] [Section 0.4].

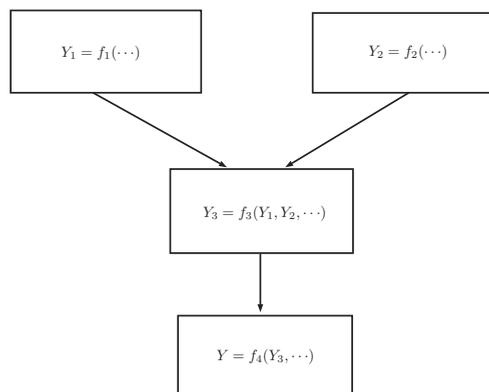


Figure 2. A measurement carried out in four stages. Arrows indicate the transfer of information about intermediate results. The unspecified function arguments ‘...’ represent external quantities that influence the procedures. This figure does not represent a particular measurement, but the four stages may be regarded as follows: realisation of unit reference standards (stages 1 and 2), calibration of an instrument using the standards (stage 3), and a measurement made with the calibrated instrument (stage 4).

2.4. A Simple Example

This section presents an example of uncertain-number data processing applied to a simple electrical network. Figure 3 shows an electrical network with three resistors in series. A voltmeter can be connected between the lower terminal and any of the three terminals above, allowing the potential difference between terminals 0 and 1, 0 and 2, or 0 and 3 to be measured (V_{10} , V_{20} , or V_{30} , respectively).

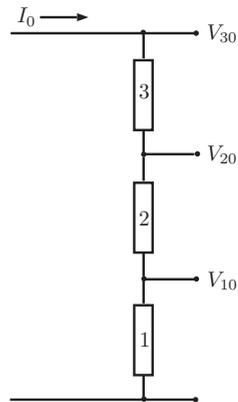


Figure 3. Three resistors in an electrical network. A voltage measurement can be performed between the bottom terminal and any terminal above.

We adopt a simple model for an imperfect voltmeter. The model has three sources of error (influence quantities) that affect the response of a meter (reading) to an input voltage V . A random error (noise), represented as E_{rnd} , affects every reading; a systematic error, E_{off} , contributes a fixed offset to every reading; a systematic relative error, E_{rel} , contributes an error proportional to the reading itself (representing imperfect scaling or non-linearity of the instrument). The relationship between the input voltage, V , and a voltmeter reading, v , is expressed by the model (already shown in Section 2.2).

$$V = v(1 - E_{\text{rel}}) - E_{\text{off}} - E_{\text{rnd}}. \quad (8)$$

The influence quantities E_{rel} , E_{off} , and E_{rnd} are unknown, and so their effects cannot be corrected. However, the displayed value v is used as an approximation for V , because we assume that the instrument is properly adjusted. This is the same as assuming that the residual errors are small enough to be considered approximately zero. The uncertainty in the value of v , due to the estimates $e_{\text{rel}} = e_{\text{off}} = e_{\text{rnd}} = 0$, can be found if the uncertainties $u(e_{\text{rel}})$, $u(e_{\text{off}})$, and $u(e_{\text{rnd}})$ are known.

For uncertain number objects representing inputs to a measurement model, we find it helpful to adopt the term *elementary uncertain number*. Elementary uncertain numbers represent influence quantities. Numeric data must be provided when defining elementary uncertain numbers during the problem initialisation phase; this includes the following: a value (the estimate), a standard uncertainty, and a number of degrees of freedom.

We can use GTC to evaluate properties of the circuit, given measured values and some information about the voltmeter's characteristics. Objects of the class `Voltmeter`, shown below, are used for data processing. During initialisation of a new `Voltmeter` (execution of `__init__()`), elementary uncertain numbers representing the two systematic errors are created and stored as instance variables (`ureal()` creates the uncertain numbers).

```
from GTC import ureal, rp, result

class Voltmeter(object):
def __init__(self,
# Default characteristics for 1 V scale
u_off=5E-3,
u_rel=8E-4,
u_rnd=1E-4
):
self.u_rnd = u_rnd
self.E_off = ureal(0.0,u_off,label="E_off")
```

```

self.E_rel = ureal(0.0,u_rel,label="E_rel")

def applied_voltage(self,v,index):
E_rnd = ureal(0,self.u_rnd,label="E_rnd_{}".format(index))
V = v*(1 - self.E_rel) - self.E_off - E_rnd
return V

```

The `applied_voltage()` method implements the model Equation (8). It returns an uncertain number for the applied voltage corresponding to a displayed value, v (the second argument, `index`, is used to create a label for the elementary uncertain number, E_rnd , which is associated with random noise). In the code below, the uncertain numbers V_{10} , V_{20} and V_{30} are obtained from measurements of V_{10} , V_{20} , and V_{30} .

```

# Displayed values, in volt
v_10 = 0.125841
v_20 = 0.385569
v_30 = 0.981950

# Voltmeter instance, using default specifications
dvm = Voltmeter()

# Uncertain numbers from displayed values
V_10 = dvm.applied_voltage(v_10,1)
V_20 = dvm.applied_voltage(v_20,2)
V_30 = dvm.applied_voltage(v_30,3)

```

We can infer circuit properties from these uncertain-number results. For example, the code below shows how accurately the voltage V_{10} was measured and the most important contributions to the uncertainty in that measured value.

```

# Utility function to display a result
def display(v,label):
print( "{}: {}".format(label,v) )

for l,u in rp.budget(v,trim=0):
print( "{}: {}".format(l,u) )

display(V_10,"V_10")

```

The function `display()` prints the measured value and a standard uncertainty in parentheses, followed by a list of components of uncertainty in order of magnitude. The label for each component is shown on the left and the magnitude of the component of uncertainty on the right. The two most significant figures of standard uncertainty are shown. Here, V_{10} has a measured value of 0.1258 volts and a standard uncertainty of 0.0050 volts.

```

V_10: 0.1258(50)
E_off: 0.005
E_rel: 0.000100672800000000002
E_rnd_1: 0.0001

```

Note that the dominant component of uncertainty is associated with the approximation made for systematic offset error E_{off} . Very similar results are obtained for V_{20} and V_{30} .

Other circuit properties can be calculated too. For example, the potential difference across resistor 2 can be found by taking the difference between V_{20} and V_{10} . Subtracting those uncertain numbers in the argument, `display(V_20-V_10, "V_20-V_10")` yields the following.

```

V_20-V_10: 0.25973(25)
E_rel: 0.000207782400000000003
E_rnd_1: 0.0001
E_rnd_2: 0.0001
E_off: 0.0

```

This is an interesting result, which illustrates the detailed underlying calculation of uncertainty that is performed automatically. The standard uncertainty in the difference here is only 0.00025 V—significantly less than the standard uncertainty in the individual measurements (both 0.0050 V). The uncertainty in this voltage difference is lower because it is insensitive to the offset E_{off} (the offset is exactly the same in both readings). The display of components of uncertainty shows that the sensitivity to E_{off} has been reduced to zero and that the influence of E_{rel} is now dominant. We might also expect a smaller contribution to uncertainty to come from relative systematic error E_{rel} . However, that component varies in proportion to the applied voltage (it is a systematic relative error), and since v_{20} is about three times larger than v_{10} , the contribution to uncertainty from E_{rel} is still about two times larger than it was in the direct measurement of V_{10} .

3. Aspects of GTC Design

Using the mathematics described in the previous section applied to a given measurement model, GTC is required to evaluate a measured value, a standard uncertainty, and a number of degrees of freedom. This data processing can involve many computational stages and hundreds of influence factors. In addition, GTC can report the components of uncertainty in a result due to the uncertainty of individual influence (input) quantities and the components of uncertainty due to uncertainty in particular intermediate results, as required. Furthermore, it can store and retrieve uncertain numbers, allowing stages along a traceability chain to be appropriately handled.

This section describes how GTC has been designed to meet these challenges. The GTC package was first released four years ago, but our experience with the uncertain-number approach reaches back more than twenty years. We have used different programming languages and changed our thinking about how to implement the technique. For instance, early versions encountered difficulties when the size and variety of the measurement problems grew, and when additional software features were requested. Some programming languages were found to be better suited than others; larger problems exposed scaling weaknesses in our designs; and additional features place strain on some of the data structures and algorithms. GTC implements what we now consider to be our ‘best’ approach.

3.1. Simultaneous Calculation of Value and Uncertainty

Section 2.2 explained that the calculation of components of uncertainty can be handled using the chain rule for partial differentiation when a measurement model is expressed in stages. GTC extends this further by decomposing stage model expressions into very basic operations, such as \times , \div , $\sin()$, $\exp()$, etc. This is effectively using a computational technique called automatic differentiation [11]. Arithmetic operator overloading and a library of mathematical functions for uncertain numbers are used to automate decomposition of mathematical expressions into simple steps and then to evaluate the value and components of uncertainty at each step.

All the basic uncertain-number functions and arithmetic operations defined in GTC are either univariate or bivariate. For a univariate function, $f_h(z)$, Equation (7) reduces to the following:

$$u_j(y_h) = \frac{\partial f_h}{\partial z} u_j(z), \tag{9}$$

and for a bivariate function, $f_h(z_1, z_2)$, Equation (7) becomes the following.

$$u_j(y_h) = \frac{\partial f_h}{\partial z_1} u_j(z_1) + \frac{\partial f_h}{\partial z_2} u_j(z_2). \tag{10}$$

Thus, for example, the uncertain-number trigonometric sine function can be handled as follows. If the value of an uncertain-number input is z , then the value of the uncertain-number result is $y_h = \sin(z)$. Furthermore, if there are two components of uncertainty associated with the input, $u_1(z)$ and $u_2(z)$, then the two corresponding components of

uncertainty associated with the result are $u_1(y_h) = c u_1(z)$ and $u_2(y_h) = c u_2(z)$, where $c = \partial f_h / \partial z = \cos(z)$ is the derivative in (9).

3.2. Unique Identifiers

GTC algorithms track the identity of elementary uncertain numbers representing influence factors. The subscript i that appears on the terms for components of uncertainty, $u_i(y)$, is the same as the subscript appearing on the influence quantities, X_i , in model Equation (1).

$$Y = f(X_1, X_2, \dots).$$

Software and digital records must somehow keep track of these i 's, even when measurements are carried out in stages, at different locations and at different times. GTC uses a simple tuple of integers as an identifier format. The first integer is kept fixed for a given session while the second integer takes the value of a counter that is incremented each time an elementary uncertain number is created. To ensure that these identifiers are unique in time and space, the first integer is a Universally Unique Identifier (UUID) formatted as a 128-bit integer.

This identifier format reveals nothing about the influence quantity, although identifiers can be arranged in order, which improves the performance of some algorithms. Would a more sophisticated type of identifier be useful? For the purposes of data processing, the only requirement is uniqueness. Nevertheless, GTC already allows text labels to be associated with nodes (used as labels for influence quantities in uncertainty budgets) and a planned enhancement to GTC will allow information about influence quantities to be held in a manifest and indexed by unique identifiers. Such a manifest could accompany a digital record of uncertain-number data. This could address any need for additional metadata about influence quantities, without the burden of minting and configuring more specialised digital objects designed to access information on the internet [12].

3.3. Node Classes

During a GTC calculation, most of the uncertain numbers that are created may be regarded as temporary objects and can be garbage-collected almost immediately. If this is not performed in large problems, the demands on memory can seriously limit performance. To address this, GTC only holds essential information about influence quantities, and information about certain intermediate results as required. This essential information is kept in small node objects, allowing the memory occupied by larger uncertain-number objects to be reclaimed when not required.

There are two classes of node: A `Leaf` is associated with elementary uncertain numbers and a `Node` is associated with uncertain numbers representing intermediate results; `Leaf` is a subclass of `Node` (Figure 4). A `Leaf` is created whenever an elementary uncertain number is declared. The information encapsulated includes the following: a value of standard uncertainty; a number of degrees of freedom; a Boolean flag, which identifies objects declared to be independent; a string label, for display purposes; a unique identifier; and two Python collection objects. One of these is the dictionary `correlation`, which holds values of $r(x_i, x_j)$ for calculations such as Equation (11). The other is the `ensemble` set, which is used to identify other nodes associated with an ensemble of closely related elementary uncertain numbers. Ensembles are used in the calculation of degrees of freedom described in Section 2.1 (see also Appendix A.1).

While a `Leaf` is created for every elementary uncertain number, there is no need to create nodes at every stage of a calculation. When intermediate components of uncertainty are required (for an intermediate result of particular significance or when an uncertain number will be stored), the function `result()` is used to create a new `Node` (as shown below in Section 3.5 and later in Appendix A.3).

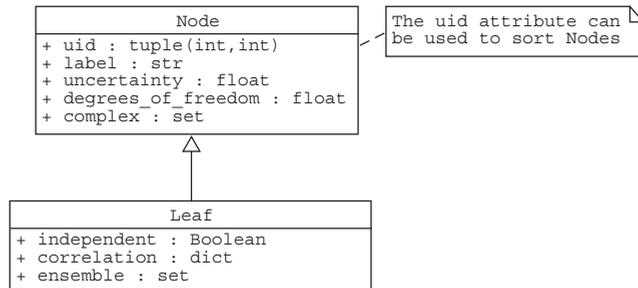


Figure 4. A UML diagram for node classes. There are two types of node: Node and Leaf. Leaf nodes are associated with elementary uncertain real numbers, while objects of the parent class Node may be associated with intermediate uncertain-number results. The uid attribute can be used to sort nodes. The first integer in uid is a common UUID value for all the nodes created in a Python session, while the second integer enumerates the nodes created in that session.

3.4. Propagating Uncertainty

During uncertainty propagation, components of uncertainty must be evaluated at each step. The cumulative effect of these computations can dominate execution time and the demands on memory can be high. Moreover, the overhead of looking up correlation coefficient values for pairs of inputs in Equation (3) during the final calculation of a standard uncertainty is inefficient.

To address this, information about components of uncertainty is stored in several sequences in uncertain-number objects (u_components, d_components and i_components in Figure 5). The elements in these sequences consist of a component of uncertainty paired with a node that holds information about the corresponding influence quantity (see Figure 4).

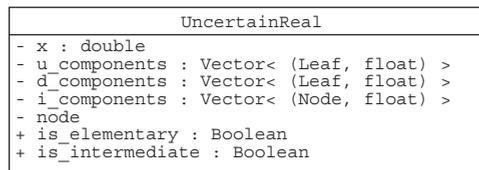


Figure 5. A UML class diagram for an uncertain real number. Vector objects contain sequences of pairs (the Node and Leaf classes in Figure 4). The Vectors u_components and d_components hold components of uncertainty for independent and dependent influences, respectively. The Vector i_components holds components of uncertainty with respect to designated intermediate results. The private node attribute will refer to a Leaf when an UncertainReal object is elementary, or to a Node when the object is an intermediate result, but otherwise, the attribute is not assigned.

During calculations, an uncertain number is created at every step. The components of uncertainty are evaluated by weighting the components of uncertainty for the inputs to the step and combining these weighted components when common influences are involved. By keeping the elements in sequences ordered, this process can be handled efficiently by stepping along the sequences and identifying any common influences. The ordering is established by the uid attribute of node objects.

For example, consider the multiplication of a pair of uncertain numbers, $f_h(Z_1, Z_2) = Z_1 Z_2$. Suppose the component-of-uncertainty sequence for the first argument contains the following elements, where only the first four digits of the UUID integer are shown:

- (7953..., 1) : 3,
- (7953..., 2) : -1,
- (7953..., 4) : 15,
- (7953..., 6) : -5

and the component-of-uncertainty elements for the second argument are as follows.

- (7953..., 1) : -1,
- (7953..., 3) : 2,
- (7953..., 4) : 2,
- (7953..., 5) : 12

To obtain the component-of-uncertainty sequence for the product, components in the first sequence are weighted by the value of the second argument and components in the second sequence are weighted by the value of the first argument. The weighted components of common influences are added together (in this case, there is a pair of common influences identified by (7953 . . . , 1) and (7953 . . . , 4)). Thus, if the value of the first argument is 5 and the value of the second is 10, the components of uncertainty for the product are as follows.

- (7953..., 1) : 25,
- (7953..., 2) : -10,
- (7953..., 3) : 10,
- (7953..., 4) : 160,
- (7953..., 5) : 60,
- (7953..., 6) : -50

When evaluating Equation (3), the number of terms to be summed grows in proportion to the square of the number of input arguments. Moreover, a value of $r(x_i, x_j)$ is needed for every pair of inputs. However, in practice, there are very few non-trivial correlation coefficients assigned (when $i = j$, $r(x_i, x_j) = 1$ and usually $r(x_i, x_j) = 0$, when $i \neq j$). Therefore, not only is the overhead of looking up correlation coefficients unnecessary but many terms in the double sum are zero. To streamline this calculation, the data for independent and dependent influences are separated in two different component-of-uncertainty sequences (`u_components` and `d_components`, in Figure 5). When elementary uncertain numbers are defined, it will be known whether correlation coefficients will be associated with the inputs; thus, dependent and independent influences can be identified and separated. The evaluation of Equation (3) can then be handled more efficiently. For instance, if there are independent estimates X_1, \dots, X_K and dependent estimates X_{K+1}, \dots, X_L , the calculation of Equation (3) can be expressed as follows:

$$u(y) = \left[\sum_{i=1}^K u_i^2(y) + 2 \sum_{i=K+1}^L \sum_{j=i+1}^L u_i(y) r(x_i, x_j) u_j(y) \right]^{1/2}, \tag{11}$$

where the double sum is now only over dependent terms.

3.5. Intermediate Results

In some problems, the interpretation of results is complicated by a large number of influences and, hence, a large number of components of uncertainty. A succinct and

often more intuitive presentation can sometimes be obtained, without sacrificing rigour, by reporting the sensitivity of a final result to uncertainty in intermediate results. This is implemented in GTC by using another sequence for components of uncertainty with respect to designated intermediate results (`i_components` in Figure 5).

To initialise the process of calculating an intermediate component of uncertainty, function `result()` must be applied to an uncertain number. This seeds a new element in `i_components`. Thereafter, propagation occurs, as before, by weighting the intermediate components of stage inputs by the partial derivatives of the stage function. The elements in the sequence `i_components` are also node-value pairs, but a `Node` rather than a `Leaf` is used (see Figure 4).

The electrical network example can be used to illustrate the use of intermediate results. Suppose the current through the series network is measured as 1.0000 mA, with a standard uncertainty of 0.0010 mA. The resistance of the resistor in the middle of the network, and a breakdown of the contributions to uncertainty in that value, is obtained simply from the following.

```
I = ureal(1E-3,0.001E-3,label="I")
display( (V_20-V_10)/I, "R2" )
```

The results are as follows.

```
R2: 259.73(36)
I: 0.2597279999999999
E_rel: 0.20778240000000003
E_rnd_1: 0.1
E_rnd_2: 0.1
E_off: 0.0
```

However, to see how resistance depends on voltage and current measurements, we declare the voltage difference to be an intermediate result and report an uncertainty budget in terms of only the current and voltage.

```
V_20-V_10 = result(V_20-V_10,label="V_20-V_10")
for l,u in rp.budget(V_20-V_10/I,influences=[I,V_20-V_10]):
print( "{l}: {u}".format(l,u) )
```

The results are as follows:

```
I: 0.2597279999999999
V_20-V_10: 0.2513434418276316
```

which shows that the contribution to uncertainty from the current measurement is comparable to the contribution from the voltage measured across the resistor. This was not so obvious from the complete list of uncertainty components obtained earlier.

3.6. Storage and Retrieval of Uncertain Numbers

Section 2.3 explained that traceable measurements can be preformed in stages and that, by reporting uncertain numbers when data are processed at each stage, the final uncertainty can be evaluated correctly. The results at one stage will be required at a later time and place. Thus, uncertain numbers must be stored somehow and their identities, which correspond to physical quantities in the actual measurement, must be retained. In GTC, an `Archive` object is used to manage storage and retrieval of uncertain numbers. The unique identifiers described in Section 3.2 keep track of uncertain-number identities in different Python sessions.

As an example, the code below saves an uncertain number for the voltage difference $V_{20} - V_{10}$ in a text file using a JSON format. Note that the GTC function `result()` is applied to designate $V_{20} - V_{10}$ as an intermediate result. This is a prerequisite for storage.

```
from GTC import pr
a = pr.Archive()
```

```
a.add( V_20_V_10 = result(V_20 - V_10) )
with open("file_name.json", "w") as f:
pr.dump_json(f,a,indent=4)
```

The uncertain number $V_{20_V_10}$ for the voltage difference is retrieved by the following:

```
with open("file_name.json", "r") as f:
a = pr.load_json(f)
display( a["V_20_V_10"] , "V_20_V_10")
```

which produces the following output.

```
V_20_V_10: 0.25973(25)
E_rel: 0.00020778240000000003
E_rnd_1: 0.0001
E_rnd_2: 0.0001
E_off: 0.0
```

This shows that the necessary information about influence quantities has been retained. Alternatively, the measurements of V_{10} and V_{20} can be saved individually.

```
a = pr.Archive()
a.add( V_10 = result(V_10) )
a.add( V_20 = result(V_20) )
with open("file_name.json", "w") as f:
pr.dump_json(f,a,indent=4)
```

Then, in a later session, the difference between the individual results can be evaluated by the following:

```
with open("file_name.json", "r") as f:
a = pr.load_json(f)
V_10 = a["V_10"]
V_20 = a["V_20"]
display(V_20 - V_10, "V_20_V_10")
```

which provides exactly the same results.

```
V_20_V_10: 0.25973(25)
E_rel: 0.00020778240000000003
E_rnd_1: 0.0001
E_rnd_2: 0.0001
E_off: 0.0
```

Although the code here suggests that uncertain-number objects are simply being saved and restored, there is more to it: information about related elementary and intermediate uncertain numbers is also included in the digital record. When an uncertain number is 'added' to an archive, objects that hold information about related influences are identified from the component-of-uncertainty sequences. For instance, we showed earlier that the calculation associated with V_{10} could be decomposed into stages (Figure 1). Referring again to that figure and thinking about data processing, the error terms, E_{rel} , E_{off} , and E_{rnd} , correspond to elementary uncertain numbers, and each of the circled mathematical operations represents an intermediate stage in the calculation. If the intention is to store the uncertain-number V_{10} , then information is also required about the voltmeter errors. Figure 6 shows objects with information that would be saved. Later, when the contents of an Archive is loaded back into a different session, this contextual information is immediately restored (nodes are created with the appropriate identifiers). This ensures that when uncertain numbers are retrieved from an archive, they behave as they would have in the context of the original session, which maintains the integrity of information in a traceability chain.

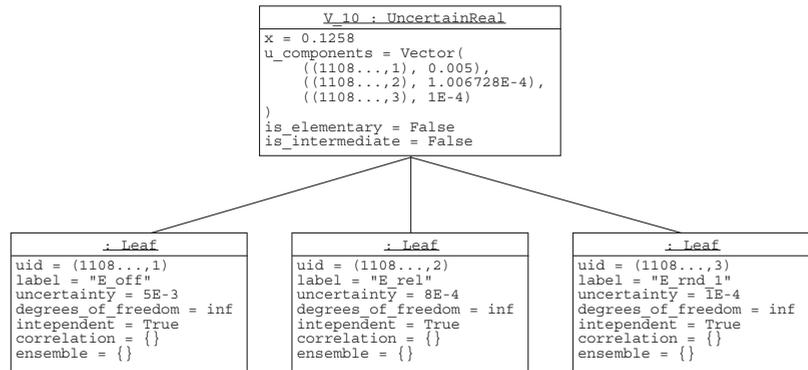


Figure 6. A UML object diagram showing information that would be collected when the uncertain number V₁₀ is stored. The three Leaf objects contain information about the error terms in the voltmeter measurement model: the fixed offset, the relative error and the random noise. Only one Leaf for the influence of noise is shown; however, if several readings had been taken there would be a different node for each reading.

4. Discussion

This paper provides some insight into the usefulness of uncertain numbers, which have the distinctive feature of providing an abstract representation for measured quantities, allowing uncertainty calculations to be automated. Recently, uncertain numbers have been applied to a goniometric measurement system for optical reflectance. The four-axis goniometric system has many configuration errors that must be considered in the measurement model to account for final measurement uncertainties [7]. The application of GTC was carefully compared with alternative computational methods, using Monte Carlo and direct mathematical analysis. GTC was found to be the preferred choice. Using the information provided by uncertain numbers, the authors were able to obtain a better understanding of the measurement system and the inherent correlations between significant measurement errors. This enabled them to significantly improve the accuracy of certain measurements.

The inherent support for metrological traceability is perhaps the most important quality of uncertain numbers. This aspect is implemented in data structures and storage formats used by GTC, which is a particular choice but other formats would be possible. One can easily imagine a more heterogeneous situation, where processing at various stages would be carried out using different software tools. To support this, the format for exchange of data between stages would need to be standardised. That is, there would need to be agreed formats for representing uncertain numbers, which would be used in digital reporting documents such as calibration reports [13].

Digitalisation should offer benefits that are not currently available. The GUM recommends that detailed information about influence quantities be reported at each stage. However, this rarely happens, because calibration certificates and other measurement reports are intended to be read by people; thus, handling the additional data would be difficult. As a consequence, information about common influence factors is rarely shared. A simple situation where this might arise is the scenario of a batch of sensors that are calibrated using a more accurate reference device. If the common reference is ignored, the accuracy of results obtainable from a survey of the sensors' readings is compromised [14]. However, uncertain-number calibration factors can track common effects and account for them when comparing readings from different sensors. This was illustrated in Section 2.4, where E_{off} contributed a common offset to single voltage readings but nothing to uncertainty in the voltage difference. It is also worth noting that a 'smart' sensor capable of reporting

uncertain-number results would not need to process a lot of information. As was the case of the simple voltmeter, a model of the sensor measurement might only require a few influence factors and the calculations would be simple.

The various stages of a traceable measurement often occur in different locations (national metrology institute, second-tier calibration laboratory, etc.) but they may also happen at different times in the same location. For example, a working standard might be calibrated in-house against an externally calibrated transfer standard. The working standard would then be used repeatedly to calibrate different instruments at different times. Importantly, measurement errors realised when at the time the working standard is calibrated should be treated as systematic effects in subsequent instrument calibrations. Performing this would allow any bias, or correlation, in downstream measurement results using those instruments to be accounted for correctly. This could be easily handled by digitalisation if uncertain-number storage and retrieval mechanisms are used to save calibration data for the working standard and later retrieve it for reuse when instruments are calibrated.

During formal international measurement comparisons, national metrology institutes (NMIs) go to much greater pains when reporting measurement data than they do for regular calibration work. These international comparisons assess the competence of NMIs in performing specific types of measurement. The more detailed reporting requirements in comparisons align with the GUM's recommendations in this case. A recent study, which explored a future scenario where an uncertain-number reporting format was used by all participants, showed that using uncertain numbers would not only provide the information required, but they would also simplify comparison analysis and comparison linking and provide additional insights into the results [6].

Measurement models are needed in order to use uncertain numbers effectively. The close correspondence between quantity terms in a model and uncertain numbers in data processing routines makes software development and testing more robust and reliable and avoids the need to explicitly derive expressions for the components of uncertainty from a model, which GTC handles automatically. However, although modelling lies at the heart of the GUM's approach, skilled metrologists are often confident in their ability to assess measurement uncertainty heuristically and frequently elide the formal modelling step. This presents a problem for digitalisation, because digital systems need a rigorous formal problem definition for autonomous operation. Some tutorial guidance on developing measurement models has been provided in a recent booklet [15] and is also the subject of another paper [16]. There is also a new supplement to the GUM, which deals with modelling [17].

One common conceptual difficulty when modelling is the omission of influence quantities estimated as zero. These terms would not be needed in conventional data processing; however, they must be modelled, because the actual (unknown) values affect the final measurement result, and so they contribute to uncertainty. Influence quantities with trivial estimates are often called residual errors. The electrical network example, in Section 2.4, included three residual errors that were all estimated as zero. These terms were represented by uncertain numbers and modelled imperfect voltmeter behaviour.

5. Conclusions

GTC is a software tool for data processing with automatic evaluation of measurement uncertainty. It follows international best-practice, described in the GUM, and offers useful extensions to those methods for important special cases. The use of uncertain numbers is a distinctive feature of GTC. The uncertain-number data-type facilitates data processing, which can be performed in a piece-wise and open-ended manner. This allows calculations to be more easily matched to the models of a measurement performed in stages. The automation of uncertainty calculations allows measurement data processing to be made more rigorous, which can lead to accuracy enhancement in some cases. The uncertain-number format significantly exceeds current paper-based practices that support traceability.

Therefore, GTC and the data structures used to implement uncertain numbers are a useful example of software that meets the requirements of a fully functional digital infrastructure for metrological traceability.

Funding: This work was funded by the New Zealand Government.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The author is grateful to Peter Saunders and Joseph Borbely for careful reviews of the manuscript.

Conflicts of Interest: The author declares no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

BIPM	International Bureau of Weights and Measures;
GUM	Guide to the Expression of uncertainty in measurement;
IEC	International Electrotechnical Commission;
IFCC	International Federation of Clinical Chemistry and Laboratory Medicine;
ILAC	International Laboratory Accreditation Cooperation;
ISO	International Organisation for Standardisation;
IUPAC	International Union of Pure and Applied Chemistry;
IUPAP	International Union of Pure and Applied Physics;
JSON	JavaScript Object Notation;
NIST	National Institute of Standards and Technology;
NMI	National Metrology Institute;
OIML	International Organization of Legal Metrology;
SI	International System of Units (<i>Système International</i>);
UUID	Universally Unique Identifier.

Appendix A. Additional Details

This Appendix includes details about a number of other aspects of GTC. Support for problems involving finite degrees of freedom with correlated inputs is discussed in Appendix A.1 and support for complex quantities is briefly covered in Appendix A.2. Appendix A.3 presents an unusual case, where counter-intuitive results are obtained due to the relationship between Python variables and underlying uncertain-number objects. This provides further insight into computational mechanisms. Appendix A.4 briefly describes the validation of GTC and Appendix A.5 compares GTC with some similar software projects.

Appendix A.1. Ensembles

As noted in Section 2.1, the Welch-Satterthwaite formula cannot be used on correlated data with finite degrees of freedom. However, there is an extension that can be applied in situations where data are deemed to come from closely related quantities with fixed interdependencies [9]. To implement this, GTC algorithms must be able to identify sets of uncertain numbers declared as representing related quantities. The `ensemble` attribute of the `Leaf` node is used for this purpose (Figure 4). An ensemble is a set of `Leaf` nodes. There are some GTC functions that declare ensembles automatically, such as functions for linear regression; in other cases an ensemble can be explicitly defined by `multiple_ureal()` (the GTC online documentation for `multiple_ureal()` shows a calculation from GUM Appendix H2 [18]).

GTC includes regression functions that estimate the parameters of a straight line passing close to a sample of data. The finite sample size means that uncertainties in estimates for the slope and intercept have finite degrees of freedom and are usually correlated.

The code below shows a least-squares regression for nine data points. The GTC function `line_fit()` returns an object with an attribute that holds a pair of uncertain numbers for the slope and intercept (`a_b`).

```
from GTC import type_a, get_correlation

x = [1,2,3,4,5,6,7,8,9]
y = [15.6,17.5,36.6,43.8,58.2,61.6,64.2,70.4,98.8]
result = type_a.line_fit(x,y)
a,b = result.a_b
print("a =",repr(a))
print("b =",repr(b))
print("r(a,b) =",get_correlation(a,b))
```

The results are as follows.

```
a = ureal(4.813888888888881,4.886206312183354,7)
b = ureal(9.408333333333335,0.8683016476563609,7)
r(a,b) = -0.888523316639
```

The slope and intercept are correlated and there are seven degrees of freedom associated with the uncertainties. However, these results may still be used to calculate the expected value y for $x = 5.5$:

```
y_p = a + b*5.5
print("y_p =", repr(y_p))
```

which produces a result with seven degrees of freedom.

```
y_p = ureal(56.55972222222223,2.2835948151943155,7.0)
```

Appendix A.2. Complex Quantities

Section 2.2 described data processing for real-valued quantities, but very similar formulae apply to complex quantities. These are also implemented in GTC. A review of measurement uncertainty for complex quantities has been given by Hall [19].

GTC can handle mathematical expressions with a mixture of real-valued and complex-valued quantities and results may be either real or complex uncertain numbers, as is appropriate. An uncertain complex number is implemented as a pair of uncertain real numbers; thus, uncertainty is represented by uncertainties in the real and the imaginary components as well as the correlation coefficient between those components. A convenient format for specifying uncertainty in a complex value is a 2×2 variance-covariance matrix. The number of degrees of freedom associated with uncertainty in the real and imaginary components is the same.

Often a complex quantity is evaluated from a small sample of data. In that case, the real and imaginary component estimates are dependent, being evaluated from the same sample; they will also have a finite number of degrees of freedom. As already mentioned, the combination of finite degrees of freedom and correlation creates problems for data processing. However, when converting from a complex quantity to a real one, the modified form of the Welch–Satterthwaite formula can be useful [9]. For example, suppose a complex number $z = x + iy$ has been evaluated from a small sample, $x = 0.20$ and $y = 0.0$, the real and imaginary components each have a variance of 0.1, there is a covariance between the components of 0.05, and there are 10 degrees of freedom. If the real-valued magnitude $|z|$ is of interest, it can be evaluated as follows:

$$|z| = \sqrt{x^2 + y^2}.$$

is of interest, it can be evaluated as follows.

```

from GTC import ucomplex, magnitude

z = ucomplex(0.20, [0.1, .05, .05, 0.1], 10)
print("mag(z) = {!r}".format( magnitude(z) ))

```

The result

```
ureal(0.2, 0.31622776601683794, 10.0)
```

is an uncertain real number with 10 degrees of freedom.

On the other hand, if the result of a calculation is an uncertain complex number, there is an alternative to the Welch–Satterthwaite formula that must be used [20]. Here is an example (from [20]):

```

from GTC import ucomplex

x1 = ucomplex( 1, (0.96, -0.34, -0.34, 0.27), 5 )
x2 = ucomplex( 1, (0.51, 0.33, 0.33, 0.31), 3 )
x3 = ucomplex( 1, (0.45, 0.28, 0.28, 1.65), 6 )

z = x1 + x2 + x3

print("z = {0.x}".format( z ))
print("u(z) = {0.u}".format( z ))
print("r(x,y) = {0.r}".format( z ))
print("df(z) = {0.df}".format( z ))

```

which displays the complex value, the standard uncertainties, the correlation coefficient, and the degrees of freedom.

```

z = (3+0j)
u(z) = StandardUncertainty(real=1.3856406460551018, imag=1.493318452306808)
r(x,y) = 0.13048503857331625
df(z) = 11.340977790491408

```

Appendix A.3. Uncertain Number Objects and References

The combination of uncertain numbers and Python language features can provide intuitive and meaningful representations of a problem domain. In particular, the distinction between random and systematic effects can be elegantly captured in object-oriented designs. However, on rare occasions, the behaviour of the Python variable names that refer to objects in memory can result in confusion. It is interesting to see how this can happen, because it provides insight into the computational processes. Here is a simple example.

Consider the following equations.

$$\begin{aligned}
 u &= x, \\
 v &= x + y, \\
 w &= u + v.
 \end{aligned}$$

What is $\partial w / \partial u$? To find the answer using GTC, we may perform the following (note, partial derivatives are evaluated when elementary uncertainty numbers are declared with an uncertainty of unity):

```

from GTC import ureal, component, result

x = ureal(0,1,label="x")
y = ureal(0,1,label="y")
u = x
v = x + y
w = u + v
print("partial derivative wrt x =", component(w,x) )
print("partial derivative wrt u =", component(w,u) )

```

which displays the following.

```
partial derivative wrt x = 2
partial derivative wrt u = 2
```

If we had in mind that $w = u + v$, this result may come as a surprise because $\partial w / \partial u = 1$ would be expected. However, it is important to remember that the terms in a calculation correspond to uncertain-number objects in memory and not the variable names in code. Both x and u refer here to the same elementary uncertain number. Therefore, equation $w = u + v$ actually corresponds to $w = 2x + y$ in terms of the underlying objects, and so $\partial w / \partial x = 2$ is correct.

Confusion is created by equating the Python variables $u = x$ if it is (incorrectly) assumed that u and x are somehow different. If we intend to take the derivative of $w = u + v$ with respect to u , a distinct uncertain-number object must be created for u (and designated as an intermediate result to allow an intermediate component of uncertainty to be calculated).

To implement this calculation, we may use the unary “+” operator to create an additional uncertain number representing u in memory. This operator copies the numerical attributes of its argument into a new uncertain number. As far as calculation is concerned, this object corresponds to a distinct term. The following code clones x and designates it as an intermediate result to allow the component of uncertainty to be evaluated.

```
x = ureal(0,1,label="x")
y = ureal(0,1,label="y")
u = result(+x,label="u")
v = x + y
w = u + v
print( "partial derivative wrt x =", component(w,x) )
print( "partial derivative wrt u =", component(w,u) )
```

This displays the following.

```
partial derivative wrt x = 2
partial derivative wrt u = 1
```

This situation is unusual. Normally, `result()` would be applied to an object produced as the result of a calculation; thus, there is almost never a need to clone uncertain numbers as is shown here.

Appendix A.4. Testing and Validation

GTC has a modular structure. It uses Python arithmetic operator overloading and mathematical function definitions to decompose mathematical expressions into basic uncertain-number operations. This makes the code amenable to unit testing. The calculation of values uses standard Python mathematical operations and processing of components of uncertainty uses automatic differentiation, which also makes use of built-in Python arithmetic and mathematical libraries. An extensive suite of test cases has been built up to verify implementation details [5]. Calculations are also checked against standard examples from appendices to the GUM and other published sources, including various forms of regression analyses with uncertainties. Moreover, GTC has been used for more than a decade at the Measurement Standards Laboratory, where it is closely scrutinised by different groups. Very few issues have been reported since the project was made publicly available on github in 2018.

Appendix A.5. Similar Software

Software to evaluate measurement uncertainty is often used alongside other data processing tools (a notable example is the web-based calculator called *The NIST Uncertainty Machine* [21]). However, the separation of data processing into different work streams is unnecessary with GTC, because uncertainty calculation is an integral part of all data

processing. GTC may be incorporated in projects to provide data processing and support for traceability. This is also the case for the C# library called *UncLib* [22], which is part of the data acquisition and data processing application called *VNA Tools II* [23], now used by many leading microwave metrology laboratories. Similarly to GTC, *UncLib* handles measurements of real-valued and complex-valued quantities and provides support for traceability by identifying input quantities and allowing them to be stored and retrieved. However, the data structures of *UncLib* have been designed to support a particular optimisation strategy, which results in some different behaviours (see Zeier et al. [22] [Section 3.3]). For instance, the evaluation of the intermediate components of uncertainty may fail unless certain preconditions are satisfied [22] [Section 3.4] and the extensions to the Welch–Satterthwaite formula and degrees of freedom support for complex quantities cannot be implemented.

A well-known Python package that calculates uncertainty is *Uncertainties* [24]. This package is intended for engineering error and sensitivity analyses, such as described by Bevington and Robinson [25]. Similarly to GTC and *UncLib*, *uncertainties* use automatic differentiation to evaluate partial derivatives during data processing. However, it does not calculate degrees of freedom, nor does it handle complex quantities or provide for storage and retrieval of results.

References

- Hackel, S.; Härtig, F.; Hornig, J.; Wiedenhöfer, T. The Digital Calibration Certificate. *PTB-Mitteilungen* **2017**, *127*, 75–81. [CrossRef]
- Thiel, F. Digital transformation of legal metrology-The European Metrology Cloud. *OIML Bull.* **2018**, *59*, 10–21.
- BIPM; IEC; IFCC; ILAC; 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, France, 2008.
- Hall, B.D. Computing with Uncertain Numbers. *Metrologia* **2006**, *43*, L56–L61. [CrossRef]
- Hall, B.D.; Borbely, J.S. GUM Tree Calculator and v1.3.6. Available online: <https://github.com/MSLNZ/GTC>; doi:10.5281/zenodo.5459479 (accessed on 13 November 2021).
- Hall, B.D.; Koo, A. Digital Representation of Measurement Uncertainty: A Case Study Linking an RMO Key Comparison with a CIPM Key Comparison. *Metrologia* **2021**, *1*, 166–181. [CrossRef]
- Molloy, E.; Saunders, P.; Koo, A. Effects of rotation errors on goniometric measurements. *Metrologia* **2021**, in press. [CrossRef]
- Walpole, R.E.; Myers, R.H.; Myers, S.L.; Ye, K. *Probability & Statistics for Engineers and Scientists*, 8th ed.; Pearson Education: Upper Saddle River, NJ, USA, 2007; p. 07458.
- Willink, R. A generalization of the Welch–Satterthwaite formula for use with correlated uncertainty components. *Metrologia* **2007**, *44*, 340. [CrossRef]
- Hall, B.D.; White, D.R. Digital representation of measurement uncertainty for metrological traceability. In *Advanced Mathematical and Computational Techniques in Metrology XII*; Pavese, F., Forbes, A.B., Chunovkina, A.G., Zhang, N.F., Eds.; *Series on Advances in Mathematics for Applied Sciences*; World Scientific: Singapore, 2021; Volume 90, pp. 262–272.
- Rall, L.B.; Corliss, G.F. An Introduction to Automatic Differentiation. In *Computational Differentiation: Techniques Applications, and Tools*; Berz, M., Bischof, C.H., Griewankpp, A., Eds.; Siam Proceedings in Applied Mathematics Series; SIAM: Philadelphia, PA, USA, 1996; Volume 89, pp. 1–17.
- Klump, J.; Huber, R. 20 Years of Persistent Identifiers—Which Systems are Here to Stay? *Data Sci. J.* **2017**, *16*, 9. [CrossRef]
- Hackel, S.; Härtig, F.; Hornig, J.; Wiedenhöfer, T. Metrologie für die Digitalisierung von Wirtschaft und Gesellschaft. 2017. Available online: <https://oar.ptb.de/files/download/5a9808d94c91840d190b3891> (accessed on 13 November 2021).
- Hall, B.D. *An Opportunity to Enhance the Value of Metrological Traceability in Digital Systems*, 2nd ed.; Workshop on Metrology for Industry 4.0 and IoT (MetroInd4.0 IoT); IEEE: Naples, Italy, 2019; pp. 16–21.
- Hall, B.D.; White, D.R. *An Introduction to Measurement Uncertainty*; Measurement Standards Laboratory of New Zealand: New Zealand, 2020. [CrossRef]
- Hall, B.D. Object-oriented software for evaluating measurement uncertainty. *Meas. Sci. Technol.* **2013**, *24*, 055004. [CrossRef]
- BIPM; IEC; ILAC; IFCC; ISO; IUPAC; IUPAP; OIML. *Guide to the Expression of Uncertainty in Measurement—Part 6: Developing and Using Measurement Models JCGM GUM-6:2020*, 1st ed.; BIPM Joint Committee for Guides in Metrology: Paris, France, 2020.
- Hall, B.D.; Borbely, J.S. GTC Documentation. Available online: <https://gtc.readthedocs.io/en/stable/> (accessed on 13 November 2021).
- Hall, B.D. Evaluating the measurement uncertainty of complex quantities: A selective review. *Metrologia* **2015**, *53*, S25–S31. [CrossRef]
- Willink, R.; Hall, B.D. A classical method for uncertainty analysis with multidimensional data. *Metrologia* **2002**, *39*, 361–369. [CrossRef]

21. Lafarge, T.; Possolo, A. The NIST Uncertainty Machine. *Ncsli Meas. J. Meas. Sci.* **2015**, *10*, 20–27. [[CrossRef](#)]
22. Zeier, M.; Wollensack, M.; Hoffmann, J. METAS UncLib-a measurement uncertainty calculator for advanced problems. *Metrologia* **2012**, *49*, 809–815. [[CrossRef](#)]
23. Wollensack, M.; Hoffmann, J.; Ruefenacht, J.; Zeier, M. VNA Tools II: S-parameter uncertainty calculation. In Proceedings of the 79th ARFTG Microwave Measurement Conference Digest, Montreal, QC, Canada, 22–22 June 2012. [[CrossRef](#)]
24. Lebigot, E.O. Uncertainties: A Python Package for Calculations with Uncertainties. Available online: <http://pythonhosted.org/uncertainties> (accessed on 13 November 2021).
25. Bevington, P.R.; Robinson, D.K. *Data Reduction and Error Analysis*; McGraw-Hill: New York, NY, USA, 2003.

Article

Designing Possibilistic Information Fusion—The Importance of Associativity, Consistency, and Redundancy

Christoph-Alexander Holst * and Volker Lohweg

inIT—Institute Industrial IT, Technische Hochschule Ostwestfalen-Lippe, Campusallee 6, 32657 Lemgo, Germany; volker.lohweg@th-owl.de

* Correspondence: christoph-alexander.holst@th-owl.de

Abstract: One of the main challenges in designing information fusion systems is to decide on the structure and order in which information is aggregated. The key criteria by which topologies are constructed include the associativity of fusion rules as well as the consistency and redundancy of information sources. Fusion topologies regarding these criteria are flexible in design, produce maximal specific information, and are robust against unreliable or defective sources. In this article, an automated data-driven design approach for possibilistic information fusion topologies is detailed that explicitly considers associativity, consistency, and redundancy. The proposed design is intended to handle epistemic uncertainty—that is, to result in robust topologies even in the case of lacking training data. The fusion design approach is evaluated on selected publicly available real-world datasets obtained from technical systems. Epistemic uncertainty is simulated by withholding parts of the training data. It is shown that, in this context, consistency as the sole design criterion results in topologies that are not robust. Including a redundancy metric leads to an improved robustness in the case of epistemic uncertainty.

Keywords: information fusion; possibility theory; information fusion system design

Citation: Holst, C.-A.; Lohweg, V. Designing Possibilistic Information Fusion—The Importance of Associativity, Consistency, and Redundancy. *Metrology* **2022**, *2*, 180–215. <https://doi.org/10.3390/metrology2020012>

Academic Editor: Simona Salicone

Received: 17 March 2022

Accepted: 6 April 2022

Published: 11 April 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The discipline of information fusion is concerned with the aggregation of uncertain information from several sources. Through the process of fusion, uncertainty is to be reduced, that is, information fusion aims at creating information of higher quality [1].

Uncertainty and ignorance manifest in many forms, such as a lack of confidence, aleatoric uncertainty, or epistemic uncertainty. A comprehensive taxonomy of ignorance is provided by Ayyub and Klir [2]. Uncertain information are modelled in various mathematical frameworks, especially probability theory, Dempster–Shafer theory, fuzzy set theory, and possibility theory [3], and each has strengths and weaknesses with regard to types of uncertainty. Possibilistic information fusion is focused on handling epistemic uncertainty, imprecise information, and incomplete information [4,5], which stem from, e.g., scarce data, repetitive data, or biased data. In possibilistic information fusion, knowledge about the state of affairs is complemented by excluding alternatives, which single information sources deem impossible.

In the following, this paper relies on the nomenclature of information items and information sources adopted from [6].

Definition 1 (Information Item). Consider an unknown entity v and a non-empty set of possible alternatives $X_A = \{x_1, \dots, x_n\}$ with $n \in \mathbb{N}_{>0}$. An information item models information in the form of plausibilities or probabilities about v regarding X_A . An information item can be, e.g., a set, an interval, a probability distribution, or a possibility distribution. Consequently, an item may be expressed with certainty ($v = x$ or, assuming $A \subset X_A$, $v \in A$), may be affected by uncertainty (v is probably x or v is possibly x), or may be expressed imprecisely ($x_1 < v.s. < x_2$).

Definition 2 (Information Source). An information source S provides information items. It is an ordered concatenation of information items $S = \{I_1, I_2, \dots, I_m\}$ with $m \in \mathbb{N}_{>0}$. Each I_j represents an information item at instance $j \in \{1, \dots, m\}$. An information source may be, for example, a technical sensor, a variable, a feature, or a human expert.

Often information fusion benefits from distributing the fusion into a multi-step piecewise process [7–10]. This means, for example, that information items are fused sequentially, in parallel, or hierarchically instead of centralised all at once. The sequence in which items are fused is often referred to as the topology or architecture. While the term architecture is often used in a broader sense to refer to complete fusion frameworks (see [11–14]), the term topology is used in this paper to describe the structure in which the fusion is arranged. Example fusion topologies are shown in Figure 1.

Designing and optimising a fusion topology is one of the main challenges in implementing an information fusion system [15]. An optimal topology reduces communicational and computational loads, increases fusion accuracy [16], and helps to detect defective sources [17]. Fusion topologies are usually designed manually, as e.g., in the dissertation of Mönks [18] or require meta-knowledge about information sources, such as in the work of Fritze et al. [19]. Automated learning processes are rare. Such a learning process is made more difficult by epistemic uncertainty due to, e.g., missing or underrepresented classes in training data or due to having few training data instances to begin with. This calls for approaches of learning topologies based on possibility theory.

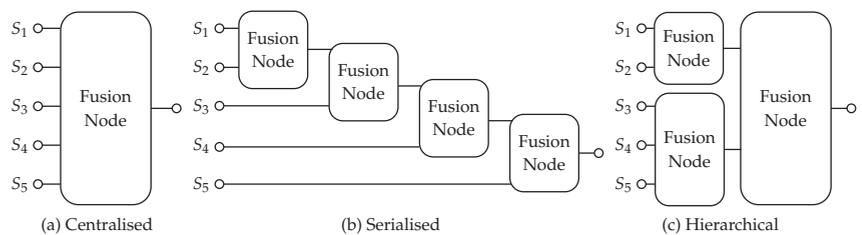


Figure 1. Three example information fusion topologies. (a) Centralised fusion, (b) serial fusion, and (c) hierarchical fusion.

Key characteristics for designing fusion topologies are the associativity of fusion rules, consistency of information items, and redundancy of information sources. Associativity allows the optimisation of a topology towards, e.g., computational load or other criteria without having to worry about distorting the fusion result. Associativity is especially crucial if a specific topology is necessitated by an application.

Information may not be available at the same time or information sources may be spatially distributed so that a centralised fusion is simply not feasible. Structuring fusion based on consistency or redundancy was proposed quite early [17,20]. The basic idea is to fuse consistent or redundant information in earlier stages and complementary information in later stages. Grouping sources in this way provides the benefits that (i) it is reasonable to conduct fusion conjunctively resulting in maximal certain information [6] and (ii) it is easier to identify defective or malfunctioning sources increasing the robustness of applications [21–24].

In this article, we contribute an approach towards a data-driven automated learning of information fusion topologies. The article focuses on information modelled within the possibility theory. As a foundation, common possibilistic fusion rules are recapitulated and analysed regarding the associativity property. Based on this analysis, design algorithms relying on consistency and redundancy are proposed and discussed. The aim of the design algorithms is to build topologies that result in maximal specific (i.e., minimal uncertain) fusion outcomes and that facilitate source defect detection. The proposed learning algorithm approaches are discussed with regard to their robustness and further improved by exploiting outlier resistant averaging possibilistic fusion rules. As a first

step in this article, an overview of the state of the art in fusion topology design is given independent of the mathematical framework.

2. Fusion Topology Design in Related Work

Information fusion systems are composed of various interacting parts and methodologies, such as information sources, information pre-processing, fusion nodes, mathematical frameworks, or fusion algorithms. This results in high-dimensional design spaces, i.e., a large amount of hyperparameters. Deciding on and designing the topology is an important subtask in fusion system design as identified by Raz et al. [16]. The authors explored the design space of a relatively simple fusion task (still $> 2 \times 10^5$ design combinations) with the help of machine-learning algorithms. Their goal was to estimate the impact of design choices on the performance of the fusion system. Among other design parameters, the topology and allocation of sources to fusion nodes were identified to be crucial to the performance. This motivated ongoing work on topology design.

A widely used approach towards designing topologies and allocating information sources is to rely on meta-knowledge about the information sources. Mönks et al. [18,25] grouped information sources (here: technical sensors) into a two-level fusion topology based on the sensor's observed objects, measured physical property, or spatial location. Semantically close (e.g., observing the same object) or spatially close sensors are assumed to be at least partly-redundant and are allocated to the same fusion node. This manual approach has been partly automated by Fritze et al. [19,26,27] who equipped sensors with a self-description containing information about the sensor's characteristics, its contextual environment, and observed objects. A rule-based system then matches and groups sensors based on their self-description. Other ontology-based approaches have been proposed by Boury-Brisset [28] and Martí et al. [29]. Both do not focus on topology design specifically but rather on designing or facilitating a fusion system. Boury-Brisset [28] discussed ontological methods for the integration in the Joint Directors of Laboratories (JDL) fusion architecture [30] including the semantic integration of information. Martí et al. [29] proposed an ontology-based adaptive sensor fusion architecture, and this architecture organises sensors and external sources into preprocessing nodes and fusion nodes depending on the task at hand. A recent application of ontology-based design of information fusion systems can be found in the field of assisted living [31]. Ontological approaches reduce the manual effort needed for structuring fusion topologies; however, they still require profound expert knowledge about the information sources and their context. Building the ontology requires manual engineering and is time-consuming [28].

Designing information fusion topologies is closely related to the *data association* step predominately but not exclusively used in the JDL fusion architecture. Solaiman and Bossé [32] refer to the task of data association as the identification of any relation between information elements and monitored objects. Waltz and Llinas [33] defined the data association problem with regard to fusion systems more specifically as the “Cross correlation of measurements and m -ary decisions to partition all measurements into sets of common origin. One can distinguish between associating a set of measurements (partitioning) and associating a measurement (or a set of measurements) to a given object. [...]”.

In this definition, the partitioning of measurements refers to preparing a fusion task in which each partition represents the input to a fusion node; hence, the relation to designing fusion topologies. Data-driven approaches for data association are given by Grabisch and Prade [34] and Ayoun and Smets [35]. Both approaches cluster sensor measurements based on quantifications of the measurements' proximities. Grabisch and Prade [34] modelled information within the possibility theory and computed the proximity based on the degree of intersection of possibility distributions. Ayoun and Smets [35] used Dempster–Shafer theory instead and clustered based on the degree of conflict between measurements. A similar approach was taken by Schubert [36,37]—although not explicitly labelled as data association—who clustered basic belief functions (evidential masses) based on their conflict and attraction with each other. All of these works ([34–37]) partition information sources

based on single instances of measurements (the current measurement) and not on historical data. More sophisticated interdependencies and interrelations between information sources can only be detected robustly in historical data. For example, for the identification and quantification of redundancies between sources, meaningful data are necessary, which spans over the sources' frame of discernment as shown by Holst and Lohweg [38,39].

Regarding the problem of data association, it has to be mentioned that more recent publications focus solely on the specific application task of visual target tracking (see for example the works of Kamal et al. and Yoon et al. [40,41]). This focus comes with a shift in interpretation of the data association problem as shown by the definition given by Khaleghi et al. [42]: “[...] the data association problem, which may come in two forms: measurement-to-track and track-to-track association. The former refers to the problem of identifying from which target, if any, each measurement is originated, while the latter deals with distinguishing and combining tracks, [...]”. Publications with this shifted focus are less related to the problem of designing fusion topologies.

In summary, in related works, the task of structuring fusion topologies has been approached based on expert knowledge, ontologies, or based on current measurements. Approaches that consequently analyse historical data or information in order to derive a fusion topology are missing. While this section considered topology design independently from the mathematical fusion framework, the remainder of this paper focuses on possibility theory.

3. Fusion within Possibility Theory

To provide a basis for a discussion on fusion topology design, the importance of associativity, and the role of consistency and redundancy, the core principles of possibility theory (PosT) are recapitulated. For this, common fusion rules are also reported in detail.

The main motivation behind PosT is that probability theory (ProbT) is not able to model epistemic uncertainty adequately—such as imprecision or missing information. Probability theory models random phenomena quantitatively; PosT handles incomplete information qualitatively [5,43]. Zadeh [44] introduced PosT based on fuzzy sets in the context of natural language processing. He interpreted fuzzy membership functions as possibility distributions allowing uncertainties in the sense of imprecisions as well as a lack of confidence in statements [45].

Consequently, PosT is mathematically close to fuzzy set theory [46]. This proximity often allows mathematical operations defined in the context of fuzzy sets—such as similarity measures or t-norms—to be applied to possibility distributions. Since Zadeh's introduction of PosT, Dubois and Prade [4,6,47–49] and YAGER [50–53] have mainly contributed to the advancement of possibility theory. If not explicitly mentioned otherwise, a numerical, real-valued representation of possibility values is assumed (cf. Dubois et al. [6] for an overview of qualitative and numerical possibility scales).

A possibility distribution is defined as a mapping of mutually exclusive and exhaustive alternative events to a numerical representation. Let the set of all alternative events be described as the *frame of discernment* X and let $v \in X$ be an imprecisely known element whose true value is unknown. Then, a possibility distribution is defined by

$$\pi_v : X \rightarrow [0, 1]. \tag{1}$$

Alternatives $x \in X$ that are assigned higher values are deemed more plausible. Alternatives with $\pi_v(x) = 0$ are considered impossible, and alternatives with $\pi_v(x) = 1$ are fully plausible. Possibility theory is strongly guided by the *minimum specificity principle*, which states that any alternative x not known to be impossible should not be disregarded [45]. Extreme cases of knowledge about v are *total ignorance* and *complete knowledge*. In the first case, $\forall x \in X : \pi_v(x) = 1$. In the case of complete knowledge, only one alternative is fully possible, and all others are impossible. A possibility distribution $\pi_v(x)$ is said to be *normal* if $\exists x \in X : \pi_v(x) = 1$. The subset $A \subseteq X$, which $\forall x \in A : \pi_v(x) = 1$ is referred to as

core of $\pi_v(x)$; if $\forall x \in A : \pi_v(x) \geq 1$, then A is referred to as *support*. In the following, the shortened notation $\pi(x) = \pi_v(x)$ is used.

Let multiple information sources $\mathbf{S} = \{S_1, \dots, S_n\}$ each provide an information item $I_i, i \in \{1, \dots, n\}$ in the form of a possibility distribution π_i regarding the same imprecisely known element $v \in X$. A possibilistic fusion operator is then defined by $fu : [0, 1]^n \rightarrow [0, 1]$ and the fused possibility distribution is obtained as $\pi^{(fu)}(x) = fu(\pi_1(x), \dots, \pi_n(x))$. Multiple information sources allow the identification of even more impossible or hardly possible alternatives for the unknown v resulting in more precise, more specific, and thus more qualitative information. In this sense, the goal in possibilistic fusion is to reach a maximal specific outcome (the most certain outcome possible) although possibility theory follows the minimum specificity principle. It is important that none of the available information is disregarded or neglected—that is, that any information source is considered by the fusion process (see also the *fairness* property postulated for fusion operators [6]). This fairness constraint represents the minimum specificity principle stating that alternatives that are not known to be impossible are not to be ruled out [45].

Over time, multiple possibilistic fusion operators have been proposed, verified, and brought to applications. We propose to categorise these operators as follows:

- **Possibilistic Pooling Fusion** has mainly been advanced by Dubois et al. [4,48]. The aim of possibilistic pooling is to find the possibility degree for each alternative x . Hence, operators work on the grades of possibilities (by applying fuzzy norms). Inside this framework, the choice of fusion rules is most often based on the state of knowledge about the reliability of the information sources involved. Depending on reliability and available knowledge, fusion operators are distinguished into *conjunctive*, *disjunctive*, and *trade-off* modes [32].
- **Possibilistic Estimation Fusion** was mainly devised and advanced by Yager [54]. In contrast to pooling, estimation operators are based on *Zadeh's extension principle* [55], which defines the use of mappings to fuzzy inputs. The goal of estimation concerns itself with finding the result that is the most compatible with all information items. Operators apply averaging functions on the frame of discernment X .
- **Majority-guided Fusion** identifies majority subsets—often based on consistency measures—and aggregates information from these subsets either exclusively or prioritised—similar to a voting procedure. Majority-guided fusion deliberately violates the fairness principle. It finds application in situations in which it is explicitly known that sources produce consistent readings, e.g., in redundantly engineered technical sensor systems [23]. The operators for majority-guided fusion are often based upon either pooling or fuzzy estimation as is shown in detail in the following.

3.1. Possibilistic Pooling Fusion

Conjunctive and disjunctive fusion is most commonly performed using triangular norms (*t-norms*) and their counterpart triangular conorms (*s-norms*)—both stemming from fuzzy set theory. Triangular norms and conorms are functions $t, s : [0, 1] \times [0, 1] \rightarrow [0, 1]$, which satisfy the properties of commutativity, associativity, and monotonicity [56]. For *t-norms*, 1 is the identity element, i.e., $t(\pi, 1) = \pi$. For *s-norms*, 0 is the identity element, i.e., $s(\pi, 0) = \pi$. Examples of *t-norms* are the minimum and the product operator. An example of an *s-norm* is the maximum operator. Although *t-norms* and *s-norms* are defined as binary functions, they can be directly applied to multiple possibility distributions because of their commutative and associative property.

In conjunctive mode, it is presumed that sources agree at least partially about the possibility of alternatives, that is, their information items are at least partially consistent. Consistency within a group of information items \mathbf{I} is defined as [4]

$$h(\mathbf{I}) = h(\pi_1, \pi_2, \dots, \pi_n) = \max_{x \in X} \left(t_{i \in \{1, \dots, n\}}(\pi_i(x)) \right). \tag{2}$$

Partially agreeing sources are characterised by items with $h(\mathbf{I}) > 0$ —that is, their possibility distributions have overlapping support. Fully agreeing sources have items with $h(\mathbf{I}) = 1$, i.e., their possibility distributions have overlapping cores. Conjunctive fusion of fully consistent information items is then achieved by directly applying a t-norm [48]:

$$\pi^{(\text{fu})}(x) = \mathop{\text{t}}_{i \in \{1, \dots, n\}} (\pi_i(x)). \tag{3}$$

As t-norms satisfy the strong zero preservation principle, i.e., $t(\pi, 0) = 0$, the conjunctive fusion excludes all alternatives, which at least one information source deems impossible. Conjunctive fusion results in the most specific outcome by eliminating alternatives. If information items are only partially consistent, then fusion based on t-norms results in subnormal possibility distributions. Renormalising the resulting possibility distribution leads to

$$\pi^{(\text{fu})}(x) = \frac{\mathop{\text{t}}_{i \in \{1, \dots, n\}} (\pi_i(x))}{\mathop{\text{h}}_{i \in \{1, \dots, n\}} (\pi_i)}, \tag{4}$$

which is only defined if sources are not completely disagreeing and if their information items not fully inconsistent, i.e., $h \neq 0$ [48].

The disjunctive fusion is appropriate if information items are completely inconsistent, i.e., sources disagree, at least one of them is wrong in its assessment, and it is not known which one. The disjunctive fusion is given by applying an s-norm:

$$\pi^{(\text{fu})}(x) = \mathop{\text{s}}_{i \in \{1, \dots, n\}} (\pi_i(x)) \tag{5}$$

keeping all available information. In general, purely disjunctive fusion is not desirable as it results in minimal specific outcomes but is necessary in disagreeing cases.

Trade-off fusion modes combine conjunctive and disjunctive fusion depending on what is known (or assumed) about the reliability of sources. Prominent fusion rules can be found in the paper of Dubois and Prade [4]. For this paper, the most important of these are fusion based on the most consistent subsets, quantified fusion, and adaptive fusion.

One prominent way to aggregate information in a two-step process is to search for *maximal consistent subsets* (MCS) [20,57]. These nonconflicting MCS are fused conjunctively prior to disjunctive fusion of intermediate results. Dubois et al. [58] proposed an algorithm that finds MCS with linear complexity. In this algorithm, all subsets of \mathbf{I} with a consistency above or equal to $\alpha \in [0, 1]$ are clustered. Let $\mathbf{I}^{\text{MCS}} \subseteq \mathbf{I}$ denote MCS subsets, then MCS fusion is formalised for a possibilistic setting as [6]:

$$\pi^{(\text{fu})}(x) = \max_{\mathbf{I}^{\text{MCS}} \subseteq \mathbf{I}} \left(\mathop{\text{t}}_{I_i \in \mathbf{I}^{\text{MCS}}} (\pi_i(x)) \right). \tag{6}$$

Later advancements in MCS fusion were proposed in multiple works [59–61].

Quantified fusion [62,63] is a similar two-step fusion process, which assumes that the number of reliable sources j is known. The quantified rule then takes all subsets of information items $\mathbf{I}_* \subseteq \mathbf{I}$ with cardinality j and fuses these conjunctively in the first step. All intermediate results are then fused disjunctively:

$$\pi^{(\text{fu})}(x) = \max_{\substack{\mathbf{I}_* \subseteq \mathbf{I} \\ |\mathbf{I}_*| = j}} \left(\min_{I_i \in \mathbf{I}_*} (\pi_i(x)) \right). \tag{7}$$

Adaptive fusion aims at progressing gradually from conjunctive to disjunctive behaviour as conflict increases. A simple adaptive fusion rule is

$$\pi^{(fu)}(x) = \max \left(\frac{\min_{i \in \{1, \dots, n\}} (\pi_i(x))}{h}, \min \left(\max_{i \in \{1, \dots, n\}} (\pi_i(x)), 1 - \frac{h}{\max_{i \in \{1, \dots, n\}} (\pi_i)} \right) \right). \quad (8)$$

It fuses all sources disjunctively (assuming one source is right) and discounts the result by $(1 - h)$. In parallel, it fuses all sources conjunctively (assuming all sources are right) and combines both intermediate results. This process does not consider situations in which more than one or less than all sources are reliable. If many sources are fused, it is likely that $h \rightarrow 0$, thus, resulting in uninformative results [4]. Dubois' adaptive fusion rule [4,48] builds upon the quantified (7) and adaptive fusion rule (8) assuming that a minimum and maximum number of reliable sources are known. The minimum and maximum number are derived from the consistency of information items I . The cardinality of the largest fully consistent subset gives the minimum number $j^- = \max(|I| \mid h(I) = 1)$; the largest partially consistent subset provides the maximum number $j^+ = \max(|I| \mid h(I) > 0)$. The adaptive fusion is then

$$\pi^{(fu)}(x) = \max \left(\frac{\pi_+^{(fu)}(x)}{h}, \min \left(\pi_-^{(fu)}(x), 1 - h_+ \right) \right) \quad (9)$$

in which $\pi_+^{(fu)}(x)$ and $\pi_-^{(fu)}(x)$ are obtained by quantified fusion (7) (with j^- and j^+ , respectively) and $h_+ = \max_{I_* \subseteq I \mid |I_*| = j^+} (h(I_*))$. In this way, completely disagreeing sources with fully inconsistent items ($h = 0$) are disregarded. Furthermore, small changes in the input possibility distributions may lead to significant changes in the fusion result [64].

Oussalah et al. [64] proposed changes to (9) improving the behaviour in the case of outliers and with regard to robustness against small changes. For their progressive fusion rule, they introduced a distance measurement with which the disjunctive fusion ($\pi_-^{(fu)}(x)$) part is adapted. Let $x_0, x_1 \in X$ be the smallest and largest element of the consensus set, then

$$d(x) = \begin{cases} \max(|x - x_0|, |x - x_1|) & \text{if } x < x_0 \text{ or } x > x_1, \\ 0 & \text{otherwise,} \end{cases}$$

measures the distance from point x to the consensus set. Let $\alpha(x) = \min\left(\frac{d(x)}{d_0}, 1\right)$ be a weighting factor. The threshold d_0 is the maximum distance until outliers are considered. Then, $\pi_-^{(fu)}(x)$ in (9) is replaced by

$$\pi_-^{(fu)}(x) = \alpha(x) \cdot \pi_+^{(fu)}(x) + (1 - \alpha(x)) \cdot \max_{i \in \{1, \dots, n\}} (\pi_i(x)). \quad (10)$$

Instead of (9), (10) considers the completely disjunctive fusion of all information items. The degree to which it considers disjunction relies on $d(x)$. The further x is from the consensus set, the more consideration is given to inconsistent items.

3.2. Possibilistic Estimation Fusion

Whereas pooling fusion aims at discarding alternatives, estimation fusion assumes that none of the sources are completely wrong and attempts to find a fusion result that is compatible with all information items [4]. Nonetheless, more specific or precise outcomes are still preferable. Estimation fusion has received less attention in the scientific community compared with pooling fusion (The higher number of citations of Dubois's paper [4] compared to Yager's paper [65] reflect the higher attention). Therefore, the following discussion takes a deeper look into the algebraic properties of estimation fusion.

Estimation fusion is based on Zadeh’s extension principle, which allows mapping functions to be used on fuzzy sets [66]. Let Y, Z be a frame of discernments and $F : Y \rightarrow Z$. Let A be a fuzzy set defined on Y and B a fuzzy set defined on Z , and then F maps the fuzzy membership function $\mu_A(y)$ with $y \in Y$ to $\mu_B(z)$ with $z \in Z$ by $\mu_B(z) = \mu_A(F^{-1}(z)) = \mu_A(y)$ with $z = F(y)$. If F results in multiple outputs for the same y , then

$$\mu_B(z) = \max_{y \in Y: F(y)=z} \mu_A(y).$$

In multi-source estimation fusion, the input possibility distributions are first pooled by a fusion function—here referred to in this context as G . The result is then mapped by the multi-parameter function $F(x_1, x_2, \dots, x_n)$ with $x_i \in X_i, i \in \{1, \dots, n\}$ onto a new frame of discernment X , i.e.,

$$\pi^{(fu)}(x) = \max_{x_i \in X_i: F(x_1, \dots, x_n) = x} \left(G(\pi_i(x_i)) \right), \tag{11}$$

for which the notation

$$\pi^{(fu)}(x) = \left\{ \frac{G(\pi_1(x_1), \pi_2(x_2), \dots, \pi_n(x_n))}{F(x_1, x_2, \dots, x_n)} \right\}. \tag{12}$$

is used in the following. The fusion rule in (11) takes the maximum of $G_i(\pi_i(x_i))$ for every n -tuple $(x_1 \in X_1, \dots, x_n \in X_n)$, which satisfies $F(x_1, \dots, x_n) = x$.

Yager [65] proposed an estimation fusion rule in which G is the minimum operator and F is defined to be an averaging operator.

Definition 3 (Averaging Operator). *An operator that satisfies the three properties of commutativity, monotonicity, and idempotency, is referred to as a mean or averaging operator [4]. Such an averaging operator $\text{avg}(\cdot)$ lies between $\min(\cdot)$ and $\max(\cdot)$, i.e., $\min(\cdot) \leq \text{avg}(\cdot) \leq \max(\cdot)$.*

Yager’s estimation fusion rule [65] is then:

$$\pi^{(fu)}(x) = \left\{ \frac{\min_{i \in \{1, \dots, n\}} (\pi_i(x_i))}{F(x_1, x_2, \dots, x_n)} \right\}. \tag{13}$$

The application of the minimum operator results in maximal specific possibility distributions, which are placed on an averaged frame of discernment. The disadvantages of estimation fusion are that (i) it requires a frame of discernment for which it is sensible to apply averaging operators on and that (ii) estimation fusion may lead to fusion results that have been deemed impossible by all sources, i.e., the results do not satisfy the zero preservation principle [4]. Regarding the first disadvantage, it is often assumed that $X \subseteq \mathbb{R}$ [65], which is also assumed for the remainder of this section.

If G is also an averaging operator, then a noteworthy interaction between estimation fusion and the frame of discernment takes place, which is relevant for practical implementations.

Proposition 1. *If G is an averaging operator other than the minimum operator and $X \subseteq \mathbb{R}$, then fusion with (13) is influenced by the borders of X . More formally, $\min_{\pi^{(fu)}(x) > 0} x$ is dependent on $\min_{x \in X} x$ and $\max_{\pi^{(fu)}(x) > 0} x$ on $\max_{x \in X} x$.*

Proof. Let $x_a = \min_{x \in X} x$ and $x_b = \max_{x \in X} x$, i.e., $X = [x_a, x_b]$. Let $x' = \min_i \min_{x_i \in X_i: \pi_i(x_i) > 0} x_i$, i.e., x' is the smallest element in X for which at least one $\pi_i > 0$. Furthermore, let $i' = \arg \min_i \min_{x_i \in X_i: \pi_i(x_i) > 0} x_i$. If $G \neq \min$, then, for at least one permutation of the n -tuple

$(x_a, x_a, \dots, x', \dots, x_a, x_a): G \left(\pi_1(x_a), \dots, \underbrace{\pi_{i'}(x')}_{>0}, \dots, \pi_n(x_a) \right) > 0$. This n -tuple defines the minimum boundary of $\pi^{(fu)}$, i.e., $\min_{\pi^{(fu)}(x) > 0} x = F(x_a, \dots, x', \dots, x_a)$. The same holds for the maximum boundary of $\pi^{(fu)}$ only that $x' = \max_i \max_{x_i \in X_i: \pi_i(x_i) > 0} x_i, i' = \arg \max_i \max_{x_i \in X_i: \pi_i(x_i) > 0} x_i$, and $\max_{\pi^{(fu)}(x) > 0} x = F(x_b, \dots, x', \dots, x_b)$. \square

An example of the effects of Proposition 1 is illustrated in Figure 2.

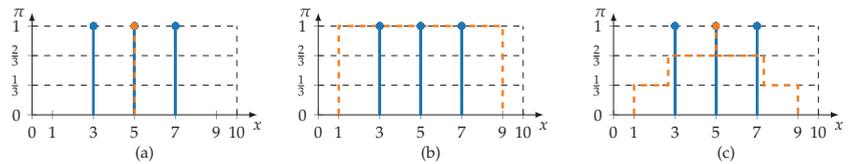


Figure 2. An example of the interaction between estimation fusion (12) and X as discussed in Proposition 1. A frame of discernment $X = [0, 10]$ and three possibility distributions are given. Each possibility distribution claims complete knowledge; $\pi_1(x = 3) = 1, \pi_2(x = 5) = 1$, and $\pi_3(x = 7) = 1$. The plots show fusion results (dashed red) in which F is the arithmetic mean and G is (a) the minimum, (b) the maximum, and (c) the arithmetic mean operator.

Corollary 1. *If X is also unbounded and F is an averaging operator other than the minimum or maximum operator, then (13) results in an unbounded $\pi^{(fu)}$. If X is half-bounded, then $\pi^{(fu)}$ is also half-bounded.*

Proof. From Proposition 1 it follows directly that, if $F \neq \max$, then $\lim_{x_a \rightarrow -\infty} \min_{\pi^{(fu)}(x) > 0} x = F(x_a, \dots, x', \dots, x_a) = -\infty$. If $F \neq \min$, then $\lim_{x_b \rightarrow \infty} \max_{\pi^{(fu)}(x) > 0} x = F(x_b, \dots, x', \dots, x_b) = \infty$. \square

Consequently, if G is an averaging operator other than the minimum operator, then it is reasonable to apply estimation fusion only on bounded X . Otherwise, (12) and (13) lead to fusion results spanning to infinity—even for very precise input possibility distributions.

3.3. Majority-Guided Fusion

In essence, fusion rules, which focus and prioritise the consensus set—often also referred to as majority observation—fall under the category of majority-guided fusion. Majority-guided fusion is particularly sensible in cases in which information sources are known to produce consistent items. Possibility distributions deviating from the consensus set are then deduced to be faulty (unreliable) instead of giving useful information about the unknown value v .

With this in mind, Dubois’ fusion rule (9) already satisfies as a majority-guided fusion rule because it ignores all inconsistent information items (although this fact is precisely one of the main points of criticism by Oussalah et al. [64]). In the specific case of assuming fully reliable sources and expecting consistency between items, it is reasonable to rely on simpler fusion rules; accordingly, it was proposed to use a purely conjunctive fusion rule [23]. Similarly simple are counting fusion functions; the result here is the alternative that most sources consider possible [5].

Estimation fusion rules, such as (13), favour the majority observation because of the averaging characteristic of the estimation operator F . A more complex majority-guided fusion rule, which is based on Yager’s estimation fusion (13), was proposed by Glock et al. [67], the *majority-opinion-guided possibilistic fusion rule* (MOGPFR). The MOGPFR replaces both

the conjunctive fusion part G and the estimation operator F with the *Implicative Importance Weighted Ordered Weighted Averaging* (IIWOWA) operator. The IIWOWA operator, as proposed by [68], is an extension of the parent class of Ordered Weighted Averaging (OWA) operators [50]. An OWA operator allows weighting inputs with $\mathbf{w} = (w_1, \dots, w_n)$, $w_i \in [0, 1]$, and $\sum_i w_i = 1$. Inputs π_i are ordered in descending order. This results in aggregation $\frac{1}{n} \sum w_i \cdot \pi_i$ and allows the aggregation to be shifted between the minimum with $\mathbf{w} = (0, 0, \dots, 1)$ and maximum $\mathbf{w} = (1, \dots, 0, 0)$. The MOGPFR is then defined as follows:

$$\pi^{(fu)}(x) = \max_i(reli_i) \cdot \hat{\pi}^{(fu)}(x) + 1 - \max_i(reli_i), \text{ with}$$

$$\hat{\pi}^{(fu)}(x) = \left\{ \frac{\lambda_{\text{IIWOWA}}(\mathbf{v}, \mathbf{w}_p, \pi_i(\mu^{(i)}))}{\lambda_{\text{IIWOWA}}(\mathbf{v}, \mathbf{w}_m, \mu^{(i)})} \right\}; \tag{14}$$

in which $\lambda_{\text{IIWOWA}}(\circ)$ denotes the IIWOWA operator, and rel_i is the reliability for each source. The MOGPFR specifically allows the control of fusion by (i) a reliability vector $\mathbf{v} = \{v_1, v_2, \dots, v_n\}$ with $v_i \in [0, 1]$, which discounts informations items and (ii) two weighting vectors, \mathbf{w}_p and \mathbf{w}_m , which control whether G and F are close to the minimum or maximum operator, respectively. The IIWOWA operator is defined only for inputs in $[0, 1]$, which necessitates the fuzzification of X so that the possibility distributions become $\pi_i(\mu^{(i)})$.

The MOGPFR facilitates the prioritisation of information items belonging the majority observation. The importance values v_i are determined by a distance function of π_i to the majority set; the possibility distribution π_i is discounted accordingly. The parameters \mathbf{w}_p and \mathbf{w}_m allow adapting fusion towards conjunctive and disjunctive behaviour. The benefit gained by the MOGPFR lies in its level of control through parametrisation.

4. Approach towards Topology Design

Associative fusion rules allow changing the sequence in which information sources are fused without altering the fusion result. Therefore, associativity is a beneficial property with regard to the topology design of distributed information fusion systems. Assuming associativity, a system designer or a design algorithm can focus on other criteria for designing a fusion system, such as spatial availability of sources or consistency as well as the redundancy of sources. In this section, we analyse the presented fusion rules regarding the associativity property and its impact on topology design. Following this, a two-layer fusion topology based on the MCS fusion rule (6) is presented. Consistency as a design criterion both increases the specificity of fusion results due to the minimum-operator [6] and to facilitate source defect detection algorithms [21,22]. This motivates the dive into the MCS fusion topologies in this article.

Some flaws and shortcomings of this consistency-based approach are discussed, which leads to several adjustments to overcome those. This includes the introduction of redundancy as a design criterion.

First, both fusion node and fusion topology are defined, and some notations introduced:

Definition 4 (Fusion Node). *A fusion node fn is a self-contained module encapsulating a fusion operator. A node takes information items as input and outputs a single fused information item. As a node is a self-contained module, a fusion node and its fusion operator have to satisfy the following additional properties:*

- **Modularity:** *A fusion node outputs a fused information item, which qualifies as a possibility distribution π (see Section 3), i.e., π is normal. This property allows self-contained intermediate results in a topology and makes fusion nodes modular. This increases the transparency of the distributed fusion topology.*
- **Self-Reproducing:** *Given a single input, a fusion node reproduces this input. It preserves its identity, i.e., $fu(I) = I$.*

Idempotency as a property is not required since idempotency restricts the fusion node in the case where a reinforcement effect is desired (e.g., via the product operator as a t-norm). A fusion node with an associative fusion operator is beneficial since it allows splitting the fusion node.

A fusion node is a modular part of a fusion topology. In order to facilitate the fusion process of the grander topology, it may output auxiliaries denoted as [AUX]. Consequently, a node is also required to be able to process [AUX] as input if necessary.

Definition 5 (Fusion Topology). Interconnected fusion nodes build up a fusion topology. Fusion nodes may be interconnected parallelly, serially, hierarchically, cascadingly, or in more complex structures. A fusion topology organises a feed forward flow of information. Recursive interconnections are excluded. A fusion topology is constructed in layers $l \in \mathbb{N}_{>0}$. In each layer, fusion nodes are indexed consecutively with $k \in \mathbb{N}_{>0}$. The k -th fusion node in layer l is denoted by $fn_{(k,l)}$, its output information item by $I_{(k,l)}$, and its auxiliary output by $[AUX]_{(k,l)}$.

Given the above definitions, Figure 3 shows a three-layer example topology to help visualise the introduced notations.

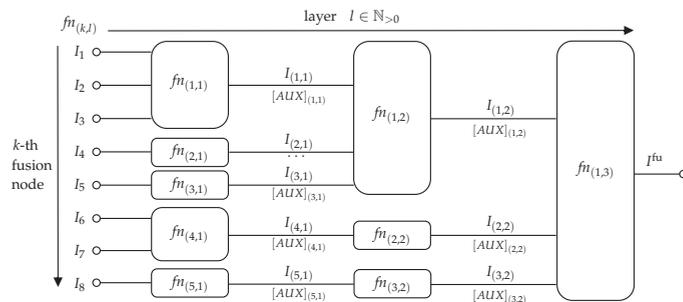


Figure 3. An example for a three-layer fusion topology. Fusion nodes are denoted with $fn_{(k,l)}$ and their output information items with $I_{(k,l)}$ together with auxiliary information $[AUX]_{(k,l)}$. The index l denotes the layer. Within a layer l , the nodes are numbered consecutively by k .

The MCS-based design presented in this article focuses on a two-layer topology by grouping consistent or redundant information sources into fusion nodes. For an easier reading of the article, fusion nodes are also denoted as $fn_{(k)}$ in a two-layer topology. Since this approach considers associative fusion rules, the basic two-layer design can be easily extended into a multi-layer version.

4.1. Associativity

In possibilistic information fusion, the fusion process is rarely considered to be distributed. As a consequence, possibilistic fusion rules are often not associative, which heavily alters the fusion results in differently structured topologies. However, in works regarding possibilistic fusion, associativity has been considered with low priority at best and neglected at worst. For instance, associativity is described as a useful property by Dubois et al. [6]; however, its absence is not considered to be a fatal flaw.

As a first step in discussing associativity, the fusion rules presented in the previous section are summarised in Table 1.

Table 1. Common fusion rules and the property of (quasi-)associativity.

Fusion Rule	Equation(s)	Associative	Proof of Associativity	Quasi-Associative	Proof of Quasi-Associativity
Conjunctive	(3)	yes	Inherited from t-norm	yes	See Proposition 2
Renormalised Conjunctive	(4)	Dependent on t-norm	Proof for nonassociativity in the case of minimum-norm and associativity in the case of product-norm given by Dubois and Prade [47]	yes	$f_{(k)}(\mathbf{I}_{(k)}) = t(\mathbf{I}_{(k)})$ and $g = \frac{1}{h(f_{(k)}(\mathbf{I}_{(k)}))}$
Disjunctive	(5)	yes	Inherited from s-norm	yes	See Proposition 2
MCS fusion	(6)	no	[61]	no	[61]
Quantified	(7)	no	Proof given in Appendix B	no	Similar to MCS fusion
Adaptive	(8), (9)	no	[69]	no	[69]
Progressive	(9), (10)	no	Inherited from adaptive fusion	no	Inherited from adaptive fusion
Estimation	(13)	yes (with restrictions)	See Proposition 3	yes (with restrictions)	See Propositions 2 and 3
MOGPFER	(14)	no	Proof given in Appendix B	no	OWA operator prevents quasi-associativity

The table also shows whether the rules satisfy the following two properties:

Definition 6 (Associativity). *A fusion operator fu is associative if the fusion outcome is independent of the sequence in which information items are fused, i.e., $fu(I_1, I_2, I_3) = fu(fu(I_1, I_2), I_3) = fu(I_1, fu(I_2, I_3))$.*

Definition 7 (Quasi-associativity). *A fusion operator fu is quasi-associative if it can be expressed as a sequence of associative steps and a final operation acting on the results of the previous associative steps [47]. Let f be an associative function and g be a function not restricted to the associativity property, then fu is quasi-associative if $fu(I_1, I_2, I_3) = g(f(f(I_1, I_2), I_3)) = g(f(I_1, f(I_2, I_3)))$.*

Proposition 2. *If a fusion operator is associative, then it is also quasi-associative.*

Proof. Let \mathbf{I} be a set of information items, let $f = fu$ and g be an identity function: $g(\mathbf{I}) = \mathbf{I}$. Then, $g(f(\mathbf{I})) = fu(\mathbf{I})$ —that is, by making use of an identity function, an associative fusion operator becomes quasi-associative. □

From this, it follows that, if a fusion rule is not quasi-associative, then it is also not associative.

Associative rules allow unrestricted topology design in the sense that sources can be freely assigned to fusion nodes without changing the overall fusion result. Quasi-associative rules require a final centralised fusion step in which the nonassociative part is computed. The associative part can be distributed to fusion nodes.

4.1.1. Pooling Fusion

As can be seen in Table 1, the simple conjunctive and disjunctive fusion rules satisfy associativity. However, depending on the applied t-norm—the renormalisation step already causes nonassociative behaviour. For the product norm ($t(\pi_1(x), \pi_2(x)) = \pi_1(x) \cdot \pi_2(x)$), fusion stays associative; however, generally, renormalisation prevents associativity [47].

MCS fusion (6) is based on the idea that consistent information items are to be fused conjunctively first before the results are fused disjunctively. MCS fusion thus specifies a sequence in which information is to be fused. Consequently, MCS fusion is not associative. It is quite easy to see that different sequences result in different outcomes (see Appendix B for an example). Quantified fusion (7) has a similar approach, meaning that it fuses conjunctively and disjunctively in two steps. Quantified fusion is—for the same reasons as MCS fusion—not associative and not quasi-associative.

More sophisticated fusion rules—such as adaptive (8), (9) and progressive (10) rules—attempt to make the most of all available information. These fusion rules rely on specific metrics, such as global consistency, consistency between specific subsets, or distances between information items. Many of these metrics are only computable if all information items are available centrally. Since all three rules (8), (9), and (10) are based on the quantified fusion rule, they inherit quantified fusion’s nonassociativity.

4.1.2. Estimation and Majority-Guided Fusion

Estimation fusion (11)–(13) as well as the majority-guided MOGPFR (14) relies on Zadeh’s extension principle.

Proposition 3. *With regard to Zadeh’s extension principle, a fusion operator $fu(\pi_1, \pi_2, \pi_3) = \left\{ \frac{G(\pi_1(x_1), \pi_2(x_2), \pi_3(x_3))}{F(x_1, x_2, x_3)} \right\}$ satisfies associativity if G and F are associative functions and G is monotonic increasing in all its arguments.*

Proof. The operator fu is associative if $fu(\pi_1, \pi_2, \pi_3) = fu(\pi_1, fu(\pi_2, \pi_3))$. With (11), this becomes

$$\max_{x_i \in X_i: F(x_1, x_2, x_3) = x} G(\pi_1(x_1), \pi_2(x_2), \pi_3(x_3)) = \max_{x_1 \in X_1, x' \in X': F(x_1, x') = x} G \left(\pi_1(x_1), \max_{x_2 \in X_2, x_3 \in X_3: F(x_2, x_3) = x'} G(\pi_2(x_2), \pi_3(x_2)) \right).$$

The frame of discernment X' contains every unique element given by $F(x_2, x_3)$ for every 2-tuple (x_2, x_3) with $x_2 \in X_2$ and $x_3 \in X_3$. In the following, the notation $\max_{x_2 \in X_2, x_3 \in X_3: F(x_2, x_3) = x'}$

is shortened to $\max_{F(x_2, x_3) = x'}$ —this also applies to similar notations.

Assume G to be monotonic increasing in all its arguments, i.e., for any $a_i, b_i \in [0, 1]$ with $i \in \{1, 2, \dots, n\}$ and $\forall i : a_i \leq b_i : G(a_1, a_2, \dots, a_n) \leq G(b_1, b_2, \dots, b_n)$. If $\pi_1(x_1) \geq G(\pi_2(x_2), \pi_3(x_2))$, then $G(\pi_2(x_2), \pi_3(x_2))$ has no influence on the term $\max(G(\pi_1(x_1), G(\pi_2(x_2), \pi_3(x_2))))$. If, on the other hand, $\pi_1(x_1) < G(\pi_2(x_2), \pi_3(x_2))$, then $G(\pi_1(x_1), G(\pi_2(x_2), \pi_3(x_2)))$ becomes maximal if $G(\pi_2(x_2), \pi_3(x_2))$ is maximal. Consequently,

$$\begin{aligned} fu(\pi_1, fu(\pi_2, \pi_3)) &= \max_{F(x_1, x') = x} G \left(\pi_1(x_1), \max_{F(x_2, x_3) = x'} G(\pi_2(x_2), \pi_3(x_2)) \right) \\ &= \max_{F(x_1, x') = x} \max_{F(x_2, x_3) = x'} G(\pi_1(x_1), G(\pi_2(x_2), \pi_3(x_2))). \end{aligned}$$

If G is also associative, then

$$fu(\pi_1, fu(\pi_2, \pi_3)) = \max_{F(x_1, x') = x} \max_{F(x_2, x_3) = x'} G(\pi_1(x_1), \pi_2(x_2), \pi_3(x_2)).$$

If F is associative, then

$$fu(\pi_1, fu(\pi_2, \pi_3)) = \max_{F(x_1, x_2, x_3) = x} G(\pi_1(x_1), \pi_2(x_2), \pi_3(x_2)).$$

□

In contrast to the estimation fusion rules, the MOGPFR (14) uses the IIWOWA operator for the functions F and G and is, therefore, not associative. The IIWOWA operator is an extension of the OWA operator. The OWA operator sorts the inputs $(\pi_1(x_1), \dots, \pi_n(x_n))$ in

descending order. It then weights the inputs with a predefined weighting vector (w_1, \dots, w_n) with $w_i \in [0, 1]$. For $(1, 0, \dots, 0)$ the OWA operator becomes the maximum operator, and for $(0, \dots, 0, 1)$, the minimum operator. In these cases, the OWA operator is associative. In all other cases, sorting the input values prevents associativity and quasi-associativity. Consequently, the IIOWA operator and the MOGPFR are nonassociative as well.

4.2. MCS-Based Topology Design

In addition to relying on associative and quasi-associative rules, there is the third option to design a fusion topology and its fusion process based on the characteristics of the information items themselves. In this case, the possibility distributions of sources are analysed, which guides the design towards desired effects. In a sense, the information provided by the multi-source system dictates the topology.

One approach to do so is to build upon the MCS fusion rule (6). It itself is not quasi-associative, and thus information items cannot be freely assigned to fusion nodes. However, by carefully searching for all the most consistent subsets, fusion can be distributed in a way that each fusion node produces the most specific intermediate result from agreeing sources, thus, emphasizing the consensus of this agreeing subset. In such a two-layer topology, all $I \in \mathbf{I}_{(k)}^{MCS-\alpha}$ are fused in separate fusion nodes $fn_{(k)}$ using, at the first level, a mix of renormalised conjunctive minimum fusion and maximum fusion:

$$\pi_{(k)}(\mu) = \begin{cases} \frac{\min_i \pi_i(\mu)}{h_i(\pi_i(\mu))} & \text{if } h_i(\pi_i(\mu)) > 0 \\ \max_i \pi_i(\mu) & \text{if } h_i(\pi_i(\mu)) = 0 \end{cases} \quad (15)$$

with i indexing $I_i \in \mathbf{I}_{(k)}^{MCS-\alpha}$. At the second level, all intermediate results are fused disjunctively using the maximum operator. An exemplary fusion topology based on the MCS fusion rule is shown in Figure 4.

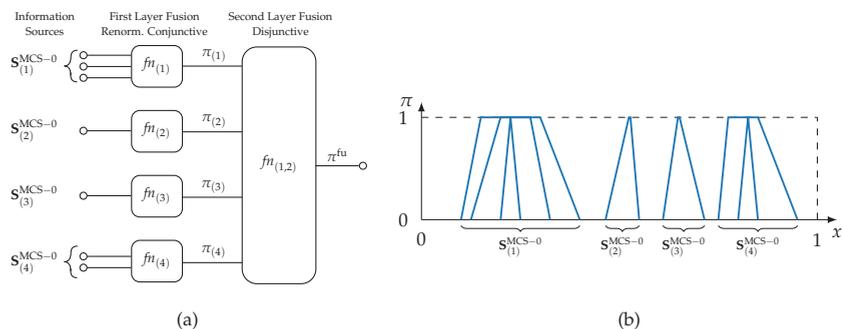


Figure 4. An example of a MCS-based fusion topology. Depicted are seven information sources fused in a two-layer topology. On the left side (a), the topology itself is shown with minimum fusion on the first layer and maximum fusion on the second layer. The right side (b) illustrates the associated possibility distributions from which the topology is constructed.

As MCS fusion analyses the consistency of information items, the inferred topology needs to be adapted for each new set of items. This is, particularly in a technical system, often not practical or feasible. Think, for example, of a technical multi-sensor system in which sensors give updated measurements in periodic time increments. In this case, the advantages of distributed fusion—such as the distribution of computing load into local nodes or lower communication loads by condensing information—are negated by the reorganisation with each measurement. Finding the MCS requires having all information items at hand in one central node rendering the distribution of the fusion process pointless. Therefore, topology design based on MCS fusion is only beneficial, if knowledge about the sources’ expected behaviour regarding consistency exists a priori. In other words, if it

is known that sources produce consistent items continually, then they are assigned to a fusion node without the need for an update with each new instance or measurement. This knowledge can be derived or learned from representative training data. Conclusions about the sources' consistency in the training data are used to build up the MCS fusion topology.

Let $\mathbf{S}_{(k)}^{\text{MCS}-\alpha}$ be a set of information sources that are assigned to fusion node $fn_{(k)}$. Furthermore, let $j = \{1, \dots, m\}$ be indices of training data, $I_{(k),j}$ be an information item produced by source $S_{(k)}$ at instance j , and $\mathbf{I}_{(k),j}$ be all information items of $\mathbf{S}_{(k)}^{\text{MCS}-\alpha}$ at instance j , then

$$S_{(k)} \in \mathbf{S}_{(k)}^{\text{MCS}-\alpha} \text{ if } \begin{cases} \forall j = \{1, \dots, m\} : h(I_{(k),j}, \mathbf{I}_{(k),j}) \geq \alpha & \text{and if } \alpha \in (0, 1], \\ \forall j = \{1, \dots, m\} : h(I_{(k),j}, \mathbf{I}_{(k),j}) > 0 & \text{and if } \alpha = 0, \end{cases} \quad (16)$$

i.e., a source $S_{(k)}$ belongs to $\mathbf{S}_{(k)}^{\text{MCS}-\alpha}$ if all its information items are consistent with the items of $\mathbf{S}_{(k)}^{\text{MCS}-\alpha}$ at least to a degree of α .

MCS-based fusion nodes are then created by Algorithm 1, which is based on the algorithm provided for finding MCS [58,61]. Algorithm 1 starts with \mathbf{S} and searches all MCS for the first data instance ($j = 1$). The found MCS are stored and themselves searched for new MCS for the next data instance and so forth.

Algorithm 1: Fast algorithm for finding subsets of information sources, which are consistent at least to degree α on every instance of training data. Each subset $\mathbf{S}_{(k)}^{\text{MCS}-\alpha}$ is assigned to fusion node $fn_{(k)}$. The algorithm relies on finding MCS of information items as defined by Dubois et al. [58,61].

Input: A set of information sources \mathbf{S} , alpha-cut-level α

Output: Set of sets \mathcal{S}_h with fusion node set $\mathbf{S}_{(k)}^{\text{MCS}-\alpha} \in \mathcal{S}_h$

$m \leftarrow$ number of training data instances;

if $j = 1$ **then**

$\mathcal{S}_h \leftarrow \{\mathbf{S}\};$

end

for $j \leftarrow 1$ **to** m **do**

$\mathcal{S} \leftarrow \{\};$

foreach $\mathbf{S}' \in \mathcal{S}_h$ **do**

$\mathbf{S}' = \text{findMCS}(\mathbf{I}_j, \alpha);$

 ; /* findMCS() as defined by Dubois et al. [58,61] */

 ;

$\mathcal{S} = \{\mathcal{S} \cup \mathbf{S}'\};$ /* \mathbf{I}_j provided by \mathbf{S}' */

end

$\mathcal{S}_h \leftarrow \mathcal{S};$

end

For the following computations, the minimum consistency in each group is stored as a reference value:

$$\alpha_{(k)}^r = \min_j h(I_{(k),j}). \quad (17)$$

In an MCS fusion topology, which is learned from training data rather than updated each j , it is not guaranteed that, for new data instances, intermediate results $I_{(k,1)}^{\text{fu}}$ are disjoint. As of this, the maximum fusion rule of the final layer as described previously is replaced with (15). This means that, in the case that the topology is learned using Algorithm 1, all fusion nodes use the same fusion rule.

Regarding parameter α , the following observation leads to maximal specific fusion results at the first layer. If $\forall j$ cores of the possibility distributions are disjoint, then fusion

with MCS-1 is equal to maximum fusion [6]. Therefore, MCS-1 fusion demands continuous mutual consistency. In contrast, MCS-0 results in minimum fusion if $\forall j$ the supports overlap and is less restrictive.

Proposition 4. MCS fusion as outlined in (15) results in the maximal specific information items if Algorithm 1 is executed with $\alpha = 0$.

Proof. With decreasing α , the condition for grouping items into fusion nodes becomes less strict—as can be seen in (16). Thus, fusion node sizes increase with decreasing α . It follows that the maximum node sizes are achieved if $\alpha = 0$. The more information items belong to a node, the more alternatives for the unknown true value are eliminated by the minimum operator in (6). Consequently, the integral $\int_{x_a}^{x_b} \pi(x) dx$ inside the specificity measure (A2) becomes minimal if $\alpha = 0$, and therefore specificity (A2) itself becomes maximal. \square

Consequently, we propose the design of MCS fusion by using $\alpha = 0$ to achieve maximal node sizes and maximal specific fusion results.

The approach presented in (16) and Alg. 1 allows the transfer of the MCS-fusion rule (6) to distributed fusion topologies. This is an alternative to designing topologies based on (quasi-)associative fusion rules, which are rare in a possibilistic setting. An MCS-based topology is aimed at producing maximal specific and precise fusion subresults. However, distributed MCS-fusion lacks robustness in the case of nonrepresentative training data or defective sources, which is detailed in the next section.

4.3. Robustness

The MCS fusion topology based on consistencies in the historic training data is prone to unexpected inconsistencies in information items. Due to the minimum operator used in the first level fusion nodes (see (15) and Figure 4), intermediate fusion results are altered significantly if items are less consistent than they are expected to be, that is $h \leq \alpha^f$. Even in large groups of sources, a single information source producing an unexpectedly inconsistent item may change the outcome significantly. An example of such an occurrence inside a fusion node using $\alpha^f = 1$ is given in Figure 5.

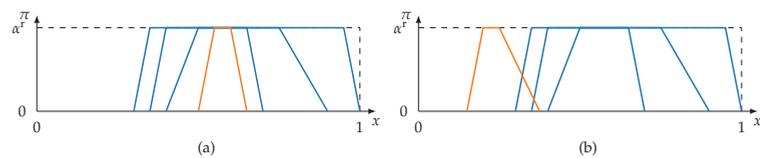


Figure 5. Information items of a fusion node with consistency level $\alpha^f = 1$. Left plot (a) shows possibility distributions with expected consistent behaviour. In the right side plot (b), a single defective information item with unexpected behaviour (marked in red) causes $h(I) < \alpha^f$. Fusion with (15) results in dissimilar possibility distributions.

Unexpected inconsistent behaviour of reliable sources occurs in two situations.

- First, incomplete information and epistemic uncertainty in the training data may lead to assessing a group of sources as consistent prematurely. Information sources may produce different (in)consistent behaviours depending on the training data’s true value and its position on the frame of discernment. Take, for example, a condition-monitoring scenario of a technical system in which sensors state the condition on a discrete frame of discernment $X = \{error1, error2, normal\}$. Two sensors may both detect two of the conditions (e.g., *error1, normal*); however, only one is able to detect the third condition (*error2*). If training data does not include data regarding *error2*, then with Algorithm 1, both sensors are falsely identified as consistent and grouped into a fusion node. If *error2* occurs later, then the sensors behave unexpectedly inconsistently. This problem relates to *spurious correlations* in probability theory [70], which describes

that, in large datasets, it is particularly likely that correlations are found between variables incorrectly.

- Second, defective sources are a cause of unexpected inconsistent behaviour. Defective sources are sources that are trustworthy and therefore have a high reliability but nonetheless start to supply incorrect information [71]. Source defects appear in different forms: Information can change suddenly, drift continuously or incrementally, or can be characterised by an increasing number of outliers [72,73]. Countermeasures are majority-guided fusion rules as applied by Ehlenbröcker et al. and Holst and Lohweg [21,23]. This requires redundant and reliable sources in a fusion node.

In the following, we propose three adaptations to the distributed MCS-based fusion topology. These adaptations aim to increase the robustness of the topology in the case of incomplete training data and defective sources.

- **Redundancy-Driven Topology Design:** To counteract non-representative training data, it must be ensured that information sources are not prematurely deemed to be consistent. For this, it must be analysed whether the consistent behaviour between sources extends over the entire frame of discernment. Therefore, instead of the consistency metric used in (16), the redundancy metric originally proposed in previous works [38,39] is adopted, which ensures that the complete frame of discernment is considered.
- **Discounting Defective Sources:** Grouping the information sources by consistency (or redundancy) eases the detection of defects [23,24]. Items detected as defective are discounted in the fusion node so that they have less influence on the output of the node. This requires an adjustment of the fusion rule (previously minimum or maximum operator) in the nodes. This defect detection step explicitly exploits the distributed topology to its advantage. This deliberately dismisses the associativity of the overall fusion.
- **Estimation-fusion-based Nodes:** Averaging information is a natural way to favour opinions of the majority. Adopting estimation fusion in nodes results in more robust behaviour against defects—such as outliers—compared to purely conjunctive fusion as applied in (6).

4.3.1. Redundancy-Driven Topology Design

In previous work [39], a redundancy metric was proposed that introduces the notion of range of a set of possibility distributions.

Definition 8 (Range [39]). *Given a frame of discernment $X = [x_a, x_b]$, the range of a set of possibility distributions \mathbf{p} quantifies how far \mathbf{p} stretches over X . Let $\mathcal{P}(\mathbf{p})$ be the power set of all possible \mathbf{p} , then the range is described by a monotonic increasing function $rge : \mathcal{P}(\mathbf{p}) \rightarrow [0, 1]$ with the following properties:*

- **Upper bound:** *If $rge(\mathbf{p}) = 1$, then $\exists \pi \in \mathbf{p} : \pi(x_a) = 1$ and $\exists \pi \in \mathbf{p} : \pi(x_b) = 1$.*
- **Lower bound:** *$rge(\mathbf{p}) = 0$ if $\forall \pi, \pi' \in \mathbf{p} : \pi = \pi'$, i.e., all possibility distributions $\pi \in \mathbf{p}$ are identical.*

The range determines whether a set of possibility distributions covers X . Together with the consistency measurement applied in (16), rge is adopted into the topology design approach. Consistency and range are balanced against each other, which results in a dual redundancy metric:

Definition 9 (Possibilistic Redundancy Metric [39]). *Let $\mathbf{S} = \{S_1, S_2, \dots, S_n\}$, i.e., a set of information sources, and $\mathcal{P}(\mathbf{S})$ be all possible combinations of sources, then a possibilistic redundancy metric ρ is a function that maps $\mathcal{P}(\mathbf{S})$ to the unit interval: $\rho : \mathcal{P}(\mathbf{S}) \rightarrow [0, 1]$. Information sources are only redundant if their information items both (i) are redundant themselves and (ii) cover the frame of discernment, i.e., have a high range (Definition 8). In accordance*

with [39], the redundancy of information items is determined via possibilistic similarity measures. Consistency (2) satisfies the requirements to serve as a similarity measure [32].

In this context and to qualify as an intuitively meaningful metric, the following requirements have to be met:

- **Boundaries:** A redundancy metric should be able to model complete redundancy and complete non-redundancy. It follows that ρ is minimally and maximally bounded. It is proposed that $\rho \in [0, 1]$.
- **Identity relation:** An information source is fully redundant with identical copies of itself: $\rho(S, S, \dots, S) = 1$. Note that sources can be redundant without necessarily being identical.
- **Symmetry:** The metric ρ is a symmetric function in all its arguments, i.e.,

$$\rho(S_1, S_2, \dots, S_n) = \rho(S_{p(1)}, S_{p(2)}, \dots, S_{p(n)})$$

for any permutation p on $\mathbb{N}_{>0}$.

The following relations between redundancy of information items and sources hold.

- If information sources are redundant, then they provide redundant information items. Consequently, $\rho(\mathbf{S})$ increases as the redundancy of information items increase.
- Redundant information items do not necessitate that their information sources are also redundant. Due to cases of incomplete information, redundant information items may be a case of spurious redundancy (similar to spurious correlation).

To capture the idea of a dual metric, ρ is designed to be a function of two pieces of evidence. The evidence against redundancy $e_c : \mathcal{P}(\mathbf{S}) \rightarrow [0, 1]$. As long as information items are redundant, $e_c(\mathbf{S}) = 0$. Determining the redundancy of information items is both based on the similarity of possibility distributions and related to the notion of possibilistic dependency. An overview of possibilistic redundancy measures for information items is provided by Holst and Lohweg [39]. Dependency measures are reviewed by Dubois et al. [74].

Evidence in favour of redundancy $e_p : \mathcal{P}(\mathbf{S}) \rightarrow [0, 1]$ quantifies the amount of epistemic uncertainty in training data. It incorporates the range of information. It indicates to what degree information is available from the complete frame of discernment. A set of information sources is only redundant if $e_p(\mathbf{S}) > 0$ and $e_c(\mathbf{S}) < 1$. The smaller value of e_p and $(1 - e_c)$ dominates the redundancy metric. In previous work [39], the geometric mean is proposed as an averaging function for e_p and e_c as follows:

$$\rho(\mathbf{S}) = \rho(e_c(\mathbf{S}), e_p(\mathbf{S})) = \sqrt{e_p(\mathbf{S}) \cdot (1 - e_c(\mathbf{S}))}. \tag{18}$$

Let the consistency measure h (2) determine the redundancy between information items and let I_j be the set of information items available at instance j , then

$$e_c(\mathbf{S}) = 1 - \operatorname{avg}_{j = \{1, \dots, m\}} (h(I_j)), \tag{19}$$

i.e., e_c averages consistencies available from training data with an averaging operator (see Definition 3). Designing MCS-based topologies (16) is based on the notion that the consistency is above a certain α for all instances. To keep this notion for the redundancy-based design, the minimum operator is used as averaging operator in (19).

The evidence e_p is computed based on the range as follows:

$$e_p(\mathbf{S}) = \frac{\operatorname{rge}(\mathbf{S}) - x_a}{x_b - x_a}. \tag{20}$$

The range itself is dependent on the position of possibility distributions on the frame of discernment, which is determined by their center of gravity [2]

$$pos(\pi) = \begin{cases} x & \text{if } \pi(x) = 1 \text{ and } \forall x' \in \{X \setminus x\} : \pi(x') = 0, \\ \frac{\int_{x_a}^{x_b} x \cdot \pi(x) dx}{\int_{x_a}^{x_b} \pi(x) dx} & \text{otherwise.} \end{cases} \quad (21)$$

The position of a set of possibility distributions \mathbf{p} is obtained by prior disjunctive fusion (5), i.e.,

$$pos(\mathbf{p}) = pos(fu(\mathbf{p})).$$

Given a set of information sources $\mathbf{S} = \{S_1, S_2, \dots, S_n\}$ providing information items $\mathbf{I}_j = \mathbf{p}_j = \{\pi_{1,j}, \pi_{2,j}, \dots, \pi_{n,j}\}$, then

$$rge(\mathbf{S}) = \max_{j,j' \in \{1, \dots, m\}} (|pos(\mathbf{p}_j) - pos(\mathbf{p}_{j'})|) = \max_{j \in \{1, \dots, m\}} (pos(\mathbf{p}_j)) - \min_{j \in \{1, \dots, m\}} (pos(\mathbf{p}_j)). \quad (22)$$

At least one pair $\mathbf{p}_j, \mathbf{p}_{j'}$ of information item sets needs to range over the frame of discernment X in order to provide evidence for a redundant behaviour, i.e., $e_p(\mathbf{S}) > 0$ if $\exists j : rge(\mathbf{p}_j) > 0$.

The redundancy metric ρ (18) is used as a decision criterion to find suitable sets of information sources $\mathbf{S}_{(k)}^\rho$ to be fused in fusion nodes $fn_{(k)}$. Algorithm 2 describes a simple approach that searches all subsets of consistency-based fusion nodes in \mathcal{S}_h (found by Algorithm 1). A set of sources is only assigned to a fusion node if $\rho \geq \eta$.

Algorithm 2: Algorithm that searches for redundancy-based fusion nodes based on \mathcal{S}_h found by Algorithm 1. The algorithm iterates over \mathcal{S}_h and searches all $\mathbf{S}' \subseteq \mathbf{S}, \mathbf{S} \in \mathcal{S}_h$ for sets meeting the redundancy criterion η .

Input: Consistency-based fusion topology found by Algorithm 1, i.e.,

$$\mathcal{S}_h = \{\mathbf{S}_{(k)}^{\text{MCS}^{-\alpha}}\}; \text{ threshold parameter } \eta$$

Output: Redundancy-based fusion topology \mathcal{S}_ρ

$\mathcal{S}_\rho \leftarrow \{\};$

$\mathcal{S}' \leftarrow \mathcal{S}_h;$

$idx \leftarrow 1;$

while $idx \leq |\mathcal{S}'|$ **do**

$\mathbf{S} \leftarrow \mathcal{S}'[idx];$

if $\rho(\mathbf{S}) \geq \eta$ **or** $|\mathbf{S}| = 1$ **then**

 /* \mathbf{S} is added to fusion topology */

if $\mathbf{S} \notin \mathcal{S}_\rho$ **then**

$\mathcal{S}_\rho.append(\mathbf{S});$

end

else

 /* create subsets of \mathbf{S} to be checked for redundancy */

foreach $S \in \mathbf{S}$ **do**

$\mathbf{S}' \leftarrow \mathbf{S} \setminus S;$

if $\mathbf{S}' \notin \mathcal{S}'$ **then**

$\mathcal{S}'.append(\mathbf{S}')$

end

end

end

$idx \leftarrow idx + 1;$

end

As motivated previously, the redundancy-based approach of Algorithm 2 results in a more robust MCS-based topology design than Algorithm 1. As (18) includes the range of information items, the effects of incomplete information and epistemic uncertainty in the training data are reduced. This leads to less detections of spurious relations.

4.3.2. Discounting Defective Sources

Information items that deviate from the expected level of consistency α^r (17) are seen as unreliable and, consequently, are discounted in each fusion node. Therefore, the degree of reliability $rel \in [0, 1]$ is determined with regard to α^r . Let \mathbf{I} be information items fused in a node and \mathbf{I}^* be the largest subset in \mathbf{I} , which has (i) $h(\mathbf{I}^*) \geq \alpha^r$ and (ii) $|\mathbf{I}^*| > 1$; then,

$$rel(I) = \begin{cases} 1 & \text{if } h(I, \mathbf{I}^*) > \alpha^r, \\ \frac{h(I, \mathbf{I}^*)}{\alpha^r} & \text{if } h(I, \mathbf{I}^*) \leq \alpha^r. \end{cases} \quad (23)$$

In the case that there is no unique \mathbf{I}^* with $h(\mathbf{I}^*) \geq \alpha^r$ and at least two elements, then all items are seen as fully reliable, and fusion needs to switch to disjunctive fusion.

Information items' possibility distributions are modified prior to fusion so that they have a lesser effect on the fusion results [4,75]. A modification function for discounting information items has to satisfy the following requirements (extended from previous work [39]).

Definition 10 (Requirements for Information Item Modification). *As modification aims at changing fusion outputs, the requirements interact with fusion rules to be applied on π :*

- *Information preservation: If $rel(I) = 1$, then the information must not be changed but instead preserved. Let π' be a modified possibility distribution based on π . If $rel(I) = 1$, then $\pi' = \pi$.*
- *Neutral element: If $rel(I) = 0$, then I needs to have no effect on the fusion. The item I needs to act as a neutral element on fusion operator fu , i.e., $fu(I, \mathbf{I}) = fu(\mathbf{I})$.*
- *Monotonicity: For increasing $rel(I)$, I needs to have a monotonic increasing effect on fu .*

Modification functions were proposed by Yager and Kelman [75]

$$\pi'(x) = rel \cdot \pi(x) + 1 - rel,$$

and Dubois and Prade [4]

$$\pi'(x) = \max_{x \in X}(\pi(x), 1 - rel).$$

Both satisfy the requirements for modification only for conjunctive fusion. A general modification function for the use with OWA operators was proposed by Larsen [68]. It is defined based on the *andness* degree $and \in [0, 1]$ of OWA fusion:

$$\pi'(x) = and + rel \cdot (\pi(x) - and). \quad (24)$$

The OWA operator results in the minimum fusion for $and = 1$ and in maximum fusion for $and = 0$. The OWA modification (24) introduces a global possibility level of and to the distribution π' . As of this, the modification satisfies the requirement of *neutral element* only if $and = 1$ or $and = 0$ but not for $0 < and < 1$.

All three modification functions raise the overall possibility level globally. As argued in previous work [39], this kind of approach towards modification functions is counterintuitive if it is considered that defective or unreliable sources may err in their estimation of the unknown value v . An unreliable source may be slightly incorrect. Raising the possibility

level globally cannot model such a situation. A modification function that widens or shrinks the possibility distribution is proposed as (adapted from previous work [39]):

$$\pi'(x) = \begin{cases} \max_{x' \in C}(\pi(x')) & \text{if minimum fusion,} \\ \min_{x' \in C}(\pi(x')) & \text{if maximum fusion,} \end{cases} \tag{25}$$

$$C = [x - (1 - rel)^\beta \cdot (x_b - x_a), x + (1 - rel)^\beta \cdot (x_b - x_a)], \text{ and}$$

$$X = [x_a, x_b].$$

This modification considers both minimum and maximum fusion as they occur in the MCS-based fusion topology but does not approach a global modification. The reliability rel and the control parameter $\beta \in \mathbb{R}_{\geq 1}$ define a vicinity around x . The new possibility $\pi'(x)$ is taken from this vicinity. This creates a widening or shrinking effect, respectively. The parameter β allows to control the size of the vicinity and, thus, the extent to which rel alters $\pi(x)$. The larger β is, the less effect rel has on $\pi(x)$. If $rel > 0$ and $\beta \rightarrow \infty$, then (25) has no widening or shrinking effect.

4.3.3. Estimation-Based Fusion Nodes

The third adaptation to increase the robustness of the proposed MCS-based fusion topology is to replace fusion in the first layer (15) with estimation fusion (13). In this way, defective sources have a lesser impact on the fusion result of a node.

Associativity needs to hold for first layer fusion nodes (see Figure 4) if multi-level fusion is to be achieved (splitting fusion nodes into smaller ones). Estimation fusion is only associative if G is associative and monotonic increasing and F is associative. In the proposed estimation-based fusion nodes, G is the minimum operator that satisfies associativity and monotonicity. The function F is defined to be an averaging operator, which is rarely associative, e.g., the arithmetic mean. Multi-level distributed fusion can still be achieved by using a fusion node’s ability to output auxiliary information (see Definition 4).

If a node outputs the number of information items that contributed to its fusion result as a weight w , then a weighted arithmetic mean operator of the form

$$F^{WAM}(x_1, \dots, x_n) = \frac{\sum_{i=1}^n w_i \cdot x_i}{\sum_{i=1}^n w_i}$$

results in associative fusion. In the following, we refer back to the notation of fusion nodes as defined in Definition 4, i.e., $\mathbf{I}_{k,l}$ denotes the set of information items that serve as input to fusion node $fn_{(k,l)}$. To achieve associativity, a weight $w_{(k,l)}$ is assigned to the output of $fn_{(k,l)}$, which is defined as

$$w_{(k,l)} = \sum_{\mathbf{I}_{(o,p)} \in \mathbf{I}_{(k,l)}} w_{(o,p)} \text{ with}$$

$$w_{(k,1)} = |\mathbf{I}_{(k,1)}|.$$

The distributed weighted average function

$$F_{(k,l)}^{WAM}(x_1, \dots, x_n) = \sum_{\mathbf{I}_{(o,p)} \in \mathbf{I}_{(k,l)}} \frac{1}{w_{(o,p)}} \sum_{\mathbf{I}_{(o,p)} \in \mathbf{I}_{(k,l)}} w_{(o,p)} \cdot F_{(o,p)}^{WAM}(x_1, \dots, x_n) \tag{26}$$

$$\text{with } F_{(k,1)}^{WAM}(x_1, \dots, x_n) = \frac{1}{w_{(k,1)}} \sum_{i=1}^{|\mathbf{I}_{(k,1)}|} x_i$$

allows splitting nodes without changing the fusion result. An overview of a distributed fusion topology based on estimation fusion rules is given in Figure 6.

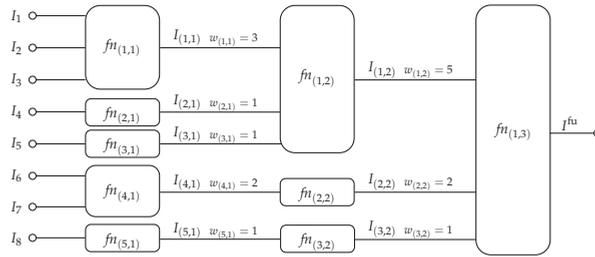


Figure 6. Example of an MCS-based fusion topology adapted with weighted estimation fusion. Previously conjunctive fusion nodes (first level fusion; see also Figure 4) are replaced with the estimation fusion rule described by (13). To preserve the associativity of the first level fusion nodes, the weighted averaging operator F^{WAM} described by (26) is applied as function F. If the estimation-based nodes are split into a multi-level topology, then F^{WAM} requires fusion nodes to communicate the number of input information items.

To keep the option of discounting defective sources, weights $w_{(k,l)}$ are modified in the case a defect is detected via (23) as follows:

$$w'_{k,l} = w_{k,l} \cdot rel(I_{(k,l)}). \tag{27}$$

If $rel(I_{(k,l)}) = 1$, then information is preserved. Otherwise, if $rel(I_{(k,l)}) = 0$ the information item is completely discounted.

4.4. Remark on Multi-Level Fusion by Splitting Nodes

The MCS-based design approach describes a two-layer fusion topology by first fusing consistent or redundant information items conjunctively and then fusing the intermediate results disjunctively. In this context, multi-layer fusion can be achieved by splitting a single fusion node into multiple smaller ones. This may be beneficial if, e.g., communication or computational loads per node need to be optimised. While this approach of splitting is feasible due to the associativity of applied fusion rules, the ability of the fusion topology to detect and discount defective sources is reduced by doing so.

Discounting information items requires finding the unique largest subset of items whose consistency is greater than α^r . If multiple sources are defective simultaneously, then—depending on the fusion node size—the largest subset may be made up by defective sources. In the worst case, the maximum number of defective sources a fusion node can handle is $\lfloor \frac{n-1}{2} \rfloor$ [24], with n being the number of sources contributing to a fusion node. As the proposed discounting approach is node-specific, the ability of a node to discount defective sources is hindered by splitting nodes. The smaller n is, the smaller is the maximum number of detectable defective sources. This has to be kept in mind in designing an MCS-based fusion topology.

5. Evaluation

The evaluation is structured into three parts in which the computational complexity, topology design approaches, and the robustness of distributed MCS fusion are focused. Distributing information fusion is motivated—as outlined in Section 1—by the assumption that computational load per distributed node is less than the load for a single centralised node. First, this assumption is examined for MCS and estimation fusion.

Subsequently, the computational complexity of design Algorithms 1 and 2 are discussed. Their performance and the effectiveness of the MCS-fusion adaptations (see Section 4.3) are then evaluated on selected real-world datasets.

5.1. Computational Complexity

The following evaluation of computational time complexity relies on the Bachmann–Landau notation $f(n) = \mathcal{O}(g(n))$, which states that a function $f(n)$ does not grow faster for $n \rightarrow \infty$ than $g(n)$. $f(n)$ is therefore asymptotically upper bounded by $g(n)$. $\mathcal{O}(g(n))$ denotes the set of all $f(n)$ such that there exist positive constants c and n_0 : $f(n) \leq c \cdot g(n), \forall n \geq n_0$ [76].

5.1.1. Fusion Rules

In the following, we evaluate whether the computational load of MCS and estimation fusion are decreased by distributing, i.e., whether each fusion node in a distributed topology has a lower load compared to a single centralised node. For MCS fusion, it is assumed that the MCS have already been found, i.e., only (6) is considered.

As (6) consists exclusively of minimum and maximum operations, centralised MCS fusion is $\mathcal{O}(n)$ with n being the number of input information sources. In a distributed two-layer fusion topology, each fusion node has $n_f \leq n$ input sources. First layer nodes operate using renormalised minimum fusion; the final layer node applies maximum fusion. Fusion in each node is therefore $\mathcal{O}(n_f)$. This simple observation shows that computational load of distributed nodes is less than in centralised fusion—for reasonable MCS fusion topologies.

For estimation fusion, the situation is not as simple. Estimation fusion, as defined in (11), (12), and (13), iterates over every n -tuple (x_1, \dots, x_n) . Thus, the computational load increases exponentially with its number of inputs n .

Proposition 5. *Let X^* be the frame of discernment with the highest cardinality in $\{X_1, \dots, X_n\}$, then the complexity of estimation fusion rule (11) is $\mathcal{O}(|X^*|^n \cdot F + |X^*|^n \cdot G + |X^*|^n)$. If G is the minimum operator and F is the arithmetic mean operator, then the complexity is $\mathcal{O}(|X^*|^n)$.*

Proof. Equation (11) is a combination of F, G , and the maximum operator. F and G need to be computed for each n -tuple (x_1, \dots, x_n) for every $x_i \in X_i$, i.e., F and G are computed $\prod_{i=1}^n |X_i|$ times. The maximum operator is computed for each $x \in X$. Its number of inputs is at worst $\prod_{i=1}^n |X_i|$. In total, the complexity of (11) is

$$\begin{aligned} & \mathcal{O}\left(\prod_{i=1}^n |X_i| \cdot F + \prod_{i=1}^n |X_i| \cdot G + \sum_{x \in X} \cdot \max\right) \\ &= \mathcal{O}\left(\prod_{i=1}^n |X_i| \cdot F + \prod_{i=1}^n |X_i| \cdot G + \sum_{x \in X} \prod_{i=1}^n |X_i|\right) \\ &= \mathcal{O}\left(\prod_{i=1}^n |X_i| \cdot F + \prod_{i=1}^n |X_i| \cdot G + \prod_{i=1}^n |X_i|\right) \end{aligned}$$

Let $X = \{X_1, \dots, X_n\}$ and $X^* = \arg \max_{X' \in X} |X'|$, then

$$= \mathcal{O}(|X^*|^n \cdot F + |X^*|^n \cdot G + |X^*|^n).$$

With G being the minimum operator and F being the arithmetic mean, this becomes

$$\begin{aligned} & \mathcal{O}(|X^*|^n \cdot n + |X^*|^n \cdot n + |X^*|^n) \\ &= \mathcal{O}(|X^*|^n \cdot n) \\ &= \mathcal{O}(|X^*|^n). \end{aligned}$$

□

Therefore, the complexity of (12) relies on the complexities of G and F ; however, it is safe to say that the growth $|X^*|^n$ leads to issues in practical implementations. Unfortunately, in this case, the lack of scalability cannot be solved by distributing the estimation fusion over several nodes.

Proposition 6. Let G be the minimum operator and F be an averaging operator as defined in (13). Assume a topology of fusion nodes using estimation fusion (13) exclusively, then fusion at the final fusion node in the last layer still grows exponentially, that is, has $\mathcal{O}(\prod_{i=1}^n |X_i|)$ or $\mathcal{O}(|X^*|^n)$, respectively.

Proof. Looking at a single fusion node with n_k inputs, F maps in worst case each tuple (x_1, \dots, x_{n_k}) to a unique point x . Then, the size of the output’s frame of discernment is $\prod_{i=1}^{n_k} |X_i|$. Let $f_{n(k,l)}$ be fusion nodes arranged in a topology so that the fusion topology outputs a single information item, i.e., there is a final fusion node $f_{n(1,L)}$, $L \in \mathbb{N}_+$. Assume all n available information items are input into a fusion node exactly once. Then, the final node has to process $2 \leq n_{\text{final}} \leq n$ input information items. The number of tuples to iterate is then $\prod_{k=1}^{n_{\text{final}}} |X_{k,L-1}|$. In a two-layer topology, $\prod_{k=1}^{n_{\text{final}}} |X_{k,L-1}| = \prod_{k=1}^{n_{\text{final}}} \prod_{i=1}^{n_k} |X_k|$. As $n_{\text{final}} = \sum_{k=1}^{\max(k)} n_k$, this is $\prod_{i=1}^n |X_i| \leq |X^*|^n$. Thus, fusion at the final node has $\mathcal{O}(|X^*|^n)$. \square

For estimation fusion, the number of elements in the frame of discernment grows with each fusion node. The final fusion node has to process in worst case $|X^*|^n$ tuples, which is the same for centralised fusion.

Yager demonstrated [65] that, if all π_i are convex and if X contains only real-valued ordered elements, then (13) (that is $G = \min$ and F is an averaging operator) can also be computed via the crisp-set α -cuts

$$A_i^\alpha = \{x \in X_i : \pi_i(x) \geq \alpha\} \text{ with } \alpha \in (0, 1]. \tag{28}$$

Definition 11. A possibility distribution π is said to be convex iff (1) each of its α -cuts A^α are a single closed interval, i.e., $A^\alpha = [a, b]$ and (2) all A^α are nested, i.e., $\forall \alpha_1 > \alpha_2 : A^{\alpha_1} \subseteq A^{\alpha_2}$.

For each α -level the crisp sets A_i^α are fused using the averaging operator F , which results in

$$\begin{aligned} A^{\text{fu}-\alpha} &= F(A_1^\alpha, \dots, A_n^\alpha) \\ &= \left[F\left(\min_{x \in A_1^\alpha} x, \dots, \min_{x \in A_n^\alpha} x\right), F\left(\max_{x \in A_1^\alpha} x, \dots, \max_{x \in A_n^\alpha} x\right) \right]. \end{aligned} \tag{29}$$

The fused possibility distribution is then obtained by taking the maximum α -level as follows:

$$\pi^{(\text{fu})}(x) = \max_{\alpha} \begin{cases} \alpha & \text{if } x \in A^{\text{fu}-\alpha}, \\ 0 & \text{if } x \notin A^{\text{fu}-\alpha}. \end{cases} \tag{30}$$

Proposition 7. The computational load of (28)–(30) grows linearly in number of input possibility distributions n , number of elements in X^* , and number of α -levels n_α , i.e., (28)–(30) have in total $\mathcal{O}(n \cdot |X^*| \cdot n_\alpha)$.

Proof. Equation (28) grows linearly in $|X_i|$. It has to be for each α -level and each input possibility distribution, i.e., (28) is $\mathcal{O}(n \cdot |X^*| \cdot n_\alpha)$.

For (29), both minimum and maximum have to be computed n times, F has to be computed two times. This has to be performed for each α -level. This results in

$$\begin{aligned} &\mathcal{O}(n_\alpha \cdot 2 \cdot (F + n \cdot \min + n \cdot \max)) \\ &= \mathcal{O}(n_\alpha \cdot (n + n \cdot |X^*| + n \cdot |X^*|)) \\ &= \mathcal{O}(n_\alpha \cdot n \cdot |X^*|). \end{aligned}$$

Equation (28) is a single maximum with n_α inputs, i.e., it is $\mathcal{O}(n_\alpha)$.

In total, (28)–(30) is $\mathcal{O}(n_\alpha \cdot n \cdot |X^*|)$. \square

In contrast to (13), the computational load is distributed over fusion nodes if (28)–(30) are distributed. Using α -cuts, neither $|X|$ nor n_α grow with each fusion node. Rather, they stay constant. Consequently, increasing the number of fusion nodes in a topology—which decreases the number of inputs per fusion node—reduces the computational load per node. In conclusion, both estimation fusion as well as MCS fusion profit from reduced computational load per node if fusion is distributed.

5.1.2. Fusion Topology Algorithms

Using (16) naively to search all possible subsets of a set of information sources \mathbf{S} for fusion nodes is computationally demanding. Such an approach grows exponentially in number of sources n . The proposed Algorithm 1 presents a computational faster approach.

Proposition 8. *The Algorithm 1 for finding consistency-based fusion nodes has complexity $\mathcal{O}(m \cdot n^2)$ with $n = |\mathbf{S}|$ and m being the number of training data instances.*

Proof. Algorithm 1 iterates over all training data instances j . For $j = 1$, it searches \mathbf{S} for all MCS. As the algorithm of [58,61] grows linearly in n , this step is $\mathcal{O}(n)$. For each subsequent iteration with $j > 1$, it searches all previously MCS found at $j - 1$ again for MCS. The maximum number of found MCS is n . The maximum number of sources belonging to an MCS is also n , i.e., each iteration at $j > 1$ grows with n^2 . Consequently, Algorithm 1 is $\mathcal{O}(m \cdot n^2)$. \square

The redundancy-based Algorithm 2 takes the fusion nodes found by Algorithm 1 as input. If an MCS does not meet the redundancy criterion, then Algorithm 2 searches within each MCS for largest subsets with $\rho \geq \eta$.

Proposition 9. *The Algorithm 2 for finding redundancy-based fusion nodes has complexity $\mathcal{O}(2^n)$ with $n = |\mathbf{S}|$.*

Proof. Algorithm 2 searches the power set of each MCS $\mathbf{S}_{(k)}^{\text{MCS}-\alpha}$. As the maximum number of sources in $\mathbf{S}_{(k)}^{\text{MCS}-\alpha}$ is n , Algorithm 2 is $\mathcal{O}(2^n)$. \square

In contrast to the consistency-based algorithm, the redundancy-based version scales in its current implementation poorly with number of sources. For reasons of practical implementation, this needs to be addressed in future works. In this regard, plausibility checks are promising as to whether subsets of $\mathbf{S}_{(k)}^{\text{MCS}-\alpha}$ can actually exhibit the required range. In such cases, it would not make sense to search these subsets at all, saving computational time.

5.2. Robustness

Fusion using the default MCS-based topology is prone to unexpected behaviour of information sources regarding their consistency (see Section 4.2). In the following, the MCS fusion design approach and topology are evaluated on selected real-world datasets regarding their robustness. First, consistency-based design is compared to the redundancy-based design approach. Following this, the adaptations of discounting and estimation fusion are evaluated. Implementation and data preprocessing are detailed to increase reproducibility.

5.2.1. Data Preprocessing

Several data preprocessing steps are performed before the implementation. These are necessary (i) to homogenise a heterogeneous frame of discernments, (ii) to reduce the effects of noise (aleatoric uncertainty) on the fusion results and topology design, and (iii) if data are not available as possibility distributions but rather as singular values or probability distributions. Preprocessing comprises the three following steps.

- If data are singular values or probability distributions, they are transferred into possibility distributions first. For this step, singular values x are interpreted as probability

distributions with $p(x) = 1$ and $x' \in X \setminus \{x\} : p(x') = 0$. Transformation is conducted by the truncated triangular probability-possibility transformation [49,77,78] resulting in $\pi(x)$.

- Second, sources providing noisy data are regarded as partially unreliable. Their possibility distribution are modified using (25) accordingly. Unreliability values for information sources are determined heuristically.
- Third, modified possibility distributions $\pi(x)$ are mapped to a common, shared frame of discernment. This X is based on fuzzy memberships μ , i.e., $X = [\mu_a, \mu_b]$. This requires a fuzzy class to be defined to which $\mu(x)$ indicates the degree of membership of x . The class membership function $\mu(x)$ can either be provided by an expert or trained automatically [18,38,39]. Here, $\mu(x)$ is trained by a parametric unimodal potential function proposed by Lohweg et al. [79]:

$$\mu(x) = \begin{cases} 2^{-d(x, \mathbf{p}_l)} & \text{if } x \leq \bar{x}, \\ 2^{-d(x, \mathbf{p}_r)} & \text{if } x > \bar{x}, \end{cases} \tag{31}$$

$$\text{with } d(x, \mathbf{p}_l) = \left(\frac{|x - \bar{x}|}{C_l} \right)^{D_l},$$

$$d(x, \mathbf{p}_r) = \left(\frac{|x - \bar{x}|}{C_r} \right)^{D_r}, \text{ and}$$

with \bar{x} being the arithmetic mean of given training data \mathbf{x} . The parameters are determined as follows: $C_l = \bar{x} - \min_{j \in \{1, 2, \dots, m\}}(x_j)$, $C_r = \max_{j \in \{1, 2, \dots, m\}}(x_j) - \bar{x}$, and $D_l, D_r \in \mathbb{N}_{>1}$. D_l and D_r are often determined empirically [21,80]. A training routine for D_l and D_r based on density estimations is given by Mönks et al. [81].

The possibility distribution $\pi(x)$ is then mapped to $\pi(\mu)$ via the extension principle as follows:

$$\pi(\mu) = \max_{x \in X: \mu(x) = \mu} \pi(x).$$

A detailed description and visualisations of these preprocessing steps are given previous work [39]. Together, the preprocessing steps allow to apply the proposed design algorithms even on heterogeneous, noisy, and nonpossibilistic data. Robustness against noise can additionally be increased by data filtering. However, since parameters of (31) rely on minimum and maximum values of training data, applying filter directly on training data \mathbf{x} would distort the borders of the unimodal potential function. For this reason, memberships $\mu(x)$ —instead of data—are filtered in the preprocessing.

5.2.2. Nonrepresentative Training Data

The effects of nonrepresentative training data on consistency-based MCS topology design and on redundancy-based design are evaluated. Consistency-based topology is obtained by Algorithm 1 with parameter $\alpha = 0$ as argued in Proposition 4. Its redundancy-based counterpart is obtained by Algorithm 2. To ensure highly redundant information sources in fusion nodes, parameter η is set to 0.6, i.e., sources are added to a fusion node if their redundancy is greater than or equal to η .

Both design approaches are applied to the Sensorless Drive Diagnosis (SDD) dataset [82,83]—a multi-class classification dataset (The SDD dataset is available for download at the University of California Machine Learning Repository [84]). Nonrepresentative training data are simulated by withholding data of certain classes from the design algorithms creating a situation of epistemic uncertainty.

For the creation of the Sensorless Drive Diagnostics data set, an electromechanical drive was monitored to detect faulty system behaviour. Data comprise features obtained from phase-related motor currents and voltages. Each feature serves as an information source in this evaluation. The dataset is particularly interesting because (i) it contains highly noisy data and (ii) data are often linearly or non-linearly correlated and thus potentially

redundant. The SDD dataset contains 11 classes in total, of which class 1 represents healthy system behaviour (henceforth referred to as the normal condition). All other classes represent various fault states, such as gear or bearing damage.

The design algorithms are executed on two subsets of the dataset. First, only data belonging to the normal condition build a reduced training dataset. This reduced set manifests epistemic uncertainty. It is nonrepresentative with regard to the complete behaviour of information sources. For comparison, the second subset is constructed to include all data, i.e., the complete dataset serves as training data.

Regarding the preprocessing steps, the unimodal potential function (31) is trained on the normal condition with parameters $D_l = 2$ and $D_r = 2$. To regard the noise in the dataset, possibility distributions are modified with reliability parameters $\forall S \in \mathbf{S} : rel(S) = 0.9$ and $\beta = 1$. Additionally, memberships are smoothed with a moving average filter using a window size of 5. As the SDD dataset provides data as singular values, the preprocessing steps result in rectangular possibility distributions.

The following behaviour is expected from the topology design approaches, which helps in verifying their output:

- For the consistency-based approach, fusion nodes trained on complete data are expected to be smaller or of equal size compared with nodes trained on reduced data. More specifically, $\forall k, \exists k' : \mathbf{S}_{(k),reduced}^{MCS-\alpha} \subseteq \mathbf{S}_{(k'),complete}^{MCS-\alpha}$ because (16) requires consistency of all data instances to be above the threshold α .
- Sources grouped by the redundancy-based approach $\mathbf{S}_{(k)}^\rho$ are expected to always be a subset of at least one consistency-based found group $\mathbf{S}_{(k)}^{MCS-\alpha}$, i.e., $\forall k, \exists k' : \mathbf{S}_{(k)}^\rho \subseteq \mathbf{S}_{(k')}^{MCS-\alpha}$ because the redundancy metric (18) is more restrictive than pure consistency. The additional range information (22) prevents sources being added to a fusion node when it is not known that they behave consistently over the complete frame of discernment.

The results of Algorithms 1 and 2 are shown in Tables 2 and 3, respectively. Both tables show found fusion nodes for the first layer of the two-layer fusion topology. Fusion nodes are shown both for reduced and complete training data along with redundancy ρ (18), range evidence e_p (20), and inconsistency evidence e_c (19).

The results of Table 2 show that the MCS-based topology meets the expectation regarding fusion node sizes. Furthermore, each set $\mathbf{S}_{(k),complete}^{MCS-\alpha}$ is a subset of at least one $\mathbf{S}_{(k),reduced}^{MCS-\alpha}$, e.g., $\mathbf{S}_{(1),complete}^{MCS-\alpha} \subset \mathbf{S}_{(7),reduced}^{MCS-\alpha}$. It is also notable that—especially but not exclusively on reduced data—some sources occur in many fusion nodes.

This relates, for example, to sources 25 and 37. Sources with little informative value are likely to be consistent with other sources because such sources provide possibility distributions, which are likely wide or even close to total ignorance. For sources 25 and 37, it is the case that both provide large possibility distributions covering a significant part of the frame of discernment. Lastly, no fusion node based on complete data is exactly similar to fusion nodes based on reduced data (which is different in the following redundancy-based approach). Fusion nodes differ significantly. This means that nonrepresentative data limits the performance of the consistent-based approach substantially, i.e., because epistemic uncertainty is not considered by Algorithm 1 fusion nodes are inflated with spuriously consistent information sources.

The results of the redundancy-based approach (Table 3) also meet the expectations formulated beforehand, i.e., $\forall k, \exists k' : \mathbf{S}_{(k)}^\rho \subseteq \mathbf{S}_{(k')}^{MCS-\alpha}$. In contrast to the consistency-based approach, sources with little informative value are not part of fusion nodes (e.g., sources 25 and 37). The computation of the range (22) penalises wide possibility distributions. This is because of the disjunctive fusion prior to computing the position of a set of distributions (21). Sets including information items close to total ignorance are given a position close to 0.5 resulting in low range values and hence low redundancies.

Similar to the consistency approach, the amount of fusion nodes decreases from reduced to complete training data. This shows that the redundancy-based approach is not able to rule out all sets showing spurious redundancy. However, the majority of nodes learned on complete data are exactly similar to nodes on reduced data. This is true for sets {10, 11, 12}, {19, 20, 21, 22, 23, 24}, {31, 32, 33}, {34, 35, 36}, and {46, 47, 48} with {7, 8, 9} coming close. This shows that the redundancy-based approach finds significant sets despite nonrepresentative training data.

Table 2. Fusion nodes and their contributing information sources as designed by Algorithm 1 with parameter $\alpha = 0$. Grouped information sources are consistent for all instances of training data (see metric e_c (19)). The left side shows fusion nodes found on reduced, highly epistemic uncertain training data, i.e., only data of the class stating normal condition were available. The right side shows nodes found on complete data. Fusion node sets on reduced training data do not meet the required redundancy threshold (i.e., $\rho \not\geq \eta$), which is due to the low range-based evidence e_p (20). Information sources are numbered as provided by the SDD dataset [82,83]. Fusion nodes with less than two information sources are omitted. In total, 24 fusion nodes were found on reduced data and 28 on complete data.

Node $fn_{(k)}$	Reduced Training Data				Complete Training Data			
	$S_{(k)}, \text{reduced}^{\text{MCS}-\alpha}, \alpha = 0$	ρ	e_p	e_c	$S_{(k), \text{complete}}^{\text{MCS}-\alpha}, \alpha = 0$	ρ	e_p	e_c
$fn_{(1)}$	{1, 3, 25, 37}	0.3647	0.1330	0	{7, 8, 9}	0.9919	0.9840	0
$fn_{(2)}$	{1, 5, 25, 37}	0.4646	0.2159	0	{10, 11, 12}	0.9923	0.9847	0
$fn_{(3)}$	{1, 6, 25, 37}	0.3460	0.1197	0	{13, 37}	0.4462	0.1991	0
$fn_{(4)}$	{1, 15, 25, 28, 37, 39, 40}	0.5329	0.2839	0	{16, 28, 37, 40}	0.4043	0.1634	0
$fn_{(5)}$	{1, 25, 28, 37, 39, 40, 42}	0.5382	0.2897	0	{18, 41}	0.9513	0.9049	0
$fn_{(6)}$	{4, 28, 40}	0.3698	0.1368	0	{19, 20, 21, 22, 23, 24}	0.7830	0.6131	0
$fn_{(7)}$	{7, 8, 9, 28, 40, 41}	0.3882	0.1507	0	{25, 28, 37, 40, 41}	0.4013	0.1610	0
$fn_{(8)}$	{10, 11, 12, 25, 28, 37, 40}	0.4415	0.1950	0	{25, 28, 37, 40, 42}	0.4043	0.1634	0
$fn_{(9)}$	{13, 16, 25, 28, 37, 40}	0.3863	0.1492	0	{31, 32, 33}	0.8896	0.7914	0
$fn_{(10)}$	{14, 38, 39}	0.5450	0.2970	0	{34, 35, 36}	0.9513	0.9049	0
$fn_{(11)}$	{15, 25, 28, 37, 38, 39, 40}	0.5314	0.2824	0	{43, 44}	0.8386	0.7033	0
$fn_{(12)}$	{17, 25, 37}	0.3657	0.1337	0	{46, 47, 48}	0.7982	0.6371	0
$fn_{(13)}$	{18, 25, 28, 37, 38, 39, 40, 41}	0.3751	0.1407	0	-	-	-	-
$fn_{(14)}$	{19, 20, 21, 22, 23, 24, 25, 28, 37, 40}	0.3555	0.1264	0	-	-	-	-
$fn_{(15)}$	{25, 27, 37, 39}	0.3508	0.1231	0	-	-	-	-
$fn_{(16)}$	{25, 28, 34, 35, 36, 37, 40}	0.3747	0.1404	0	-	-	-	-
$fn_{(17)}$	{25, 28, 37, 38, 39, 40, 41, 42}	0.3234	0.1046	0	-	-	-	-
$fn_{(18)}$	{25, 28, 37, 39, 40, 41, 42, 43, 44, 45}	0.3545	0.1257	0	-	-	-	-
$fn_{(19)}$	{25, 31, 32, 33, 37}	0.3285	0.1079	0	-	-	-	-
$fn_{(20)}$	{25, 37, 46, 47, 48}	0.2907	0.0845	0	-	-	-	-

Therefore, it copes better than the consistency approach in situations with high epistemic uncertainty because the evidence e_p (20) quantifies epistemic uncertainty. Nonetheless, it is advisable to update and adapt fusion nodes and topology with newly available data. This reduces risk of nodes with spurious redundancy.

Figure 7 depicts scatter plots of selected information sources to visualise the shortcomings of the consistency-based approach and to show the effects of epistemic uncertainty. Information items may be close to each other—and therefore be consistent—for parts of the training data (see plots (a), (b), and (c)). This is indicated by the fact that the positions of items are clustered in the upper right corners for reduced training data. This does not mean that consistent behaviour carries over to complete data (which is only true for (c)).

Table 3. Fusion nodes and their contributing information sources as designed by Algorithm 2 with parameters $\alpha = 0$ and $\eta = 0.6$. Grouped information sources are consistent for all instances of training data and range over a significant part of the frame of discernment. The left side shows fusion nodes found on reduced, highly epistemic uncertain training data. The right side shows nodes found on complete data. Information sources (features) are numbered as provided by the SDD dataset [82,83]. Fusion nodes with less than two information sources are omitted. In total, 29 fusion nodes were found on reduced data and 31 on complete data.

Node	Reduced Training Data				Complete Training Data			
	$S_{(k),reduced}^{MCS-\alpha, \alpha = 0}$	ρ	e_p	e_c	$S_{(k),complete}^{MCS-\alpha, \alpha = 0}$	ρ	e_p	e_c
$fn_{(1)}$	{1, 15}	0.6239	0.3893	0	{7, 8, 9}	0.9919	0.9840	0
$fn_{(2)}$	{1, 39, 42}	0.6228	0.3879	0	{10, 11, 12}	0.8896	0.7914	0
$fn_{(3)}$	{7, 8, 9, 41}	0.6721	0.4517	0	{18, 41}	0.9513	0.9049	0
$fn_{(4)}$	{10, 11, 12}	0.6302	0.3971	0	{19, 20, 21, 22, 23, 24}	0.7830	0.6131	0
$fn_{(5)}$	{13, 16}	0.6429	0.4133	0	{31, 32, 33}	0.8896	0.7914	0
$fn_{(6)}$	{14, 38, 39}	0.6148	0.3780	0	{34, 35, 36}	0.9513	0.9049	0
$fn_{(7)}$	{19, 20, 21, 22, 23, 24}	0.6916	0.4783	0	{43, 44}	0.8386	0.7033	0
$fn_{(8)}$	{27, 39}	0.6415	0.4115	0	{46, 47, 48}	0.7981	0.6371	0
$fn_{(9)}$	{31, 32, 33}	0.6367	0.4054	0	-	-	-	-
$fn_{(10)}$	{34, 35, 36}	0.6777	0.4593	0	-	-	-	-
$fn_{(11)}$	{38, 42}	0.6077	0.3693	0	-	-	-	-
$fn_{(12)}$	{39, 41, 43, 44, 45}	0.6134	0.3763	0	-	-	-	-
$fn_{(13)}$	{39, 42, 43, 44, 45}	0.6228	0.3879	0	-	-	-	-
$fn_{(14)}$	{41, 42, 43, 44, 45}	0.6100	0.3722	0	-	-	-	-
$fn_{(15)}$	{46, 47, 48}	0.6068	0.3682	0	-	-	-	-

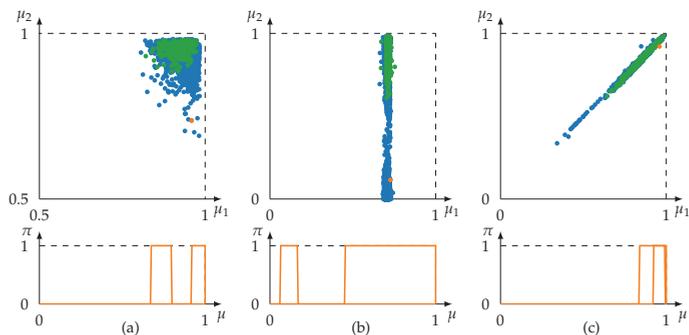


Figure 7. Information items of selected information sources belonging to reduced training data (green) and complete training data (blue). Data belongs to the Sensorless Drive Diagnosis dataset [82,83]. Subplot (a) shows information sources (features) {1, 5}, (b) {25, 10}, and (c) {43, 45}. Each point in the scatter plots represents the position or centre of gravity of a possibility distribution obtained by (21). Possibility distributions of a single pair are plotted below each scatter plot to give an intuition about the size of the distributions. In the case of reduced training data, information sources (a) {1, 5} and (b) {25, 10} belong to fusion nodes in the consistency-based approach (see Table 2) but not in the redundancy-based approach (see Table 3). Without the additional information provided by the range metric (22), the consistency-based approach considers sources, which result in being inconsistent on complete training data. Sources (c) {43, 45} are given as an example in which information items are consistent over the complete training data. Both the consistency-based as well as the redundancy-based approach consider {43, 45} in fusion nodes. Note that the scatter plot in (a) is zoomed in for better visibility.

5.2.3. Defective Sources

Regarding defective sources, two adaptations to the MCS topology were proposed in this paper. Both adaptations—(i) discounting defective sources and (ii) estimation-fusion-based nodes—were evaluated on data with purposely engineered source defects.

The Typical Sensor Defects (TSD) dataset [21] provides such defective sources (The TSD dataset is available for download at <https://zenodo.org/record/56358> (accessed on 9 March 2022)). The TSD dataset contains data of a storage container for hazardous and flammable materials measured, e.g., by temperature, smoke, and gas sensors. The dataset comprises several files, which each include a specific simulated source defect, such as incremental drift or outlier readings. For this evaluation, the files “data_standard.csv” and “data_drift_0_001.csv” are used.

The first provides unaltered data without defects. The second one contains the same data with the exception that a temperature sensor (feature 15) drifts with $1\% \text{ h}^{-1}$ of its base value. Regarding preprocessing, the parameters for the unimodal potential function (31) are provided as metadata in the dataset. As data are hardly affected by noise, sources are fully reliable, and no averaging filter is applied. Data are provided with an error margin of $\pm 2\%$ of the sensor’s measurement range [21] creating a uniform probability density function. Thus, preprocessing results in triangular possibility distributions.

The fusion topology is learned on unaltered data using the consistency-based approach of Algorithm 1—again with $\alpha = 0$. This creates three fusion nodes on the first layer:

- $fn_{(1)}$ with $\mathbf{S}_{(1)}^{\text{MCS}-0} = \{17, 14, 21, 16, 18, 22, 12, 19, 13, 11\}$,
- $fn_{(2)}$ with $\mathbf{S}_{(2)}^{\text{MCS}-0} = \{12, 15, 19, 13, 11, 20\}$, and
- $fn_{(3)}$ with $\mathbf{S}_{(3)}^{\text{MCS}-0} = \{10, 9, 8, 1, 2, 3, 4, 5, 6, 7\}$.

Their fusion results are fused at the final node $fn_{(1,2)}$ using MCS fusion (6). For the first layer nodes, the following fusion rules are used and evaluated:

- renormalised conjunctive fusion based on (15),
- discounted renormalised conjunctive fusion extending (15) with (23), (25),
- estimation fusion (13), and
- weighted estimation fusion (27).

Intermediate and final fusion outputs are computed for each of these fusion rules. The results of the same fusion rule on unaltered (standard) and drifted data are compared regarding their similarity. As similarity measure the possibilistic Jaccard index [32,85]

$$sim(\mathbf{p}) = \frac{\int_0^1 \min(\pi_{(k),\text{standard}}^{\text{fu}}(\mu), \pi_{(k),\text{drift}}^{\text{fu}}(\mu)) \, d\mu}{\int_0^1 \max(\pi_{(k),\text{standard}}^{\text{fu}}(\mu), \pi_{(k),\text{drift}}^{\text{fu}}(\mu)) \, d\mu} \tag{32}$$

is applied. Similarities $sim \in [0, 1]$ with $sim = 1$ indicating full similarity. Table 4 lists the minimum, arithmetic mean, and maximum values of the computed similarity values for $fn_{(2)}$ and $fn_{(1,2)}$. High similarities show robust behaviour against the defective source. As $fn_{(1)}$ and $fn_{(3)}$ contain no defective sources, they are omitted from the table.

It can be seen from the results that renormalised conjunctive fusion, which is the default rule in MCS fusion, was affected the most by the drifting source. Measures against defective sources are therefore reasonable.

The approach of detecting and discounting by widening inconsistent possibility distributions improved the robustness slightly but not substantially. The ineffectiveness is due to two reasons. First, widening with (25) does shift the fusion result toward reliable sources but does not guarantee that the original fusion result is restored. It is reasonable to assume that parameter β has a substantial impact, which needs to be investigated in further works. Second, a drifting possibility distribution may actually drift into other possibility distributions creating a false most consistent subset in the process.

This may lead to situations in which the wrong source is discounted. It is assumed that the risk of this happening decreases with the number of sources in a fusion node.

Estimation fusion nodes showed, on the other hand, a significant increase in robustness evidenced by the higher min- and mean-values. Weighted estimation fusion demonstrated the best performance. Due to its averaging nature, estimation fusion reduces the effects of defective sources the better the higher the number of sources.

Table 4. Similarity between fusion node outputs on unaltered (standard) dataset and drift affected dataset. The table shows the minimum, arithmetic mean, and maximum of similarities computed on each data instance. The drift affected source belongs to $fn_{(2)}$. Therefore, $fn_{(1)}$ and $fn_{(3)}$ are not explicitly listed. Similarity is increased by proposed countermeasures to defective sources.

Fusion Approach	$fn_{(2)}$ Similarity (32)			$fn_{(1,2)}$ Similarity (32)		
	min	mean	max	min	mean	max
Renormalised Conjunctive (15)	0.0007	0.3166	1	0.0162	0.5335	1
Discounted Renormalised Conjunctive (15), (23), (25)	0	0.5064	1	0.0162	0.6205	1
Estimation (13)	0.2686	0.6824	1	0.1336	0.9023	1
Weighted Estimation (27)	0.3240	0.8051	1	0.6288	0.9742	1

6. Conclusions

Choosing a topology is one of the main challenges in information fusion system design. Associativity, consistency, and redundancy play key roles in the performance of a topology. In this article, we detailed and discussed a data-driven design approach resulting in a two-layer topology inspired by MCS fusion. Due to the associativity of fusion rules in the first layer nodes, the topology can be extended to multiple layers without affecting the fusion results.

The basic design approach relies on the consistency of information items to find MCS nodes. The resulting consistency-based topology was susceptible to unexpected behaviour from information sources caused by unrepresentative training data or defective sources. We proposed adaptations to the basic design comprising the inclusion of a redundancy metric, the automated discounting of defective sources, and the application of outlier robust estimation fusion.

In the evaluation, we demonstrated that the redundancy-enhanced design resulted in more robust topologies in the case of epistemic uncertainty. Furthermore, evaluation showed that discounting defective sources and estimation fusion reduced the effects of defective sources. Estimation fusion outperformed the discounting approach in this regard mainly because, in certain situations, the discounting approach incorrectly identified sources as defective. Further work is required to improve this.

While the consistency-based approach found MCS in linear time regarding the number of sources and number of data instances, the redundancy-enhanced version searched the power set of all MCS. Although $\forall k : |\mathbf{S}_{(k)}^{MCS-\alpha}| \leq |\mathbf{S}|$ and although, in practical applications, it is reasonable to assume $\forall k : |\mathbf{S}_{(k)}^{MCS-\alpha}| \ll |\mathbf{S}|$, the scalability of the redundancy-based approach needs to be improved in further works. Another topic that should be addressed in further works is to adapt the design approaches so that they are able to update a topology on streamed data. With new dates becoming available, the epistemic uncertainty is reduced. Updating a topology has the potential to improve the fusion results continuously in small steps.

Author Contributions: C.-A.H. conceptualised the methodology, conducted the research, and wrote the article. V.L. supervised the research activity and revised the article. All authors have read and agreed to the published version of the manuscript.

Funding: This work was partly funded by the German Federal Ministry of Education and Research (BMBF) within the project ITS.ML (grant no. 01IS18041D) and the Ministry of Economic Affairs, Innovation, Digitalisation and Energy of the State of North Rhine-Westphalia (MWIDE) within the project ML4Pro² (grant no. 005-1807-0090).

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Appendix A. Specificity as a Measure of Information Content

This Appendix section recaps specificity as a measure of the information content of a possibility distribution. Specificity has been mathematically defined by ZADEH [44], DUBOIS et al. [49], and MAURIS et al. [78] as a relative quantity between two information items (π_1 is more specific than π_2 if $\forall x \in X : \pi_1(x) < \pi_2(x)$). Absolute measures for specificity have been formalised by YAGER [51–53] as well as HIGASHI and KLIR [86,87].

A specificity measure $spec(\pi) \in [0, 1]$ has to satisfy four conditions:

1. $spec(\pi) = 0$ in the case of total ignorance, i.e., $\forall x \in X : \pi(x) = 1$.
2. $spec(\pi) = 1$ in the case of complete knowledge, i.e., only one unique event is totally possible and all other events are impossible.
3. A specificity measure de- and increases with the maximum value of $\pi(x)$, i.e., let π_k be the k -th largest possibility degree in $\pi(x)$, then $\frac{dspec(\pi)}{d\pi_k} > 0$.
4. $\forall k > 2 : \frac{dspec(\pi)}{d\pi_k} \leq 0$, i.e., the specificity decreases as the possibilities of other values approach the maximum value of $\pi(x)$.

The measure of possibilistic specificity is a counterpart of Shannon’s probabilistic entropy [45,86].

A measure of specificity for a real-valued, continuous frame of discernments is given by Yager [51–53]:

$$spec(\pi) = \alpha_{\max} - \frac{1}{(x_b - x_a)} \cdot \int_0^{\alpha_{\max}} \left(\max_{x \in A_\alpha} x - \min_{x \in A_\alpha} x \right) d\alpha, \tag{A1}$$

with x_a and x_b being the borders of X ($X = [x_a, x_b]$). For (A1), it is proven by Yager [51–53] that the measure satisfies the four requirements for specificity measures. The integral in (A2) is equivalent to the area under A [50]. Therefore, (A1) is equal to

$$\begin{aligned} spec(\pi) &= \alpha_{\max} - \frac{1}{(x_b - x_a)} \cdot \int_{x_a}^{x_b} \pi(x) dx \\ &= \max_{x \in X} \pi(x) - \frac{1}{(x_b - x_a)} \cdot \int_{x_a}^{x_b} \pi(x) dx. \end{aligned} \tag{A2}$$

Appendix B. Proofs of (Non-)Associativity of Fusion Rules

Proposition A1. *The quantified fusion rule as formalised in (7) is not associative, i.e., $fu(I_1, I_2, I_3) \neq fu(fu(I_1, I_2), I_3)$ with $I_i = \pi_i$.*

Proof by Counterexample. Assume three possibility distributions defined by key points $(x, \pi(x))$ as follows: $\pi_1 = ((0.1, 0), (0.2, 1), (0.6, 1), (0.75, 0))$, $\pi_2 = ((0.2, 0), (0.3, 1), (0.5, 1), (0.55, 0))$, and $\pi_3 = ((0.35, 0), (0.7, 1), (0.8, 1), (0.9, 0))$. With $j = 2$, fusion results using (7) $\pi_1^{(fu)} = fu(I_1, I_2, I_3)$ and $\pi_2^{(fu)} = fu(fu(I_1, I_2), I_3)$ are clearly different as shown in Figure A1. This example proves that (7) is not associative. □

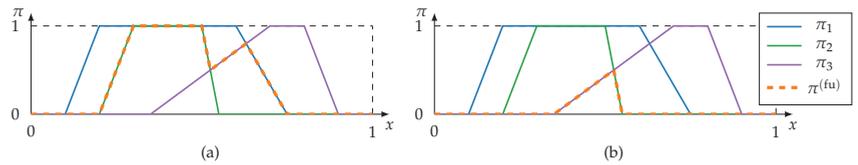


Figure A1. Three possibility distributions fused by the quantified fusion rule (7). Plot (a) shows $\pi_1^{(fu)} = fu(I_1, I_2, I_3)$. Plot (b) shows $\pi_2^{(fu)} = fu(fu(I_1, I_2), I_3)$.

Proposition A2. The majority-opinion-guided possibilistic fusion rule (14) is not associative, i.e., $fu(I_1, I_2, I_3) \neq fu(fu(I_1, I_2), I_3)$ with $I_i = \pi_i$.

Proof by Counterexample. Assume three possibility distributions defined by key points $(x, \pi(x))$ as follows: $\pi_1 = ((0.1, 0), (0.2, 1), (0.6, 1), (0.75, 0))$, $\pi_2 = ((0.2, 0), (0.3, 1), (0.5, 1), (0.55, 0))$, and $\pi_3 = ((0.35, 0), (0.7, 1), (0.8, 1), (0.9, 0))$. With $rel = (1, 1, 1)$, $\mathbf{v} = (1, 1, 1)$, $\mathbf{w}_p = (0, 0, 1)$, and $\mathbf{w}_m = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$, fusion results using (14) $\pi_1^{(fu)} = fu(I_1, I_2, I_3)$ and $\pi_2^{(fu)} = fu(fu(I_1, I_2), I_3)$ are clearly different as shown in Figure A2. This example proves that (14) is not associative. □

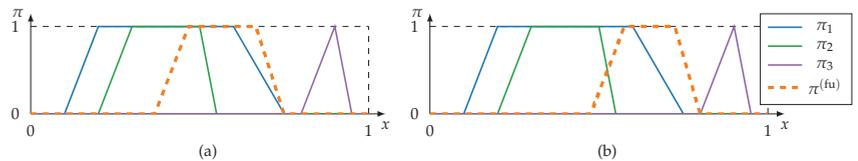


Figure A2. Three possibility distributions fused by the majority-opinion-guided possibilistic fusion rule (14). Plot (a) shows $\pi_1^{(fu)} = fu(I_1, I_2, I_3)$. Plot (b) shows $\pi_2^{(fu)} = fu(fu(I_1, I_2), I_3)$.

References

- Hall, D.L.; Llinas, J.; Liggins, M.E. (Eds.) *Handbook of Multisensor Data Fusion: Theory and Practice*, 2nd ed.; The Electrical Engineering and Applied Signal Processing Series; CRC Press: Boca Raton, FL, USA, 2009.
- Ayyub, B.M.; Klir, G.J. *Uncertainty Modeling and Analysis in Engineering and the Sciences*; Chapman & Hall/CRC: Boca Raton, FL, USA, 2006.
- Dubois, D.; Prade, H. On the use of aggregation operations in information fusion processes. *Fuzzy Sets Syst.* **2004**, *142*, 143–161. [CrossRef]
- Dubois, D.; Prade, H. Possibility theory in information fusion. In Proceedings of the Third International Conference on Information Fusion, Paris, France, 10–13 July 2000; Volume 1, pp. PS6–PS19. [CrossRef]
- Dubois, D.; Everaere, P.; Konieczny, S.; Papini, O. Main issues in belief revision, belief merging and information fusion. In *A Guided Tour of Artificial Intelligence Research: Volume I: Knowledge Representation, Reasoning and Learning*; Marquis, P.; Papini, O.; Prade, H., Eds.; Springer International Publishing: Cham, Switzerland, 2020; pp. 441–485. [CrossRef]
- Dubois, D.; Liu, W.; Ma, J.; Prade, H. The basic principles of uncertain information fusion. An organised review of merging rules in different representation frameworks. *Inf. Fusion* **2016**, *32*, 12–39. [CrossRef]
- Varshney, P.K. Multisensor data fusion. *Electron. Commun. Eng. J.* **1997**, *9*, 245–253. [CrossRef]
- Mitchell, H.B. (Ed.) *Multi-Sensor Data Fusion: An Introduction*; Springer: Berlin/Heidelberg, Germany, 2007. [CrossRef]
- Castanedo, F.; Ursino, D.; Takama, Y. A review of data fusion techniques. *Sci. World J.* **2013**, *2013*, 704504. [CrossRef] [PubMed]
- Bakr, M.A.; Lee, S. Distributed multisensor data fusion under unknown correlation and data inconsistency. *Sensors* **2017**, *17*, 2472. [CrossRef] [PubMed]
- Durrant-Whyte, H.F. Sensor models and multisensor integration. *Int. J. Robot. Res.* **1988**, *7*, 97–113. [CrossRef]
- Elmenreich, W. *An Introduction to Sensor Fusion*; Technical Report; Vienna University of Technology: Vienna, Austria, 2002.
- Elmenreich, W. A review on system architectures for sensor fusion applications. In *Software Technologies for Embedded and Ubiquitous Systems*; Obermaisser, R., Nah, Y., Puschner, P., Rammig, F.J., Eds.; Springer: Berlin/Heidelberg, Germany, 2007; pp. 547–559. [CrossRef]
- Ben Ayed, S.; Trichili, H.; Alimi, A.M. Data fusion architectures: A survey and comparison. In Proceedings of the 2015 15th International Conference on Intelligent Systems Design and Applications (ISDA), Marrakech, Morocco, 14–16 December 2015; pp. 277–282. [CrossRef]

15. Sidek, O.; Quadri, S.A. A review of data fusion models and systems. *Int. J. Image Data Fusion* **2012**, *3*, 3–21. [[CrossRef](#)]
16. Raz, A.K.; Wood, P.; Mockus, L.; DeLaurentis, D.A.; Llinas, J. Identifying interactions for information fusion system design using machine learning techniques. In Proceedings of the 2018 21st International Conference on Information Fusion (FUSION), Cambridge, UK, 10–13 July 2018; pp. 226–233. [[CrossRef](#)]
17. Luo, R.C.; Kay, M.G. Multisensor integration and fusion in intelligent systems. *IEEE Trans. Syst. Man Cybern.* **1989**, *19*, 901–931. [[CrossRef](#)]
18. Mönks, U. Information Fusion Under Consideration of Conflicting Input Signals. In *Technologies for Intelligent Automation*; Springer: Berlin/Heidelberg, Germany, 2017. [[CrossRef](#)]
19. Fritze, A.; Mönks, U.; Holst, C.A.; Lohweg, V. An approach to automated fusion system design and adaptation. *Sensors* **2017**, *17*, 601. [[CrossRef](#)]
20. Rescher, N.; Manor, R. On inference from inconsistent premisses. *Theory Decis.* **1970**, *1*, 179–217. [[CrossRef](#)]
21. Ehlenbröker, J.F.; Mönks, U.; Lohweg, V. Sensor defect detection in multisensor information fusion. *J. Sensors Sens. Syst.* **2016**, *5*, 337–353. [[CrossRef](#)]
22. Holst, C.A.; Lohweg, V. A conflict-based drift detection and adaptation approach for multisensor information fusion. In Proceedings of the 2018 IEEE 23rd International Conference on Emerging Technologies and Factory Automation (ETFA), Turin, Italy, 4–9 September 2018; pp. 967–974. [[CrossRef](#)]
23. Holst, C.A.; Lohweg, V. Improving majority-guided fuzzy information fusion for Industry 4.0 condition monitoring. In Proceedings of the 2019 22nd International Conference on Information Fusion (FUSION), Ottawa, ON, Canada, 2–5 July 2019.
24. Holst, C.A.; Lohweg, V. Feature fusion to increase the robustness of machine learners in industrial environments. *Automation* **2019**, *67*, 853–865. [[CrossRef](#)]
25. Mönks, U.; Lohweg, V.; Dörksen, H. Conflict measures and importance weighting for information fusion applied to Industry 4.0. In *Information Quality in Information Fusion and Decision Making*; Bossé, É., Rogova, G.L., Eds.; Information Fusion and Data Science; Springer International Publishing: Cham, Switzerland, 2019; pp. 539–561. [[CrossRef](#)]
26. Fritze, A.; Mönks, U.; Lohweg, V. A support system for sensor and information fusion system design. *Procedia Technol.* **2016**, *26*, 580–587. [[CrossRef](#)]
27. Fritze, A.; Mönks, U.; Lohweg, V. A concept for self-configuration of adaptive sensor and information fusion systems. In Proceedings of the IEEE 21st International Conference on Emerging Technologies and Factory Automation (ETFA), Berlin, Germany, 6–9 September 2016. [[CrossRef](#)]
28. Boury-Brisset, A.C. Ontology-based approach for information fusion. In Proceedings of the Sixth International Conference of Information Fusion, Cairns, QLD, Australia, 8–11 July 2003; pp. 522–529. [[CrossRef](#)]
29. Martí, E.; García, J.; Molina, J.M. Adaptive sensor fusion architecture through ontology modeling and automatic reasoning. In Proceedings of the 2015 18th International Conference on Information Fusion (Fusion), Washington, DC, USA, 6–9 July 2015; pp. 1144–1151.
30. Steinberg, A.N.; Bowman, C.L. Revisions to the JDL Data Fusion Model. In *Handbook of Multisensor Data Fusion*; Hall, D.L., Llinas, J., Eds.; The Electrical Engineering and Applied Signal Processing Series; CRC Press: Boca Raton, FL, USA, 2001; pp. 2-1-2-19.
31. Smoleń, M.; Augustyniak, P. Assisted living system with adaptive sensor's contribution. *Sensors* **2020**, *20*, 5278. [[CrossRef](#)] [[PubMed](#)]
32. Solaiman, B.; Bossé, É. *Possibility Theory for the Design of Information Fusion Systems*; Information Fusion and Data Science; Springer International Publishing: Cham, Switzerland, 2019.
33. Waltz, E.; Llinas, J. *Multisensor Data Fusion*; Artech House: Boston, MA, USA, 1990; Volume 685.
34. Grabisch, M.; Prade, H. The correlation problem in sensor fusion in a possibilistic framework. *Int. J. Intell. Syst.* **2001**, *16*, 1273–1283. [[CrossRef](#)]
35. Ayoun, A.; Smets, P. Data association in multi-target detection using the transferable belief model. *Int. J. Intell. Syst.* **2001**, *16*, 1167–1182. [[CrossRef](#)]
36. Schubert, J. Clustering belief functions based on attracting and conflicting metalevel evidence using Potts spin mean field theory. *Inf. Fusion* **2004**, *5*, 309–318. [[CrossRef](#)]
37. Schubert, J. Clustering decomposed belief functions using generalized weights of conflict. *Int. J. Approx. Reason.* **2008**, *48*, 466–480. [[CrossRef](#)]
38. Holst, C.A.; Lohweg, V. A redundancy metric based on the framework of possibility theory for technical systems. In Proceedings of the 2020 25th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA), Vienna, Austria, 8–11 September 2020. [[CrossRef](#)]
39. Holst, C.A.; Lohweg, V. A redundancy metric set within possibility theory for multi-sensor systems. *Sensors* **2021**, *21*, 2508. [[CrossRef](#)]
40. Kamal, A.T.; Bappy, J.H.; Farrell, J.A.; Roy-Chowdhury, A.K. Distributed multi-target tracking and data association in vision networks. *IEEE Trans. Pattern Anal. Mach.* **2016**, *38*, 1397–1410. [[CrossRef](#)]
41. Yoon, K.; Du Kim, Y.; Yoon, Y.C.; Jeon, M. Data association for multi-object tracking via deep neural networks. *Sensors* **2019**, *19*, 559. [[CrossRef](#)] [[PubMed](#)]
42. Khaleghi, B.; Khamis, A.; Karray, F.O.; Razavi, S.N. Multisensor data fusion: A review of the state-of-the-art. *Inf. Fusion* **2013**, *14*, 28–44. [[CrossRef](#)]

43. Lohweg, V.; Voth, K.; Glock, S. A possibilistic framework for sensor fusion with monitoring of sensor reliability. In *Sensor Fusion*; Thomas, C., Ed.; IntechOpen: Rijeka, Croatia, 2011. [\[CrossRef\]](#)
44. Zadeh, L.A. Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets Syst.* **1978**, *1*, 3–28. [\[CrossRef\]](#)
45. Denœux, T.; Dubois, D.; Prade, H. Representations of uncertainty in artificial intelligence: Probability and possibility. In *A Guided Tour of Artificial Intelligence Research: Volume I: Knowledge Representation, Reasoning and Learning*; Marquis, P., Papini, O., Prade, H., Eds.; Springer International Publishing: Cham, Switzerland, 2020; pp. 69–117. [\[CrossRef\]](#)
46. Salicone, S.; Prioli, M. *Measuring Uncertainty within the Theory of Evidence*; Springer Series in Measurement Science and Technology; Springer: Cham, Switzerland, 2018. [\[CrossRef\]](#)
47. Dubois, D.; Prade, H. Representation and combination of uncertainty with belief functions and possibility measures. *Comput. Intell.* **1988**, *4*, 244–264. [\[CrossRef\]](#)
48. Dubois, D.; Prade, H. Possibility theory and data fusion in poorly informed environments. *Control Eng. Pract.* **1994**, *2*, 811–823. [\[CrossRef\]](#)
49. Dubois, D.; Foulloy, L.; Mauris, G.; Prade, H. Probability-possibility transformations, triangular fuzzy sets, and probabilistic inequalities. *Reliab. Comput.* **2004**, *10*, 273–297. [\[CrossRef\]](#)
50. Yager, R.R. On ordered weighted averaging aggregation operators in multicriteria decisionmaking. *IEEE Trans. Syst. Man Cybern.* **1988**, *18*, 183–190. [\[CrossRef\]](#)
51. Yager, R.R. On the specificity of a possibility distribution. *Fuzzy Sets Syst.* **1992**, *50*, 279–292. [\[CrossRef\]](#)
52. Yager, R.R. Measures of specificity. In *Computational Intelligence: Soft Computing and Fuzzy-Neuro Integration with Applications*; Kaynak, O., Zadeh, L.A., Türkşen, B., Rudas, I.J., Eds.; Springer: Berlin/Heidelberg, Germany, 1998; pp. 94–113.
53. Yager, R.R. Measures of specificity over continuous spaces under similarity relations. *Fuzzy Sets Syst.* **2008**, *159*, 2193–2210. [\[CrossRef\]](#)
54. Yager, R.R. Aggregation operators and fuzzy systems modeling. *Fuzzy Sets Syst.* **1994**, *67*, 129–145. [\[CrossRef\]](#)
55. Zadeh, L.A. The concept of a linguistic variable and its application to approximate reasoning—II. *Inf. Sci.* **1975**, *8*, 301–357. [\[CrossRef\]](#)
56. Klement, E.P. *Triangular Norms*; Springer eBook Collection Mathematics and Statistics; Springer: Dordrecht, Germany, 2000; Volume 8. [\[CrossRef\]](#)
57. Benferhat, S.; Dubois, D.; Prade, H. Reasoning in inconsistent stratified knowledge bases. In Proceedings of the 26th IEEE International Symposium on Multiple-Valued Logic (ISMVL'96), Compostela, Spain, 29–31 May 1996; pp. 184–189. [\[CrossRef\]](#)
58. Dubois, D.; Fargier, H.; Prade, H. Multi-source information fusion: A way to cope with incoherences. In Proceedings of the French Days on Fuzzy Logic and Applications (LFA), Paris, France, 21 October 2000; pp. 123–130.
59. Liu, W.; Qi, G.; Bell, D.A. Adaptive merging of prioritized knowledge bases. *Fundam. Inform.* **2006**, *73*, 389–407.
60. Hunter, A.; Liu, W. A context-dependent algorithm for merging uncertain information in possibility theory. *IEEE Trans. Syst. Man Cybern. Part A Syst. Hum.* **2008**, *38*, 1385–1397. [\[CrossRef\]](#)
61. Destercke, S.; Dubois, D.; Chojnacki, E. Possibilistic information fusion using maximal coherent subsets. *IEEE Trans. Fuzzy Syst.* **2009**, *17*, 79–92. [\[CrossRef\]](#)
62. Yager, R.R. Aggregating evidence using quantified statements. *Inf. Sci.* **1985**, *36*, 179–206. [\[CrossRef\]](#)
63. Dubois, D.; Prade, H.; Testemale, C. Weighted fuzzy pattern matching. *Fuzzy Sets Syst.* **1988**, *28*, 313–331. [\[CrossRef\]](#)
64. Oussalah, M.; Maaref, H.; Barret, C. From adaptive to progressive combination of possibility distributions. *Fuzzy Sets Syst.* **2003**, *139*, 559–582. [\[CrossRef\]](#)
65. Yager, R.R. A general approach to the fusion of imprecise information. *Int. J. Intell. Syst.* **1997**, *12*, 1–29. [\[CrossRef\]](#)
66. Zadeh, L.A. Fuzzy sets. *Inf. Control* **1965**, *8*, 338–353. [\[CrossRef\]](#)
67. Glock, S.; Voth, K.; Schaede, J.; Lohweg, V. A framework for possibilistic multi-source data fusion with monitoring of sensor reliability. In Proceedings of the World Conference on Soft Computing, San Francisco, CA, USA, 23–26 May 2011.
68. Larsen, H.L. Efficient importance weighted aggregation between min and max. In Proceedings of the ninth Conference on Information Processing and Management of Uncertainty in Knowledge-based Systems, Annecy, France, 1–5 July 2002; pp. 1203–1208.
69. Oussalah, M. Study of some algebraical properties of adaptive combination rules. *Fuzzy Sets Syst.* **2000**, *114*, 391–409. [\[CrossRef\]](#)
70. Calude, C.; Longo, G. The deluge of spurious correlations in big data. *Found. Sci.* **2017**, *22*, 595–612. [\[CrossRef\]](#)
71. Delmotte, F.; Borne, P. Modeling of reliability with possibility theory. *IEEE Trans. Syst. Man Cybern. Part A Syst. Hum.* **1998**, *28*, 78–88. [\[CrossRef\]](#)
72. Žliobaitė, I. Learning under concept drift: An overview. *arXiv* **2010**, arXiv:1010.4784. [\[CrossRef\]](#)
73. Ramírez-Gallego, S.; Krawczyk, B.; García, S.; Woźniak, M.; Herrera, F. A survey on data preprocessing for data stream mining: Current status and future directions. *Neurocomputing* **2017**, *239*, 39–57. [\[CrossRef\]](#)
74. Dubois, D.; Del Cerro, L.F.; Herzig, A.; Prade, H. An ordinal view of independence with application to plausible reasoning. In *Uncertainty Proceedings 1994*; Mantaras, R.L.d., Poole, D., Eds.; Morgan Kaufmann: Burlington, MA, USA, 1994; pp. 195–203. [\[CrossRef\]](#)
75. Yager, R.R.; Kelman, A. Fusion of fuzzy information with considerations for compatibility, partial aggregation, and reinforcement. *Int. J. Approx. Reason.* **1996**, *15*, 93–122. [\[CrossRef\]](#)
76. Knuth, D.E. Big omicron and big omega and big theta. *ACM Sigact News* **1976**, *8*, 18–24. [\[CrossRef\]](#)

77. Lasserre, V.; Mauris, G.; Foulloy, L. A simple possibilistic modelisation of measurement uncertainty. In *Uncertainty in Intelligent and Information Systems*; Bouchon-Meunier, B., Yager, R.R., Zadeh, L.A., Eds.; World Scientific Publishing Co. Pte. Ltd.: Singapore, 2000; Volume 20, pp. 58–69. [[CrossRef](#)]
78. Mauris, G.; Lasserre, V.; Foulloy, L. Fuzzy modeling of measurement data acquired from physical sensors. *IEEE Trans. Instrum. Meas.* **2000**, *49*, 1201–1205. [[CrossRef](#)]
79. Lohweg, V.; Diederichs, C.; Müller, D. Algorithms for hardware-based pattern recognition. *EURASIP J. Appl. Signal Process.* **2004**, *2004*, 1912–1920. [[CrossRef](#)]
80. Voth, K.; Glock, S.; Mönks, U.; Lohweg, V.; Türke, T. Multi-sensory machine diagnosis on security printing machines with two-layer conflict solving. In Proceedings of the SENSOR+TEST Conference 2011, Nuremberg, Germany, 7–9 June 2011; pp. 686–691. [[CrossRef](#)]
81. Mönks, U.; Petker, D.; Lohweg, V. Fuzzy-Pattern-Classifier training with small data sets. In *Information Processing and Management of Uncertainty in Knowledge-Based Systems. Theory and Methods*; Hüllermeier, E., Kruse, R., Hoffmann, F., Eds.; Springer: Berlin/Heidelberg, Germany, 2010; pp. 426–435. [[CrossRef](#)]
82. Paschke, F.; Bayer, C.; Bator, M.; Mönks, U.; Dicks, A.; Enge-Rosenblatt, O.; Lohweg, V. Sensorlose Zustandsüberwachung an Synchronmotoren. In *23. Workshop Computational Intelligence*; Hoffmann, F., Hüllermeier, E., Mikut, R., Eds.; KIT Scientific Publishing: Karlsruhe, Germany, 2013; Volume 46, pp. 211–225.
83. Lessmeier, C.; Enge-Rosenblatt, O.; Bayer, C.; Zimmer, D. Data acquisition and signal analysis from measured motor currents for defect detection in electromechanical drive systems. In Proceedings of the PHM Society European Conference, Nantes, France, 8–10 July 2014. [[CrossRef](#)]
84. Dua, D.; Graff, C. *UCI Machine Learning Repository*; University of California, School of Information and Computer Science: Irvine, CA, USA, 2020. Available online: <http://archive.ics.uci.edu/ml> (accessed on 14 March 2022).
85. Charfi, A.; Bouhamed, S.A.; Bossé, É.; Kallel, I.K.; Bouchaala, W.; Solaiman, B.; Derbel, N. Possibilistic similarity measures for data science and machine learning applications. *IEEE Access* **2020**, *8*, 49198–49211. [[CrossRef](#)]
86. Higashi, M.; Klir, G.J. Measures of uncertainty and information based on possibility distributions. *Int. J. Gen. Syst.* **1982**, *9*, 43–58. [[CrossRef](#)]
87. Higashi, M.; Klir, G.J. On the notion of distance representing information closeness: Possibility and probability distributions. *Int. J. Gen. Syst.* **1983**, *9*, 103–115. [[CrossRef](#)]

Article

The Obtainable Uncertainty for the Frequency Evaluation of Tones with Different Spectral Analysis Techniques

Salvatore Dello Iacono, Giuseppe Di Leo, Consolatina Liguori * and Vincenzo Paciello

Department of Industrial Engineering, Università degli Studi di Salerno, Via Giovanni Paolo II, 132, 84084 Fisciano, Italy; sdelloiacono@unisa.it (S.D.I.); gdileo@unisa.it (G.D.L.); vpaciello@unisa.it (V.P.)

* Correspondence: tliguori@unisa.it

Abstract: Spectral analysis is successfully adopted in several fields. However, the requirements and the constraints of the different cases may be so varied that not only the tuning of the analysis parameters but also the choice of the most suitable technique can be a difficult task. For this reason, it is important that a designer of a measurement system for spectral analysis has knowledge about the behaviour of the different techniques with respect to the operating conditions. The case that will be considered is the realization of a numerical instrument for the real-time measurement of the spectral characteristics of a multi-tone signal (amplitude, frequency, and phase). For this purpose, different signal processing techniques can be used, that can be classified as parametric or non-parametric methods. The first class includes those methods that exploit the a priori knowledge about signal parameters, such as the spectral shape of the signal to be processed. Thus, a self-configuring procedure based on a parametric algorithm should include a preliminary evaluation of the number of components. The choice of the right method among several proposals in the literature is fundamental for any designer and, in particular, for the developers of spectral analysis software, for real-time applications and embedded devices where time and reliability constraints are arduous to fulfil. Different aspects should be considered: the desired level of accuracy, the available elaboration resources (memory depth and processing speed), and the signal parameters. The present paper details a comparison of some of the most effective methods available in the literature for the spectral analysis of signals (IFFT-2p, IFFT-3p, and IFFTc, all based on the use of an FFT algorithm, while improving the spectral resolution of the DFT with interpolation techniques and three parametric algorithms—MUSIC, ESPRIT, and IWPA). The methods considered for the comparison will be briefly described, and references to literature will be given for each one of them. Then, their behaviour will be analysed in terms of systematic contribution and uncertainty on the evaluated frequencies of the spectral tones of signals created from superimposed sinusoids and white Gaussian noise.

Citation: Dello Iacono, S.; Di Leo, G.; Liguori, C.; Paciello, V. The Obtainable Uncertainty for the Frequency Evaluation of Tones with Different Spectral Analysis Techniques. *Metrology* **2022**, *2*, 216–229. <https://doi.org/10.3390/metrology2020013>

Academic Editor: Simona Salicone

Received: 31 August 2021

Accepted: 17 February 2022

Published: 14 April 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: digital signal processing; spectral resolution; frequency domain analysis; frequency-domain interpolation; frequency uncertainty

1. Introduction

The spectral analysis of signals is successfully adopted in several fields—from electrical [1,2] to typical industrial fields—for speed and fault detection on motors and bearings [3–5] in military applications [6], submarine applications [7], and medical applications [8]. Despite the adaptability of frequency analysis for varied applications [9], similar cases might require different approaches, requirements, and constraints; for this reason, they differ in the tuning of the analysis parameters, and the choice of the most suitable technique can be a difficult task to accomplish. For this reason, the designer of a measurement system for spectral analysis must have knowledge about the behaviour of the different techniques concerning the operating conditions. The case that will be considered is the realisation of a numerical instrument for the real-time measurement of the spectral components of a signal: amplitude, frequency, and phase.

For this purpose, different signal processing techniques can be used that can be generally classified in parametric and non-parametric methods. The first class includes those methods that exploit the a priori knowledge about the signal parameters, such as the number of signal spectral components. Thus, a self-configuring procedure based on a parametric algorithm should include a preliminary evaluation of the number of components. The choice of the proper method among the several available approaches is fundamental for designing a procedure based on signal spectral analysis. Different aspects should be considered: the desired level of accuracy, the available elaboration resources (memory depth and processing speed), and the signal nature.

This paper will compare the most effective methods available in the literature for the spectral analysis of signals [10–25]. The considered methods for the comparison will be briefly described, and references from the literature will be given for each one of them. Their behaviour will be analysed in terms of obtainable uncertainty on the frequency evaluation. Residual errors and repeatability of the measured frequency directly influence the uncertainty of other tone properties, such as phase and amplitude.

The procedure and the criteria adopted for the comparison are described and, eventually, the results are reported and commented upon. Numerical simulations have been run in conditions similar to the real-world operation of a measurement system by studying the effects of added Gaussian noise or quantisation noise to the signal, and their results are shown in this article.

2. Considered Methods

In this section, the algorithms considered for the comparison will be briefly described. At first, some non-parametric algorithms will be presented (IFFT-2p, IFFT-3p, IFFTc), they are all based on the use of a fast Fourier transform (FFT) algorithm, but they improve the spectral resolution of the discrete Fourier transform (DFT) algorithm, with interpolation techniques. Then three parametric algorithms will be introduced (MUSIC, ESPRIT, and IWPA), based on approaches different from the DFT evaluation. Since almost any signal can be represented as a multi-tone signal (1), composed of the sum of N_s sinusoids with amplitude, A_i , and phase, ϕ_i ; all the algorithms will be compared with respect to this signal family.

$$x(t) = \sum_{i=1}^{N_s} A_i \sin(2\pi f_i t + \phi_i). \tag{1}$$

2.1. Non-Parametric Methods

Considering the multi-tone signal in (1), sampled with a T_s sampling period, the obtained signal is described by:

$$x(n) = \sum_{i=1}^{N_s} A_i \sin(2\pi f_i n T_s + \phi_i) \quad n = 1 \dots N. \tag{2}$$

Non-parametric methods are based on the DFT algorithm, where the spectrum samples are evaluated as follows:

$$X(k) = \frac{1}{G} \sum_{n=0}^{N-1} w(n)x(n)e^{-\frac{2j\pi}{N}kn} \quad k = 0 \dots N - 1, \tag{3}$$

where $x(n)$ is the sampled signal (2) and $w(n)$ are the window samples with gain G , and k is the spectral bin index, also known as the bin number. If the sampled signal is coherent with the module of the sampled sequence DFT (3), then $M(k) = |X(k)|$ presents N_s peaks, corresponding to the N_s tone frequencies; the i -th peak is located exactly at index k_i .

When coherent sampling conditions are not assured, a quantization error arises in the frequency estimation [26]; the tone module is underestimated because of the spectral leakage. Moreover, harmonic interference is present, causing an error in parameters

estimation when the sampled signal presents two tones with a small frequency difference compared with the frequency resolution, or when it has only one tone but the frequency is less than two times the frequency resolution Δf .

To correct errors on frequency estimation, phase, and amplitude estimation, several non parametric methods have been exploited in the literature [10]. In the following sections, some non parametric methods will be briefly treated, in particular the interpolated FFT (IFFT) on two points and three points, and the corrected IFFT.

2.1.1. IFFT-2p

Interpolated FFT algorithms [10,11] have been known in the literature for several years, and those based on a two-point interpolation are the most common. The frequency, f_i , of the i -th tone can be evaluated as: $f_i = (k_i + \delta_i)\Delta f$, where Δf is the DFT frequency resolution ($\Delta f = f_s/N$), k is the integer part of the bin ($f/\Delta f$), and $\delta_i \in [-1/2, +1/2]$ is the fractional bin deviation. The fractional bin deviation, δ_i , is evaluated from the ratio between the two largest samples closest to the peak: $\alpha_i = \frac{|X(k_i + \epsilon_i)|}{|X(k_i)|}$, where: $\epsilon_i = \text{sign}(|X(k_i + 1)| - |X(k_i - 1)|)$. Considering the sampled spectrum of the window function, $W(k)$, the following is obtained [9]:

$$\alpha_i = \frac{|W(\epsilon_i - \delta_i)|}{|W(-\delta_i)|} = \frac{|W(k_i - \delta_i)|}{|W(k_i)|} \tag{4}$$

The value of δ_i can be evaluated from the latter relationship, given the window function and its analytical expression.

2.1.2. IFFT-3p

The interpolated three-point DFT algorithm [12–15] is based on an interpolation of the DFT results of the signal, windowed by cosine windows, and using three points for each tone peak. Considering the multi-tone signal, with N_s spectral components of (1), like the IFFT-2p, the frequency f_i of the i -th tone is evaluated as $f_i = (k_i + \delta_{3i})\Delta f$; in this case, δ_{3i} is evaluated considering the three largest samples of the peak:

$$\alpha_{3i} = \frac{|X(k_i - 1)| + |X(k_i + 1)|}{|X(k_i - 1)| + 2|X(k_i)| + |X(k_i + 1)|} \tag{5}$$

$$\delta_{3i} = K * \alpha_{3i}, \tag{6}$$

where K is a proportional factor that depends on the used windowing function; in the case of an Hanning window, this is $K = 2$.

2.1.3. IFFTc

The corrected interpolated FFT algorithm, presented in [16–18], is based on an IFFT-2p, but includes further processing to correct the effects of the harmonic interference between spectral components. Concerning the multi-tone signal of (1), it has been shown that the DFT value closest to the peak of the i -th spectral component can be written as: $X(k_i) = \frac{V_i}{S} W(-\delta_i) + F_i$, where $V_i = \frac{A_i}{2} e^{j\phi_i}$, $S = \sum_{n=0}^{N-1} w(n)$, and the contribution of the harmonic interference of other components on the i -th one can be taken into account by the term F_i . Similar considerations can be made for the second strongest bin: $X(k_i + \epsilon_i) = \frac{V_i}{S} W(\epsilon_i - \delta_i) + B_i$.

The α_i becomes: $\alpha'_i = \frac{|W(\epsilon_i - \delta_i)|}{|W(-\delta_i)|} = \frac{|X(k_i + \epsilon_i) - B_i|}{|X(k_i) - F_i|}$. The correction factors, F_i and B_i , depend on the frequency, amplitude, and phase of the signal tones. The proposed solution consists of using the values of frequency, amplitude, and phase measured with a preliminary two-point IFFT to evaluate the correction factors (IFFTc). In the presence of a low-frequency tone, the frequency image contribution can be corrected with the same relationships [20]. This step could be iterated further, but without any significant improvement in terms of estimation error reduction.

2.2. Parametric Methods

Numerous parametric methods exist in the literature; however, in this article, only the three algorithms presenting the best compromise in terms of computational requirements and estimation performance have been considered—MUSIC, ESPRIT, and IWPA.

2.2.1. MUSIC

This parametric algorithm (multiple signal classification) [19–21] determines the frequencies of the tones in a signal by performing a decomposition of the covariance matrix of the sequence of signal samples, $x(n)$. We modelled the input data as a N_s -tone signal and a superimposed noise, as follows:

$$x(n) = \sum_{i=1}^{N_s} A_i \sin(2\pi f_i n T_s + \phi_i) + z(n), \tag{7}$$

where $z(n)$ is the noise signal. The covariance of the signal is $R_x = E\{xx^H\}$, and can be numerically computed using signal samples, $x[n]$. If the noise is considered to be white Gaussian noise, then the signal can be decomposed in order to separate the signal from the noise orthogonal subspaces. The frequencies of the signal tones can be estimated from this decomposition [19]. To compute the MUSIC algorithm, the number of signal tones, N_s , must be known in advance; the same applies to the number of signals eigenvectors to be found with the decomposition.

2.2.2. ESPRIT

This parametric algorithm (estimation of signal parameter via rotational invariance technique), introduced in [22–24], exploits the rotational invariance property, which is valid for the signal eigenvectors (\mathbf{x}) of the sample sequence covariance matrix. Similar to the MUSIC algorithm, ESPRIT needs an estimation of the signal covariance matrix. Thanks to the knowledge of the number of components, the eigenvectors corresponding to signal components can be separated from the noise eigenvectors. Each signal eigenvector can be written as:

$$\begin{aligned} \mathbf{x}_k &= [x(0), x(1), \dots, x(N-2), x(N-1)] \\ &= \mathbf{A}_k \times [1, e^{j\omega_k}, e^{j2\omega_k}, \dots, e^{j(N-1)\omega_k}] \\ &= [s_1, x(N-1)] = [x(0), s_2], \end{aligned} \tag{8}$$

where \mathbf{A}_k is the coefficients vector. The $s_2 = s_1 e^{j\omega_1}$ rotational invariance property is valid, so all the signal eigenvectors and the signal components can be collected into the matrix, U , as well as into the signal components in the following matrices:

$$\Gamma_1 = [I_{M-1} | 0_{M-1}]_{(M_1) \times M} \tag{9a}$$

$$\Gamma_2 = [0_{M-1} | I_{M-1}]_{(M_1) \times M} \tag{9b}$$

Considering the rotational invariance property for each signal eigenvector, the selection matrices, Γ_1 (9a) and Γ_2 (9b), can be used to obtain the following system:

$$[\Gamma_1 U] \Phi = \Gamma_2 U, \tag{10}$$

where $\Phi = \text{diag}\{e^{j\omega_1}, e^{j\omega_2}, \dots, e^{j\omega_{N_s}}\}$. It is possible to obtain the frequencies of the components belonging to the signal solving this system with a least square technique.

2.2.3. IWPA

This method, proposed in [25], is based on the iteration of the weighted phase average algorithm (WPA). Considering the case of a signal with only one spectral component, $x(t) = A_0 \cos(2\pi f_0 t + \phi_0)$, a coarse estimation, \hat{f}_0 , of the frequency, f_0 , can be obtained in the first place, as the maximum of the amplitude of the DFT sequence, $X(k)$. The

signal is then divided into M non-overlapping segments of length P : $x_s(n) = x(n + s \cdot P)$, $0 \leq n \leq P - 1$.

In the simple but effective case of two segments and $P = N/2$, the spectra of the two segments, x_1 and x_2 , are evaluated at frequency \hat{f}_0 , and it can be shown that the fractional bin deviation, δ , can be estimated as:

$$\delta = \frac{N}{2\pi \cdot P} \left(\angle X_1(\hat{f}_0) - \angle X_2(\hat{f}_0) \right). \tag{11}$$

The IWPA algorithm, at each iteration, applies the WPA to obtain the frequency estimation of the strongest component, while amplitude and phase of this component are obtained through a least square technique. In the next step, the estimated component is subtracted from the samples of the previous iteration in the time domain. The number of iterations has to be equal to the number of components, so that a new component can be estimated at each iteration. The IWPA algorithm can be easily converted into a non-parametric algorithm by iterating its processing steps until the level of the residual decreases below a threshold.

3. Residual Errors

Due to approximations, the considered methods may exhibit a bias between the estimated and actual values of the signal tones, even when noise is not superimposed to the signal, and their expected values are not equal to their actual values. Such behaviour can be associated with interharmonic interference; as for IFFTc and IWPA, the behaviour can be associated with inadequate knowledge of the values required by parametric methods or the finite word length of the data processing.

To evaluate the proposed methods and produce a clear comparison of their performance, the multi-frequency signals described by (1) were considered; the tests are made for different values of the number of tones (N_s), the number of samples (N), the frequency (f_i), the amplitude (A_i), and the phase (ϕ_i). All the simulations have been made supposing an observation window longer than two periods of the signal. The major effects analysed are the frequency resolution, the signal dynamic range, and the harmonic interference [10].

$$f_i = (k_i + \delta_i)\Delta f \tag{12a}$$

$$A_i = \beta_i \cdot A_0 \tag{12b}$$

$$d_{ij} = \frac{f_j - f_i}{\Delta f} \tag{12c}$$

Given Equations (12a)–(12c), where k_i is the frequency bin index corresponding to f_i and δ_i is the fractional bin deviation, the simulations are made at changing values of δ_i , d_{ij} , β_i , ϕ_i , and N in order to analyse the dependence of the harmonic interference effects on the signal characteristics and the measurement system settings. In order to evaluate the interference effects on the different methods, tests with only two tones with the same amplitude ($A_1 = A_2$), corresponding to more substantial interference on both tones, refs. [3–5] are carried out at changing distances between tones, d_{12} ; with d_1 always greater than 20.

The logarithms of the absolute errors on the fractional bin deviations are calculated as the difference between the measured ($\hat{\delta}_i$) and the real value (δ_i), as follows:

$$E_{\delta_i} = |\hat{\delta}_i - \delta_i| \tag{13}$$

In Figure 1 the estimation error (13) for the first tone versus the distance between tones (d_{12}) is reported for the considered methods; similar results are obtained with the second tone. Interpolated FFT algorithms use the Hanning window, while for MUSIC and ESPRIT, $M = N/4$ was posed, and the matrix covariance was calculated using the samples with no noise added.

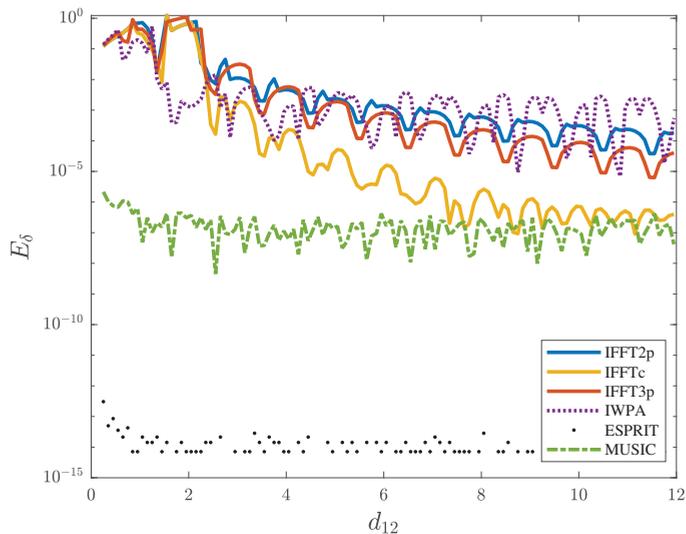


Figure 1. Absolute errors on δ obtained for a two-tone signal versus the distance, d_{12} , between tones.

Some considerations can be outlined, as follows:

- For each distance, the best performance is obtained by the ESPRIT method, that exhibits the lowest error at any distance between the tones since the error due to the frequency quantization is negligible.
- When the distance between tones is small ($d_{12} < 2\text{bins}$), the non-parametric approaches detect only one tone and the errors on the detected tone are significant (comparable with δ). Even if the IWPA method is able to estimate both tones and its errors are lower than those of the parametric approach, the error is still high.
- The tone distance slightly influences the algorithms based on the autocorrelation (MUSIC and ESPRIT): only for d_{12} lower than one bin is the MUSIC algorithm affected by a highest residual error.
- The performance of IWPA and IFFT are comparable, but for small tone distances, the IWPA gives better estimations—vice versa occurs for larger distances d_{12} .
- The IFFTC algorithm for tone distance greater than 8 bin gives results comparable with MUSIC: errors of the order of 10^{-6} are measured for both tones.

Since parametric algorithms require the knowledge of the number of spectral components, but the information can not be obtained in some applications, a characterization of all the algorithms will be reported for the case in which a different and generally wrong number of spectral components (N_{s_0}) is specified. For instance, Figure 2 shows the errors on δ versus the specified number of tones, N_{s_0} , in the case of a five-tone signal ($N_s = 5$) for the considered algorithms. The results refer to a signal with all the tones at the same amplitude ($A_i = 1$) and uniformly spaced with $d_{i,j-1} = 3$. For the cases where $N_{s_0} < 5$, the error on δ for a non-detected tone is evaluated with respect to the closest detected tone.

As expected, the algorithms based on IFFT, being non-parametric algorithms, are not influenced by N_{s_0} , and the residual errors are quite similar for each tone. Parametric methods MUSIC and ESPRIT manifest a different behaviour: errors are very high for each tone as long as N_{s_0} is lower than the actual number of tones. In other words, if N_{s_0} is lower than N_s , then the estimated frequencies are significantly different (at least $\Delta f/2$) from the actual frequencies of each of the five tones. When $N_{s_0} \geq N_s$, the ESPRIT method gives the best performance: it does not show residual errors, and small differences (less than 10^{-15}) are only caused by the finite word length of the preprocessing unit (CPU); MUSIC shows greater errors (about 10^{-7}), but these are negligible with respect to the other

methods. IWPA is less sensitive to an underestimated number of tones ($N_{s_0} \geq N_s$): in these cases, the frequency estimations are better than the other parametric methods, while for a $N_{s_0} < N_s$, its estimation deteriorates, since noise components are considered erroneously as signal tones until N_{s_0} components are detected. In the results of Figure 2, when $N_{s_0} < N_s = 5$, the errors for the undetected components are evaluated as the absolute difference between the actual value for that component and the estimated value for the closest component.

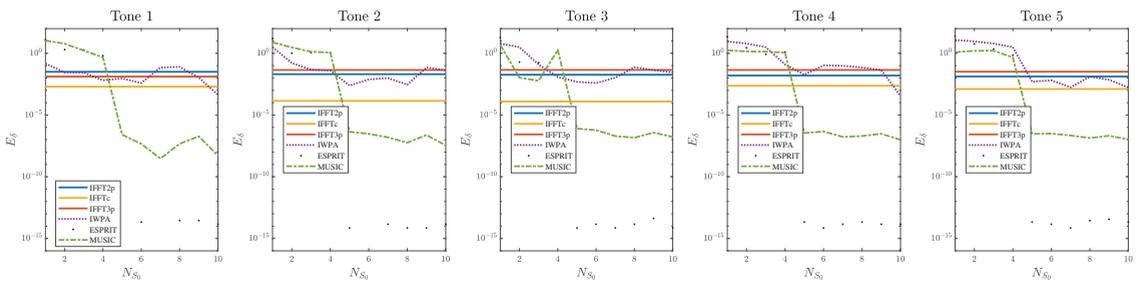


Figure 2. Errors on δ or a 5-tone signal ($N_s = 5$), versus the specified number of tones, N_{s_0} . Each figure refers to a single tone starting from tone 1 (on the left) to tone 5 (on the right).

Further tests were carried out to highlight the sensitivity of the different methods to the number of processed samples; in particular, the trends (not reported here for the sake of brevity) of the errors on δ_i , versus the bin distance, and versus the tone amplitudes, do not change when the number of acquired samples changes from 128 to 2048. This is expected for the error on δ_i , which is a kind of relative error and is different from the error on the frequency. Once the sampling frequency has been set, the greater the number of samples, the lower the spectral resolution, and, consequentially, the lower the error on frequency will be. However, a small reduction in the residual errors is measured only for the IWPA and IFFT methods when N increases (from about $E_\delta = 3 \times 10^{-3}$ with $N = 128$ to $E_\delta = 2 \times 10^{-4}$ for $N = 4096$).

4. Repeatability under Noisy Conditions

Some amount of noise always corrupts real-life signals, so the considered methods have to be evaluated when applied with noisy signals, since their performance may worsen significantly. The tests are carried out by changing the signal characteristics to estimate each method’s sensitivity to the tone composition; only two-tone signals are considered. Once the signal and the measurement parameters have been fixed, a Gaussian noise is added, noisy signal samples are generated, and the algorithms process these points in order to estimate the signal characteristics. For each signal, configuration, and noise level, the tests are repeated 1000 times; the mean and the standard deviation of the results of the algorithms are calculated. For the three algorithms, based on the FFT interpolation, a Hanning window is used.

4.1. Sensitivity to the First Tone Distance

Figure 3 reports the behaviour of the algorithms respect to a signal composed of two tones very close in frequency ($d_{12} = 3$ bins) and with the same null phase. The measured mean square error (MSE) versus the signal-to-noise ratio (SNR) for the different methods are reported, where the Cramér–Rao bound (CRB) [14] is also reported, since it gives information about the best theoretical performance (minimum variance of the quantity of interest) achievable with an ideal estimator, versus the level of superimposed noise. It has to be highlighted that the MSE considers both the random variability and the systematic effects [13]. The adopted CRB values are obtained with relationships valid in the specific case of a single-tone signal. However, the CRB estimation can be considered a kind of lower

limit, and the goodness of the estimation of a proposed method can be evaluated through the closeness of the resulting MSE to the CRB.

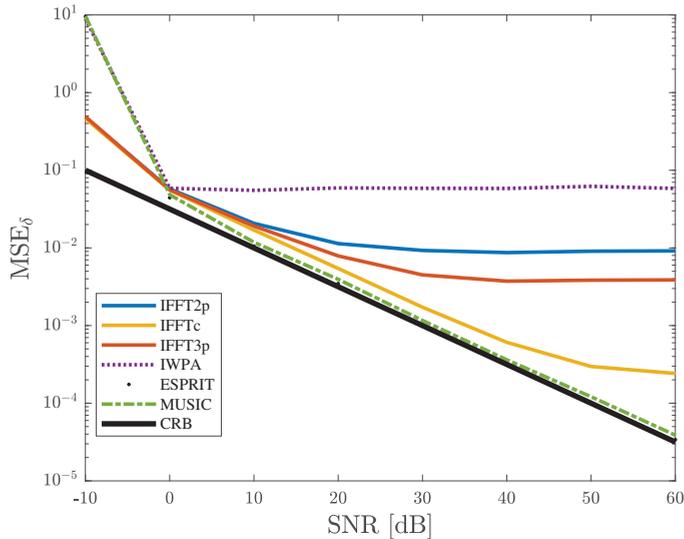


Figure 3. Mean square errors (MSE) versus the SNR for a two-tone signal with $A_1 = A_2 = 1$, $N = 256$, $f_1/\Delta f = 40.2$ bins, $d_{12} = 3$ bins, and zero phase difference.

Analysing these results, it is possible to state that IFFT and IFFTC algorithms are less sensitive to a high noise level than the other algorithms. In particular, IFFTC shows an MSE on δ less than 0 dB for SNR less than zero, while the errors can reach 20 dB for the other algorithms. For higher SNR, MUSIC, and ESPRIT show the best performance, but the results of IFFTC and IWPA are comparable with those of the other two methods when the phases are equal to zero. In presence of phase difference, not reported here for simplicity, the performance of ESPRIT and MUSIC does not change while IFFT deteriorates slightly (about 2 dB) for SNR values between 0 dB and 20 dB; the MSE on δ of the IFFTC algorithm declines of about 3 dB for high SNR (greater than 40 dB) when residual systematic effects on the phase estimation become predominant, and IWPA remarkably loses its estimation capability at the point that it can hardly be adopted.

Figure 4 reports the MSE on δ versus the relative distance between the two tones of a signal. The improvement of interpolation of the IFFTC over IFFTs is evident since IFFTC keeps good performance from d_{12} equal to 3 onwards. However, the lowest values of MSE_δ are reached by ESPRIT and MUSIC.

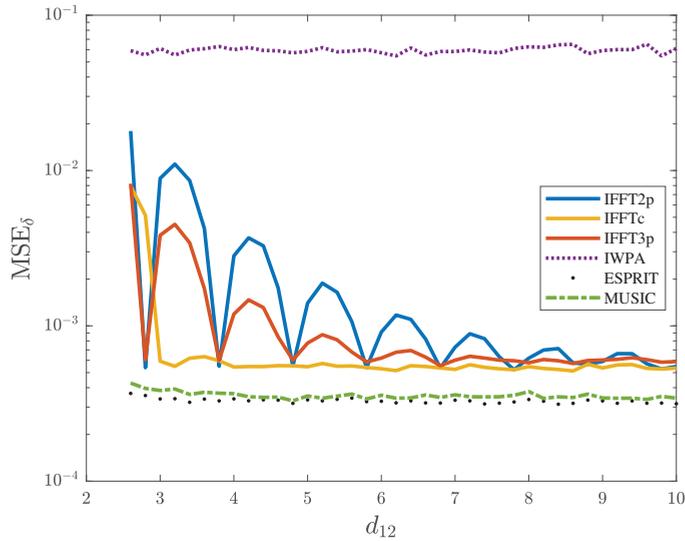


Figure 4. MSE of δ for the first tone versus the normalized tone distance d_{12} , with $A_1 = A_2 = 1$, $N = 256$, $f_1 / \Delta f = 40.2$, random phases, and SNR = 40 dB.

4.2. Sensitivity to the Tone–Amplitude Ratio

In Figure 5 the trends of the MSE in the estimation of the bin deviations for a two-tone signal with very close frequencies ($d_{12} = 3$) and with random phases are reported, versus the amplitude of the second tone (β_2 changes in the range $[0.1, 2]$, while $\beta_1 = 1$) for two different SNRs (5 dB and 40 dB). The figures show only the performance of the parametric method ESPRIT and the non-parametric algorithm based on IFFT, since the results of MUSIC are very similar to those of ESPRIT, while IWPA introduces very high errors in presence of phase variations.

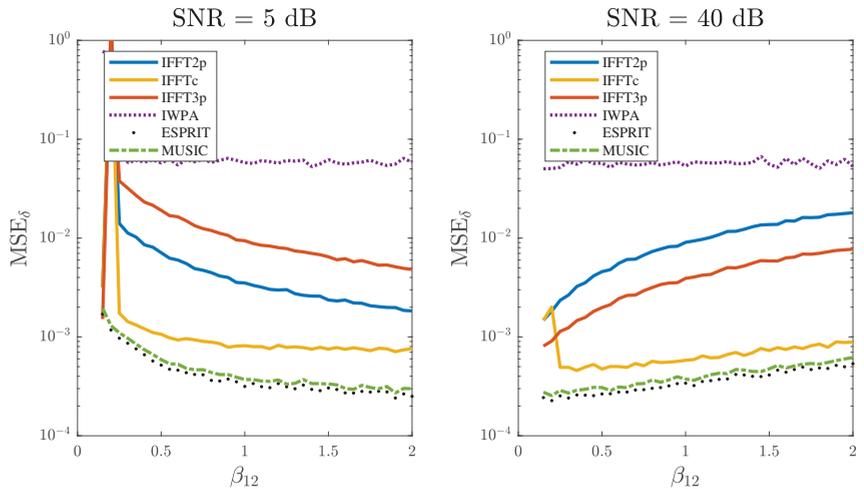


Figure 5. MSE of δ versus the amplitude of the second tone with $\beta_1 = 1$ fixed at two different SNR values 5 dB (on the left) and 40 dB (on the right). $N = 256$, $f_1 / \Delta f = 40.2$, $d_{12} = 3$, and random phases.

ESPRIT algorithm exhibits worse performance in the estimation of the the second tone frequency when β_2 is low, due to the low values of SNR at the second tone especially in the case of the lowest of the two SNR values (5 dB), while the MES value decreases for increasing amplitudes of the second tone. The estimation of the highest tone is not influenced by the amplitude of the lowest one. On the other hand, the IFFTc method is slightly influenced by the change in amplitude. Moreover, the variability obtained with all the methods on δ_i is comparable (IFFTc is characterized by a standard deviation σ_{δ_i} a bit greater than the others) and the same behaviour is observed when the second tone amplitude becomes significantly greater than the noise ($\beta_2 > 0.5$).

4.3. Sensitivity to the Number of Samples

In Figure 6 the standard deviations of the errors versus N are reported, for a given signal with two tones of the same amplitude and for two different noise levels. As expected, the standard deviations decrease when the number of processed samples increases. For both noise levels, the effect of the tone distance is less significant for high N . In the case of the lowest SNR value (5 dB), the trend is quite the same for all the methods, since the variability due to noise is comparable to the systematic effect of IFFT; meanwhile, for the highest SNR value (40 dB), the parametric algorithms and IFFTc show better performances.

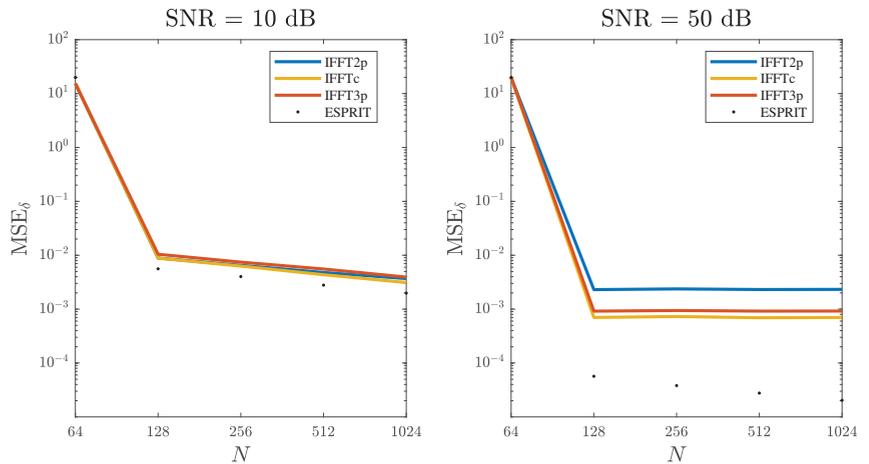


Figure 6. MSE of δ versus the number of processed samples for a two-tone signal with $A_1 = A_2 = 1$, $d_{12} = 3$, $N = 256$, $f_1/\Delta f = 40.2$, random phases, SNR 10 dB (on the left), and 50 dB (on the right).

5. Uncertainty Evaluation

As evidenced in Section 3, all the analysed methods present a residual error, that can be negligible or not function depending on the signal characteristics and the processing parameters. The residual contribution cannot be corrected since it strictly depends on the signal characteristics and the uncertainty evaluation has to be taken into account. To this aim, it is possible to write the following:

$$\delta = \delta_m + C_\delta, \tag{14}$$

where δ is the corrected bin value, δ_m is the evaluated bin value, and C_δ is the correction that can be modelled as random variable with mean value equal to zero and standard deviation, σ_C , different from zero. Applying the ISO GUM [27] the measurement uncertainty is equal to:

$$u_\delta = \sqrt{\sigma_\delta^2 + \sigma_C^2}. \tag{15}$$

As for the standard deviation of the correction value, the measurement uncertainty can be estimated considering group of signals with similar characteristics. In Table 1 the measurement uncertainty u_s on the tone frequencies evaluated for a two-tone signal with tones at varying distance (used as the index of the table) and for different values of β_{12} is reported. The uncertainty is evaluated considering for each configuration 1000 simulations with random phase and varying d_{12} between the 2 tones, and the FFT is made on 256 samples. The measurement uncertainty is reported for the six considered methods. By looking at these data, it is possible to have an idea of the order of magnitude of the uncertainty, given the signal characteristics d_1, d_2, β_{12} for a given algorithm; the uncertainty of the second tone for d_{12} between 3 and 4 when β_{12} is equal to 0.01 are not reported because it is not ever correctly detected with the algorithms based on FFT.

Table 1. Measurement uncertainty, u_s , evaluated for two-tone signals with tones at varying distance for different values of β_{12} , changing d_{12} from d_1 to d_2 . The simulations were repeated 1000 times, randomizing the tone phases with a 256-sample signal.

		$\beta_{12} = 0.01$											
d_1	d_2	Tone 1						Tone 2					
		IFFT2p	IFFTc	IFFT3p	IWPA	ESPRIT	MUSIC	IFFT2p	IFFTc	IFFT3p	IWPA	ESPRIT	MUSIC
3	4	6.7×10^{-5}	1.5×10^{-4}	2.4×10^{-5}	1.1×10^{-3}	1.1×10^{-14}	1.6×10^{-7}	-	-	-	1.1×10^{-1}	9.6×10^{-13}	1.6×10^{-7}
4	5	3.0×10^{-5}	6.0×10^{-5}	9.1×10^{-6}	1.1×10^{-3}	1.1×10^{-14}	1.8×10^{-7}	6.6×10^{-1}	5.8×10^{-1}	6.7×10^{-1}	1.0×10^{-1}	9.3×10^{-13}	1.8×10^{-7}
5	6	1.6×10^{-5}	1.0×10^{-5}	4.3×10^{-6}	1.1×10^{-3}	1.1×10^{-14}	1.9×10^{-7}	1.2×10^{-1}	3.6×10^{-5}	5.5×10^{-2}	1.0×10^{-1}	9.9×10^{-13}	1.9×10^{-7}
6	7	9.5×10^{-6}	3.4×10^{-6}	2.3×10^{-6}	1.1×10^{-3}	1.0×10^{-14}	1.6×10^{-7}	7.7×10^{-2}	3.2×10^{-5}	2.6×10^{-2}	1.0×10^{-1}	9.9×10^{-13}	1.7×10^{-7}
7	12	3.5×10^{-6}	5.6×10^{-7}	6.7×10^{-7}	1.1×10^{-3}	1.0×10^{-14}	1.7×10^{-7}	2.9×10^{-2}	2.8×10^{-5}	7.4×10^{-3}	1.0×10^{-1}	9.4×10^{-13}	1.7×10^{-7}
12	20	7.8×10^{-7}	1.8×10^{-7}	9.4×10^{-8}	1.1×10^{-3}	1.0×10^{-14}	1.7×10^{-7}	6.5×10^{-3}	2.2×10^{-5}	9.8×10^{-4}	9.7×10^{-2}	9.3×10^{-13}	1.7×10^{-7}
		$\beta_{12} = 0.1$											
d_1	d_2	Tone 1						Tone 2					
		IFFT2p	IFFTc	IFFT3p	IWPA	ESPRIT	MUSIC	IFFT2p	IFFTc	IFFT3p	IWPA	ESPRIT	MUSIC
3	4	6.8×10^{-4}	2.3×10^{-4}	2.4×10^{-4}	1.3×10^{-3}	1.1×10^{-14}	1.6×10^{-7}	5.3×10^{-2}	2.3×10^{-4}	3.5×10^{-2}	1.0×10^{-1}	1.2×10^{-14}	1.6×10^{-7}
4	5	3.0×10^{-4}	3.8×10^{-5}	9.1×10^{-5}	1.2×10^{-3}	1.1×10^{-14}	1.8×10^{-7}	2.4×10^{-2}	4.1×10^{-5}	1.2×10^{-2}	1.0×10^{-1}	1.2×10^{-14}	1.8×10^{-7}
5	6	1.6×10^{-4}	9.9×10^{-6}	4.3×10^{-5}	1.2×10^{-3}	1.1×10^{-14}	1.9×10^{-7}	1.3×10^{-2}	1.2×10^{-5}	5.2×10^{-3}	1.0×10^{-1}	1.2×10^{-14}	1.9×10^{-7}
6	7	9.5×10^{-5}	3.4×10^{-6}	2.3×10^{-5}	1.2×10^{-3}	1.1×10^{-14}	1.6×10^{-7}	7.8×10^{-3}	5.3×10^{-6}	2.6×10^{-3}	1.0×10^{-1}	1.3×10^{-14}	1.7×10^{-7}
7	12	3.6×10^{-5}	6.0×10^{-7}	7.0×10^{-6}	1.1×10^{-3}	1.1×10^{-14}	1.7×10^{-7}	3.1×10^{-3}	3.0×10^{-6}	7.9×10^{-4}	9.3×10^{-2}	1.2×10^{-14}	1.7×10^{-7}
12	20	1.9×10^{-6}	1.9×10^{-7}	1.5×10^{-7}	1.1×10^{-3}	1.1×10^{-14}	1.6×10^{-7}	1.6×10^{-4}	1.4×10^{-6}	1.5×10^{-5}	9.7×10^{-2}	1.3×10^{-14}	1.6×10^{-7}
		$\beta_{12} = 1.0$											
d_1	d_2	Tone 1						Tone 2					
		IFFT2p	IFFTc	IFFT3p	IWPA	ESPRIT	MUSIC	IFFT2p	IFFTc	IFFT3p	IWPA	ESPRIT	MUSIC
3	4	6.8×10^{-3}	2.3×10^{-4}	2.4×10^{-3}	4.0×10^{-3}	1.1×10^{-14}	1.6×10^{-7}	5.4×10^{-3}	2.3×10^{-4}	3.5×10^{-3}	1.0×10^{-1}	1.1×10^{-14}	1.6×10^{-7}
4	5	3.0×10^{-3}	3.8×10^{-5}	9.1×10^{-4}	4.2×10^{-3}	1.0×10^{-14}	1.8×10^{-7}	2.4×10^{-3}	4.1×10^{-5}	1.2×10^{-3}	1.0×10^{-1}	1.1×10^{-14}	1.8×10^{-7}
5	6	1.6×10^{-3}	9.9×10^{-6}	4.3×10^{-4}	2.9×10^{-3}	1.1×10^{-14}	1.9×10^{-7}	1.3×10^{-3}	1.1×10^{-5}	5.2×10^{-4}	1.0×10^{-1}	1.1×10^{-14}	1.9×10^{-7}
6	7	9.5×10^{-4}	3.4×10^{-6}	2.3×10^{-4}	3.0×10^{-3}	1.1×10^{-14}	1.7×10^{-7}	7.8×10^{-4}	4.0×10^{-6}	2.6×10^{-4}	1.0×10^{-1}	1.1×10^{-14}	1.7×10^{-7}
7	12	3.6×10^{-4}	7.3×10^{-7}	7.0×10^{-5}	2.2×10^{-3}	1.1×10^{-14}	1.7×10^{-7}	3.1×10^{-4}	8.5×10^{-7}	7.9×10^{-5}	9.3×10^{-2}	1.0×10^{-14}	1.7×10^{-7}
12	20	7.4×10^{-5}	3.0×10^{-7}	9.4×10^{-6}	1.6×10^{-3}	1.1×10^{-14}	1.7×10^{-7}	6.5×10^{-5}	2.4×10^{-7}	9.7×10^{-6}	9.6×10^{-2}	1.0×10^{-14}	1.7×10^{-7}

In order to verify the proposed approach, Table 1 is used to evaluate the expected uncertainty for three different signals that is compared with the measured one, evaluated with a type B approach. The analysed signals refer to different conditions: close-frequency tones (d_{12}), one of these with significantly lower amplitude (β_{12}); low-noise (Case 1 and Case 2); tones of the same amplitude with high noise (Case 3); tones with a high enough SNR (Case 4). In the first case, a two-tone signal with $\beta_{12} = 0.1, d_{12} = 3.6$, and SNR = 40 dB has been used; Case 2 reports the same kind of signal with $\beta_{12} = 0.1, d_{12} = 4.5$, and SNR = 80 dB; for Case 3, $\beta_{12} = 1, d_{12} = 5.2$, and SNR = 10 dB; meanwhile, in the last case, the signal uses the parameters $\beta_{12} = 1, d_{12} = 7.9$, and SNR = 60 dB.

In Table 2 the uncertainty of both tones is synthesized for the three algorithms—IFFTc, IFFT3p, and ESPRIT. Generally, one or two digits are enough to express the uncertainty value; however, in Table 2, more digits are used to clearly highlight the differences between the reported methods. It can be seen that there is, for all the signals, high similarity between the measured and the expected uncertainties. Even under different conditions, where the uncertainty components—due to the residual error and the noise—have different contributions, in all cases, the estimation of the uncertainty is accurate and can be an a priori alternative to the measured value. A little overestimation for the IFFT3p algorithm is

observed for Case 2, when the contribution—due to the residual error—is prevalent; this is due to the high dependence of the residual error on the tone frequency value, but in our estimation, a medium value is considered.

Table 2. Comparison of the expected uncertainty and the measured uncertainty for three different cases of a two-tone signal with changing parameters: β_{12} , d_{12} , and SNR.

		Case 1			Case 2			Case 3			Case 4	
		$\beta_{12} = 0.1, d_{12} = 3.6,$			$\beta_{12} = 0.1, d_{12} = 4.5,$			$\beta_{12} = 1.0, d_{12} = 5.2,$			$\beta_{12} = 1.0, d_{12} = 7.9,$	
		SNR = 40 dB			SNR = 80 dB			SNR = 10 dB			SNR = 60 dB	
		IFFT3p	ESPRIT	IFFTc	IFFT3p	ESPRIT	IFFTc	IFFT3p	ESPRIT	IFFTc	IFFT3p	ESPRIT
u_{d_1}	meas.	1.03×10^{-2}	1.72×10^{-2}	6.29×10^{-3}	6.99×10^{-3}	8.88×10^{-3}	4.25×10^{-3}	8.98×10^{-3}	1.03×10^{-2}	5.43×10^{-3}	2.84×10^{-2}	3.16×10^{-2}
	exp.	1.03×10^{-2}	1.72×10^{-2}	6.28×10^{-3}	6.99×10^{-3}	7.52×10^{-3}	4.24×10^{-3}	8.97×10^{-3}	1.03×10^{-2}	5.43×10^{-3}	2.84×10^{-2}	3.16×10^{-2}
u_{d_2}	meas.	2.96×10^{-3}	3.52×10^{-2}	1.74×10^{-3}	1.58×10^{-3}	2.05×10^{-3}	1.01×10^{-3}	1.28×10^{-3}	1.44×10^{-3}	6.92×10^{-4}	5.26×10^{-3}	6.52×10^{-3}
	exp.	2.98×10^{-3}	3.27×10^{-3}	1.74×10^{-3}	1.58×10^{-3}	1.22×10^{-2}	1.01×10^{-3}	1.28×10^{-3}	1.20×10^{-3}	6.93×10^{-4}	5.26×10^{-3}	6.50×10^{-3}

It is almost possible to observe an invariability of the uncertainty on both the first and second tone frequencies at the various conditions for the IWPA, ESPRIT, and MUSIC algorithms, with the same order of magnitude for both the tones for a given algorithm. The ESPRIT algorithm shows again the lowest uncertainty compared with the other parametric algorithms; the IWPA shows the worst performance in all cases. The IWPA shows better performance compared with non-parametric algorithms in almost no cases. For higher ratio of d_{12} the performance in terms of measurement uncertainty on the second tone of the non-parametric algorithms starts to be two orders of magnitude better than the IWPA algorithm. Only in the case of a low ration β_{12} and low d_{12} IWPA could be considered a good choice with respect to a non-parametric algorithm. Comparing the algorithms based on FFT, the IFFTc is able to correct the effect of the interfering tone almost in all cases (see tone 2 uncertainty with β_{12} less than 1).

6. Concluding Remarks

By comparing the obtained results, some useful conclusions can be drawn in order to guide a designer in choosing a method for the spectral analysis. Methods belonging to the class of parametric algorithms require a priori knowledge and allow the accurate estimation of the frequency only, but on the other hand, their performance is remarkable, with respect to the non-parametric algorithms. Specifically, ESPRIT shows very high performance even with a relatively small number of samples, so if there is any constraint on the number of samples, then the ESPRIT algorithm can be suggested as an optimal choice. If the number of tones (N_s) is precisely known, ESPRIT is not affected by systematic errors and is slightly affected by harmonic interference. It has excellent performance in the cases of high SNR values. On the other hand, the execution times are acceptable only when the autocorrelation matrix has a reduced size.

With the use of the proposal in Table 1, it is possible to estimate uncertainty a priori for numerous real-world conditions, without the need of extensive simulation, field acquisition, or data elaboration that needs expensive equipment or requires long time to be executed. Compared with the use of parametric approaches—where the need of a priori knowledge is indispensable in obtaining the optimal performance—this approach gives an indication of how good the result will be under certain circumstances.

As far as the other parametric approaches are considered, the performance of the MUSIC algorithm can be compared with that of the ESPRIT method, but its systematic effects are worse than those of ESPRIT when the noise level is low. Due to its zero searching strategy, the IWPA method achieves the worst performance in the estimation of frequencies in the presence of phases difference between the tones. Among the considered non-parametric algorithms, IFFTc shows the best behaviour because it achieves a decent trade-off between metrological performance and elaboration times; the IFFT algorithm is the fastest one, but in the presence of harmonic interference, the residual error is significant.

In conclusion, IFFTc is the best choice for real-time applications whenever the elaboration time is a strong requirement, but if there are constraints on the number of samples, then ESPRIT should be chosen. Furthermore, hybrid solutions—based on a pre-processing algorithm for a preliminary estimation of the signal tones and the superimposed noise, followed by a decision algorithm to select the signal processing algorithm—could be taken into account to allow the minimum uncertainty on the frequency evaluation, and to obtain the best trade-off for different configurations of tone number, SNR ratio, required spectral resolution, and real-time needs; the latter are strictly associated with the analysed bandwidth.

Author Contributions: All authors have equally contributed to the article drafting, in particular: conceptualization, C.L.; methodology V.P.; software, S.D.I.; validation, G.D.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Wen, H.; Li, C.; Yao, W. Power System Frequency Estimation of Sine-Wave Corrupted with Noise by Windowed Three-Point Interpolated DFT. *IEEE Trans. Smart Grid* **2018**, *9*, 5163–5172. [[CrossRef](#)]
2. Jin, T.; Zhang, W. A Novel Interpolated DFT Synchrophasor Estimation Algorithm with an Optimized Combined Cosine Self-Convolution Window. *IEEE Trans. Instrum. Meas.* **2021**, *70*, 1–10. [[CrossRef](#)]
3. Song, X.; Li, X.; Zhang, W.G.; Zhou, W. The new measurement algorithm of the engine speed base on the basic frequency of vibration signal. In Proceedings of the 2010 International Conference on Computer, Mechatronics, Control and Electronic Engineering, Changchun, China, 24–26 August 2010; Volume 5, pp. 273–277. [[CrossRef](#)]
4. Betta, G.; Liguori, C.; Pietrosanto, A. A multi-application FFT analyzer based on a DSP architecture. *IEEE Trans. Instrum. Meas.* **2001**, *50*, 825–832. [[CrossRef](#)]
5. Ugwiri, M.; Carratu, M.; Paciello, V.; Liguori, C. Spectral negentropy and kurtogram performance comparison for bearing fault diagnosis. In Proceedings of the 17th IMEKO TC 10 and EUROLAB Virtual Conference “Global Trends in Testing, Diagnostics and Inspection for 2030”, Dubrovnik, Croatia, 19–22 October 2020; Number 165411. pp. 105–110.
6. Kim, B.S.; Jin, Y.; Lee, J.; Kim, S. High-Efficiency Super-Resolution FMCW Radar Algorithm Based on FFT Estimation. *Sensors* **2021**, *21*, 4018. [[CrossRef](#)] [[PubMed](#)]
7. Wang, X.; Jiang, J.; Duan, F.; Liang, C.; Li, C.; Sun, Z.; Lu, R.; Li, F.; Xu, J.; Fu, X. A method for enhancement and automated extraction and tracing of Odontoceti whistle signals base on time-frequency spectrogram. *Appl. Acoust.* **2021**, *176*, 107698. [[CrossRef](#)]
8. Chien, Y.R.; Wu, C.H.; Tsao, H.W. Automatic Sleep-Arousal Detection with Single-Lead EEG Using Stacking Ensemble Learning. *Sensors* **2021**, *21*, 6049. [[CrossRef](#)] [[PubMed](#)]
9. Betta, G.; Liguori, C.; Paolillo, A.; Pietrosanto, A. A DSP-based FFT-analyzer for the fault diagnosis of rotating machine based on vibration analysis. *IEEE Trans. Instrum. Meas.* **2002**, *51*, 1316–1322. [[CrossRef](#)]
10. Schoukens, J.; Pintelon, R.; Van Hamme, H. Schoukens, J.; Pintelon, R.; Van Hamme, H. The interpolated fast Fourier transform: A comparative study. *IEEE Trans. Instrum. Meas.* **1992**, *41*, 226–232. [[CrossRef](#)]
11. Agrez, D. Weighted multipoint interpolated DFT to improve amplitude estimation of multifrequency signal. *IEEE Trans. Instrum. Meas.* **2002**, *51*, 287–292. [[CrossRef](#)]
12. Belega, D.; Petri, D. Frequency estimation by two- or three-point interpolated Fourier algorithms based on Cosine windows. *Signal Process.* **2015**, *117*, 115–125. [[CrossRef](#)]
13. Kumar, B.R.; Mohapatra, A.; Chakrabarti, S. Combined Two-Point and Three-Point Interpolated DFT for Frequency Estimation. In Proceedings of the 2020 21st National Power Systems Conference (NPSC), Gandhinagar, India, 17–19 December 2020; pp. 1–6. [[CrossRef](#)]
14. Wang, K.; Wen, H.; Li, G. Accurate Frequency Estimation by Using Three-Point Interpolated Discrete Fourier Transform Based on Rectangular Window. *IEEE Trans. Ind. Inform.* **2021**, *17*, 73–81. [[CrossRef](#)]
15. Wu, K.; Ni, W.; Andrew Zhang, J.; Liu, R.P.; Jay Guo, Y. Refinement of Optimal Interpolation Factor for DFT Interpolated Frequency Estimator. *IEEE Commun. Lett.* **2020**, *24*, 782–786. [[CrossRef](#)]
16. Liguori, C.; Paolillo, A.; Pignotti, A. An intelligent FFT analyzer with harmonic interference effect correction and uncertainty evaluation. *IEEE Trans. Instrum. Meas.* **2004**, *53*, 1125–1131. [[CrossRef](#)]

17. Liguori, C.; Paolillo, A. Implementing uncertainty auto-evaluation capabilities on an intelligent FFT-analyzer. In Proceedings of the 19th IEEE Instrumentation and Measurement Technology Conference (IEEE Cat. No.00CH37276) (IMTC/2002), Anchorage, AK, USA, 21–23 May 2002; Volume 1, pp. 657–662. [CrossRef]
18. Liguori, C.; Paciello, V.; Paolillo, A. Parameter estimation of spectral components of a signal: Comparison of techniques. In Proceedings of the 2007 IEEE Instrumentation Measurement Technology Conference IMTC 2007, Warsaw, Poland, 1–3 May 2007; pp. 1–6. [CrossRef]
19. Kusuma, J. Parametric frequency estimation: Esprit and music. Rice University, 2002. Available online: <https://cnx.org/contents/WIub0IP5@4/Parametric-frequency-estimation> (accessed on 22 February 2022).
20. Pisarenko, V.F. The Retrieval of Harmonics from a Covariance Function. *Geophys. J. Int.* **1973**, *33*, 347–366. [CrossRef]
21. Schmidt, R. Multiple emitter location and signal parameter estimation. *IEEE Trans. Antennas Propag.* **1986**, *34*, 276–280. [CrossRef]
22. Roy, R.; Kailath, T. ESPRIT-estimation of signal parameters via rotational invariance techniques. *IEEE Trans. Acoust. Speech Signal Process.* **1989**, *37*, 984–995. [CrossRef]
23. Roy, R.; Paulraj, A.; Kailath, T. Estimation of Signal Parameters via Rotational Invariance Techniques—ESPRIT. In Proceedings of the MILCOM 1986—IEEE Military Communications Conference: Communications-Computers: Teamed for the 90's, Monterey, CA, USA, 5–9 October 1986; Volume 3, pp. 41.6.1–41.6.5. [CrossRef]
24. Roy, R.; Paulraj, A.; Kailath, T. ESPRIT—A subspace rotation approach to estimation of parameters of cisoids in noise. *IEEE Trans. Acoust. Speech, Signal Process.* **1986**, *34*, 1340–1342. [CrossRef]
25. Kay, S.M. *Fundamentals of Statistical Signal Processing: Estimation Theory*; Prentice-Hall PTR: Upper Saddle River, NJ, USA, 1993.
26. Gabriele D'Antona, A.F. *Digital Signal Processing for Measurement Systems: Theory and Applications*; Springer Nature: New York, NY, USA, 2005.
27. Williams, J.H. Guide to the Expression of Uncertainty in Measurement (the GUM). In *Quantifying Measurement*; Morgan and Claypool Publishers: San Rafael, CA, USA, 2016; pp. 6-1–6-9. [CrossRef]

Article

Three-Dimensional Point Cloud Task-Specific Uncertainty Assessment Based on ISO 15530-3 and ISO 15530-4 Technical Specifications and Model-Based Definition Strategy

Gorka Kortaberria *, Unai Mutilba, Sergio Gomez and Brahim Ahmed

Department of Mechanical Engineering, Tekniker Research Centre, 20600 Eibar, Spain

* Correspondence: gorka.kortaberria@tekniker.es

Abstract: Data-driven manufacturing in Industry 4.0 demands digital metrology not only to drive the in-process quality assurance of manufactured products but also to supply reliable data to constantly adjust the manufacturing process parameters for zero-defect manufacturing processes. Better quality, improved productivity, and increased flexibility of manufacturing processes are obtained by combining intelligent production systems and advanced information technologies where in-process metrology plays a significant role. While traditional coordinate measurement machines offer strengths in performance, accuracy, and precision, they are not the most appropriate in-process measurement solutions when fast, non-contact and fully automated metrology is needed. In this way, non-contact optical 3D metrology tackles these limitations and offers some additional key advantages to deploying fully integrated 3D metrology capability to collect reliable data for their use in intelligent decision-making. However, the full adoption of 3D optical metrology in the manufacturing process depends on the establishment of metrological traceability. Thus, this article presents a practical approach to the task-specific uncertainty assessment realisation of a dense point cloud data type of measurement. Finally, it introduces an experimental exercise in which data-driven 3D point cloud automatic data acquisition and evaluation are performed through a model-based definition measurement strategy.

Keywords: uncertainty assessment; three-dimensional point clouds; ISO 15530; data-driven metrology; model-based definition; virtual twin

Citation: Kortaberria, G.; Mutilba, U.; Gomez, S.; Ahmed, B. Three-Dimensional Point Cloud Task-Specific Uncertainty Assessment Based on ISO 15530-3 and ISO 15530-4 Technical Specifications and Model-Based Definition Strategy. *Metrology* **2022**, *2*, 394–413. <https://doi.org/10.3390/metrology2040024>

Academic Editor: Simona Salicone

Received: 22 July 2022

Accepted: 22 September 2022

Published: 27 September 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Metrology is considered a fundamental tool in the context of Industry 4.0, where reliable data are needed to realise data-driven manufacturing strategies [1–3]. As far as metrology is moving from the lab to the shop floor where the manufacturing of goods takes place, it is breaking the stigma of non-productive activity and gaining a position as an enabling technology that adds value to every step of the production process [4]. This perception is becoming more evident in Industry 4.0, where measurement data from several sensors are required, including dimensional data, for the monitoring of complete manufacturing processes and real-time adjustment of process parameters, including the creation and use of metrological digital twins [2,5–9].

Massive integration of 3D optical sensors within manufacturing processes is occurring nowadays, replacing traditional Coordinate Measurement Machines (CMM) within the automotive, aerospace and power generation industries, among the leading industries in the adoption of MBD [10]. However, while the delivery of millions of points in a matter of seconds is assumed by 3D optical sensors, the process of automatically converting dense data into meaningful information and assuring the quality of these data remains a challenge [11].

This research article presents a practical approach to addressing both challenges. While the process of converting dense data into meaningful information is solved through

a Quality Information Framework (QIF)–Model-based Definition (MBD) based measurement post-processing strategy, the assurance of the quality of the data that relies on the establishment of metrological traceability is assessed by the combination of the ISO 15530-3 and ISO 15530-4 technical specifications through which the establishment of metrological traceability, which requires (a) evaluation of the measurement uncertainty and (b) the realisation of an unbroken chain of calibrations to relate a measurement result to a reference value [12], is realised. Thus, the article introduces a task-specific uncertainty assessment of a dense point cloud type of data acquisition in the absence of reliable numerical simulation models for optical systems.

Considering the evaluation of the measurement uncertainty, the *Guide to the Expression of Uncertainty in Measurement* (GUM) JCGM 100:2008 [13] establishes general rules for evaluating and expressing uncertainty in measurements that are intended to be applied to a broad spectrum of measurements. The GUM-proposed general measurement procedure seems to be clear and easy to adopt but it can be extremely difficult to implement when a complex measurement system is evaluated. As stated by Dury et al., in the broad study about 3D optical systems characterisation performed in the National Physical Laboratory (NPL) “National FreeForm centre” [11,14–16], there are many potential uncertainty error sources such as the light condition, measurand surface properties, system orientation and resolution, ambient temperature, measurement volume, chromatic effects, etc. that complicate to a high extent the reliable characterisation of those systems.

Compared with traditional CMMs, 3D optical systems are a relatively new technology, and their measurement error sources are still being researched. Even though the German guideline for optical 3D measuring systems, the VDI/VDE 2634 series (parts 2 and 3) [17,18], attempts to provide a procedure for comparing the performance of different systems for the acceptance and re-verification of these systems, it does not consider all the potential uncertainty sources while operating in unfavourable environments. Therefore, the lack of measurement procedures to fully understand how 3D optical systems behave under different measurement scenarios limits to a high extent the development of mathematical modelling for those systems [11], and therefore, the development of a digital metrology twin.

The challenge of converting dense data into meaningful information in a matter of seconds involves providing real-time automatic decision-making capability and therefore constantly adjusting process parameters for a zero-defect manufacturing scenario. However, when a 3D optical system is integrated into a manufacturing process and captures millions of points in seconds, “faster data processing” remains a challenge. Thus, the recent publication of the ISO Standard 23952:2020 “Automation systems and integration—Quality Information Framework (QIF)—An integrated model for manufacturing quality information” [19] opens the door to real-time automatic in-line quality control. This Standard suggests a new XML Schema Definition Language that defines, organises and associates the quality and metrology information needed in manufacturing systems and therefore, it allows the effective exchange of metrology data throughout the entire manufacturing quality measurement process—from product design to inspection planning to execution to analysis and reporting. For product definition, QIF includes the ISO QIF part 3: QIF Model-based Definition (MBD) [20–22], which defines a digital data format to convey part geometry (typically called the “CAD” model) and information to be consumed by downstream manufacturing quality processes, such as Product Manufacturing Information (PMI) [21–23]. This means that MBD allows the attachment of Geometric Dimensioning and Tolerancing (GD&T) information to a CAD model, typically with full “smart” associativity, to create a semantic model. This semantic CAD model allows metrology software to automatically create either an inspection plan or decision-making results (angles, distances, GD&T tolerances, etc.) from available 3D point data. Thus, the QIF MBD information model allows converting the captured dense data into meaningful information using automatic data processing methodologies [1,8,21,24,25]. Therefore, in general terms, MBD is a digital-product model that defines the requirements and specifications of the product

and is the cornerstone for Model-Based Enterprise (MBE) since MBE uses MBD to define the product requirements and specifications instead of paper-based documents as the data source for all engineering activities, including the metrology activities during the manufacturing of the product, throughout the product lifecycle [20–23,26–28].

The state-of-the-art of uncertainty assessment to point cloud measurement shows that task-specific uncertainty assessment has not been frequently applied to dense point cloud measurements. Different approaches were suggested for the uncertainty assessment of point clouds, such as the approach introduced by Ding et al., based on spatial feature registration analysis [29]. Senin et al. suggested a method based on fitting Gaussian random fields to high-density point clouds produced by measurement repeats where the fitted field delivers a depiction of the spatial distribution of random measurement error over a part geometry [30]. Yang et al. investigated the point cloud registration step as a major uncertainty source in the laser scanning-aided aircraft assembly process [31]. Zhang et al. also appointed the reconstruction of every point cloud acquisition process as a critical uncertainty source [32]. Forbes et al. presented an uncertainty assessment method associated with the position, size and shape of point cloud data [33]. Another important approach for the uncertainty assessment of point clouds is the mathematical modelling of the measurement instruments, mainly optical systems, employed in the data acquisition process. Mohammadikaji et al. suggested an approach to categorise and model the dominant sources of uncertainty and study the probabilistic propagation of the uncertainties in a 3D inspection using laser line scanners [34]. Zhao et al. suggested the use of a structured light system including the instrument itself, data acquisition, data processing, and other factors as a black model for the uncertainty assessment of 3D point clouds [35]. Some researchers also presented experimental methods to model the systematic errors pertinent to laser scanners [36,37]. Xi et al. suggested various scanner-to-surface distances and inclination angles raise systematic uncertainties for optical sensors [38,39]. Finally, the use of physical artifacts combined with a Design of Experiment (DOE) method was also suggested for the uncertainty assessment of optical systems [40–45].

2. Methods

2.1. Practical Approaches to the Uncertainty Assessment within Production Metrology

In cases where potential uncertainty sources for a measurement process can be ascertained, it is relatively easy to follow the prescription of the GUM JCGM 100:2008 [13] uncertainty framework. However, this is not the case for CMMs or 3D optical systems, in which it is extremely difficult to understand how every potential uncertainty source affects the final result. In these cases, different approaches were applied to estimate the uncertainty of the coordinate measurement. In the case of CMMs, the prevailing guidance for users is given in the ISO 15530 technical specifications. While part 1 is very informative and tutorial but not intended to provide operative evaluation tools, parts 3 and 4 are the procedures followed by the manufacturing industry for the uncertainty assessment of coordinate measurement [46,47]. While Section 3 defines an experimental comparison method using a calibrated workpiece, Section 4 suggests a computer simulation approach to provide task-specific uncertainty assessment. The project “Evaluating the Uncertainty in Coordinate Measurement” (EUCOM—under grant agreement n° 17NRM03) project within the European Metrology Programme for Innovation and Research (EMPIR) program has performed the research to develop the two missing parts of the ISO 15530 series: part 2 on a repetition and reversal method and part 5 on a method based on prior information and expert judgement.

In the case of 3D optical systems, system manufacturers employ VDI 2634 parts 2 and 3 [17,18] to characterise and run the product acceptance test before product delivery, but this does not mean that complete system characterisation is performed for a robust measurement uncertainty assessment.

2.1.1. ISO 15530-3 Technical Specification

The ISO 15530-3 [46] technical specification is a substitution method that simplifies the uncertainty evaluation exercise through the similarity between the dimensions and shapes of the workpiece and one calibrated reference part. It is based on a statistical evaluation of the measurement errors observed concerning the calibrated value of the reference part. The user must perform a relevant number (>20) of measurements under various conditions that they might expect while measuring real workpieces. This approach appears to be straightforward from the viewpoint of the user and attempts to cover intrinsic and extrinsic uncertainty contributors. However, in practice, it is fraught with difficulties. Any divergence between the master and measured parts can lead to uncertainties. Because of the similarity requirement between the produced workpiece and the calibrated standard, this approach is very arduous and expensive for large-scale metrology, where the storage, maintenance and calibration of large components is a major expense. However, it is a reliable approach for serial production, usually for small- and medium-sized components because it is affordable to manufacture and calibrate a reference part for uncertainty assessment purposes. It is usually employed for medium-size component uncertainty assessments in CMMs or Machine Tools (MT). This approach determines four input quantities as explained below [46]:

- u_b : standard uncertainty associated with the systematic error of the measurement process;
- u_p : standard uncertainty associated with the measurement procedure;
- u_{cal} : standard uncertainty associated with the uncertainty of the workpiece calibration;
- u_w : standard uncertainty associated with material and manufacturing variations.

$$U = k * \sqrt{u_p^2 + u_{cal}^2 + u_w^2 + u_b^2} \tag{1}$$

Finally, the law of uncertainty propagation is applied to obtain the combined standard uncertainty according to GUM JCGM 100:2008 [13] and the result is multiplied by an appropriate coverage factor to yield an expanded uncertainty, according to Equation (1). Figure 1 shows the practical approach to this method.

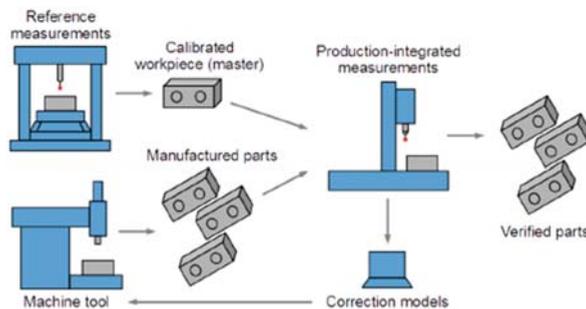


Figure 1. A practical approach to the ISO 15530-3 [5].

According to the ISO 15530-3 technical specification, the standard uncertainty (u_p) is determined using Equation (2). The standard uncertainty associated with the systematic error is given by Equation (3). Moreover, if the measurement result is not corrected by the systematic error, the error fully contributes to the uncertainty budget; thus, (u_b) = b . Thus,

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i \quad u_p = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2} \tag{2}$$

$$b = \bar{y} - x_{cal} \tag{3}$$

wherein:

- \bar{y} : mean value of the measurement results;
- Y : measured value;
- n : number of measurement results;
- x_{cal} : calibrated value of the calibrated workpiece.

2.1.2. ISO 15530-4 Technical Specifications

The ISO 15530-4 [47] technical specification introduces a “task-specific uncertainty” assessment method based on computer simulation. Measuring instruments such as CMMs and 3D optical systems are multi-purpose instruments which means that potential measurement uncertainties vary with the task being performed, environment, operator, or chosen measurement methodologies. The “task-specific uncertainty” in coordinate measurement is the measurement uncertainty that results when a specific feature is measured using a specific inspection plan. The approach is similar to GUM but instead of using an analytical approach based on a complete closed-form mathematical model, it uses a simulation method (for example, the Monte Carlo method) run on a computer to estimate the uncertainty statement for a particular measurement task. This is even more complex than the GUM approach because an initial model of the measurement instrument and process is required to run the simulation. The simulation model or virtual instrument model generates a perturbed point that represents an estimate of what a particular measurement instrument would have reported when measuring that commanded point. This process is performed several times, running as many measurements as the simulation iterations (hundreds or thousands), which enables the creation of simulation results of the measurement uncertainty.

The current state-of-the-art shows that the ISO 15530-4 approach is already being applied to measuring instruments such as CMMs or laser trackers, while modelling of optical sensors is still being developed by the research community and therefore VCMM for optical sensors is not still commercially available. The popular name for the method applied to CMMs is the so-called Virtual Coordinate Measuring Machine (VCMM) [48], which performs a point-by-point simulation of measurements, emulating the measurement strategy, measuring conditions, and physical behaviour of the CMM with dominant uncertainty contributions disturbing the measurement [48–50]. For spherical measuring instruments such as laser trackers or laser scanners, a basic spherical error model is considered in combination with a Gaussian Probability Density Function (PDF) to apply the law of uncertainty propagation. Figure 2 shows the VCMM approach, where the thick black lines show the data flow for a normal CMM measurement, while the thick grey lines show the additional data flow that is employed to achieve a VCMM estimate. Wilhelm et al. [50] presented a description of the complete VCMM workflow, as shown in Figure 2.

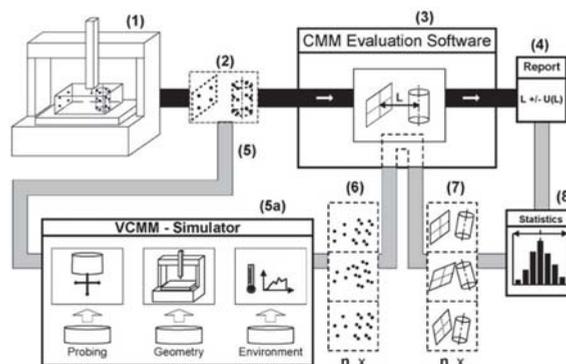


Figure 2. The VCMM approach for coordinate measuring machines [50].

The so-called Digital Metrology/Measurement Twin (D-MT) is the virtual representation of either a measurement instrument or the complete measurement procedure [51–53] and it is frequent to use a similar mathematical model to that developed within the ISO 15530-4 approach to run the simulation. Here, Artificial Intelligence (AI) algorithms such as machine learning, deep learning or neural networks are being researched for the development of those D-MT and uncertainty assessment tasks [54–56].

2.2. MBD-Based Metrology

The MBD approach will allow meeting the challenge of converting dense data into meaningful information in a matter of seconds within the production line which will allow a quick decision-making process within the production environment. However, the current state of MBD industrial implementation shows that manufacturers have applied MBD to product definition for some time, whereas aerospace and defence customers have played the role of leaders with a slower adoption in other industries [22].

From the CAD suppliers' point of view, MBD is seen as the cornerstone of creating a functioning digital thread. While the goal is to have a single source of truth for downstream operations in making a part, most CAD suppliers provide MBD in a proprietary format which means that interoperability between systems remains a challenge. Previously, a universal CAD format already exists, the ISO 10303 STEP format [57] with its accompanying AP 242 extension which includes 3D model data representation, geometric tolerance and PMI to enable global design and manufacturing collaboration [23]. However, several questions remain regarding the full definition of MBD. Standards such as ASME Y14.41 [58] and ISO 16792 [59] still exist to document how a model should be defined with annotations. These standards also help in understanding how to interpret the data within the model but the standards do not document the required amount of information that the model must contain [60].

From an MBD-based metrology point of view, MBD is allowing an automatic quality assurance workflow, allowing the automation of either the measurement program creation or the data evaluation process stages [24]. While the former is already available within the main CMM commercial software, the latter can be applied for any point cloud if the measurand MBD model and the MBD software are available. Model-based inspection has not been paid much attention to within the metrology community since the 1990s [61–64].

The MBD-based metrology process starts by creating a 3D CAD model with semantic PMI information that should be both human and machine-readable [65]. The 3D model with PMI shall contain all GD&T geometric information related to the component under measurement as well as the information related to the Bill of materials (BOM), Surface finish, weld symbols, manufacturing or measurement process plan data, metadata and notes, history of an engineering change order, legal/proprietary/export control notices and other definitive digital data [65].

The associativity between the CAD model and MBD is required to have a fully semantic smart model that allows automatic part programming and post-processing. The ability of downstream programs to read MBD models and create measurement programs is as important as creating CAD models with an attached semantic MBD. Thus, the CMM is virtually configured, and once the MBD file is imported and a set of rules is applied and matched to the configured CMM, a part program is automatically generated aided by these a priori digital approaches [24,66,67]. Typically, a second optimisation is performed to reduce the number of probe changes and minimise the CMM path length.

In the post-processing stage, the MBD concept allows fast point cloud analysis and evaluation of the measurement data. During the automatic evaluation process, the acquired point cloud is aligned to the CAD model, and an automatic segmentation process is performed using the available MBD data. At this stage, each measured point is associated with its corresponding geometric features. Then, the geometric features were adjusted using linear regression methods, rejecting possible outliers. Finally, the real relationships among the adjusted features were estimated (dimension, form error, relative positioning, etc.)

through the fully automatic interpretation and evaluation of previously defined GD&Ts. Thus, the process of converting dense data into meaningful metrology-rich information is executed automatically in seconds.

2.3. The Methodology and Its Experimental Implementation

The following lines describe the experimental exercise performed on the dummy part to realise the practical implementation of the previously mentioned technological concepts (see Figure 3). In general terms, the suggested workflow is based on three main steps: (1) measurement of the dummy part, including the data acquisition process ($\times 10$ repetitions) and automatic data process strategy (MBD) (box 1); (2) reference measurement of the dummy part on a CMM according to the ISO 15530-4 technical specification (box 2), and (3) the task-specific (GD&T) uncertainty assessment process according to the ISO 15530-3 technical specification (box 3).

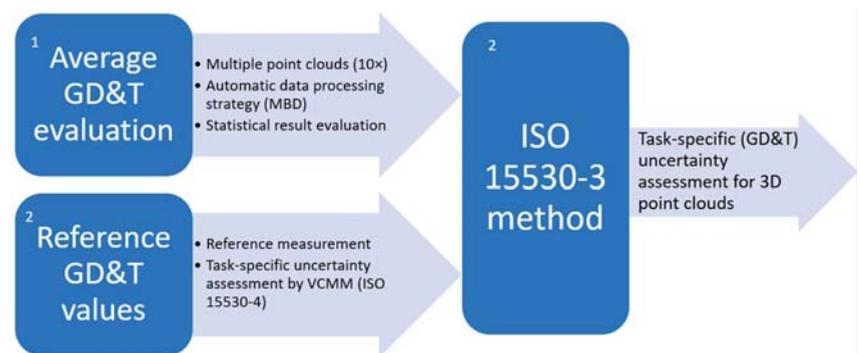


Figure 3. The point cloud measurement uncertainty assessment workflow.

The three main steps of the methodology are explained next and linked to the uncertainty contributors that comprise the uncertainty budget presented in the third step.

1. *Measured GD&T evaluation:* Automatic 3D point cloud measurement, evaluation, and statistical analysis of multiple GD&T results based on the MBD-based approach are performed. From these data, the standard uncertainty associated with the measurement process variability (u_p) is obtained.
2. *Reference GD&T values:* The dummy part is calibrated in an MMC according to the ISO 15530-4 technical specification [47]. The ZEISS VCMMTM tool is used to assess the task-specific uncertainty value for every calibrated feature. From these data, the standard uncertainty associated with the uncertainty of the MMC calibration (u_{cal}) is obtained.
3. *ISO 15530-3 method:* The task-specific uncertainty assessment of every GD&T value obtained from the 3D point cloud measurement is performed according to the ISO 15530-3 technical specification [46]. From these data, the standard uncertainty associated with the systematic error of the measurement process is obtained (u_b).

The experimental implementation of the suggested methodology is explained next point by point:

2.3.1. Measured GD&T Evaluation

A dense 3D point cloud data acquisition process is performed using a GOM ATOS III Triple ScanTM 3D optical system on a medium-sized geometric-type dummy part. The data acquisition process is fully automatic by combining an automatic rotary table with the manual triggering of the measuring instrument. Thus, any potential error source derived

from external sources, such as the alignment process between partial scanning or operator influence, is avoided.

The experimental exercise is performed in a metrology laboratory at a temperature of 20 ± 1 °C. In this way, thermal stability during the data acquisition process is guaranteed; and therefore, the geometric variation of the dummy part caused by thermal drift is avoided. Figure 4 shows the dummy part employed during the experimental exercise.



Figure 4. The medium-size dummy part from Metrologic™ is shown: (a) the CAD model and (b) the physical part placed on the rotary table.

The measurement instrument configuration is set at a working volume of $320 \times 240 \times 240$ mm³ so that the measurement resolution was optimised for the specific dummy part under measurement. The automatic data acquisition process is realised by eight angular rotary table equidistant positions, where partial scans are performed and stitched together automatically to reconstruct the overall 3D point cloud. Thus, the entire measurement process is executed within 45 s, and a point cloud comprised of 1 million points is obtained.

The fiducial targets are attached to the rotary table and unequivocally identified in each partial scan. In this manner, the automatic partial point cloud registration problem is solved, and an automatic data merging process between partial scans is performed. Once the reconstructed 3D point cloud is obtained, it is converted into a mesh using the Delaunay triangulation method. In addition to the XYZ information of every point within the point cloud, this mesh also contains information related to the surface normal value for every point which makes the final MBD-based automatic point cloud segmentation process smarter and more robust.

Figure 5 depicts the measurement scenario, comprising a GOM ATOS III Triple Scan™ 3D optical system in combination with the automatic rotary table and the dummy part on it.



Figure 5. Automatic measurement data acquisition process set-up.

The batch of experiments comprises ten 3D point cloud measurement repetitions on the dummy part. Once they are completed, an automatic data-processing approach based on the MBD strategy is applied. Thus, the dense point cloud is automatically processed and converted into meaningful GD&T information.

The software employed at this point is GOM Inspect™ metrology software. It allows MBD-type post-processing of data which means that it can digitally establish a relationship between nominal GD&T information and geometric elements within the captured data. This workflow is aligned with the PMI concept and interpreted using the ISO 1101 standard [68]. A different option to run MBD-type post-processing is to define the MBD data within the CAD model by adding the GD&T information to the available CAD file (for example, a catpart file in SolidWorks). Once this nominal MBD data-based file is prepared, the automatic evaluation of every GD&T can be performed. The workflow is as follows.

- Step 1: Point cloud-to-mesh data conversion: The measured point cloud is converted into a mesh format to make the following data management and processing steps more robust and precise. The mesh format estimates and adds the surface normal values to the point cloud format, enabling it to achieve higher accuracy results through posterior segmentation operations (step 3 below).
- Step 2: The 3D mesh is aligned with the available CAD model, which is crucial to ensure the accuracy and robustness of the MBD-based data evaluation method because it determines the correct parameterisation within the point cloud segmentation method. Thus, accurate alignment is required to achieve reliable results. In this study, the best-fit alignment method is used as an accurate method (acceptance criteria below a few microns).
- Step 3: Automatic geometric feature segmentation is performed, and the mesh is split into multiple point clouds corresponding to each geometric feature with the aid of CAD nominal feature information. In this step, the point coordinates, surface normal data (real and nominal values), and surface curvature parameters are employed to support the point cloud segmentation algorithms and reinforce their robustness.
- Step 4: Real geometric feature adjustment process: At this point, the previously obtained geometric-specific point cloud segmentation data are fitted to the corresponding geometric features by linear regression methods, rejecting possible outliers. The elimination of noisy points is established using suitable filters that estimated the 3D distance of each point concerning the fitted geometric feature. If the point-to-element distance parameter is higher than the standard deviation value (2σ) of the input points during the geometric feature adjustment process, this input point is detected as a non-suitable point and consequently removed from the process.
- Step 5: GD&T evaluation: Once the previous step is successfully performed, an automatic evaluation of every GD&T for the fitted features (measured values) is performed with the help of nominally defined annotations and relationships (ISO 1101 standard [68]). Because the software already knows the theoretical relationships among the geometric features and datum objects by the previously recognised annotations, it can estimate the real GD&T values.

Following this process, the standard uncertainty associated with the measurement process variability (u_p) is obtained for each feature. The ten available 3D point cloud measurement repetitions are statistically processed, and the uncertainty (u_p) is given by the standard deviation parameter according to the ISO 15530-3 technical specification. In addition, the average value is reported at this step for the evaluation of the systematic error (u_b) value within the next step.

Figure 6 shows the experimental results for the dummy part of the GOM Inspect™ metrology software.

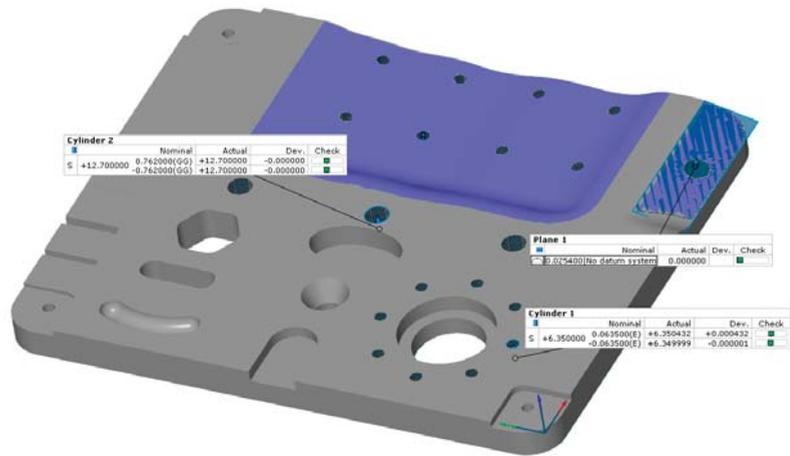


Figure 6. GD&T evaluation (real and nominal values) and representation on the CAD model.

2.3.2. Reference GD&T Values

The aim is to calibrate the measured dummy part and obtain the reference values for each evaluated feature to realise the uncertainty budget according to the ISO 15530-3 technical specification [46]. Thus, the dummy part is measured in a ZEISS UPMC 850 CARAT CMM, in which the ZEISS VCMM™ tool is available for task-specific uncertainty assessment according to the ISO 15530-4 technical specification [47]. The ZEISS VCMM™ tool considers the mathematical model of the UPMC 850 CARAT CMM to perform the task-specific uncertainty assessment process through multiple iterations ($\times 1000$ repetitions). To feed the mathematical model running within the ZEISS VCMM™ tool, the error and influence factors, as well as their variability affecting the measurement accuracy, were previously characterised and introduced into the model. Figure 7 shows the dummy calibration process.



Figure 7. Dummy part calibration on the ZEISS UPMC 850 CARAT CMM.

In this manner, a selection of the GD&T features to be considered within the experimental implementation is performed. The types of geometric elements and tolerances that were considered are explained next, according to the type of tolerance and the number of features measured.

- Size: Cylinder diameter (20× divided into groups by diameter);
- Form: Flatness–planes (3×) and complex surfaces (2×);
- Location and orientation: Positioning and composed positioning of cylinders (divided into three groups).

The positioning tolerances are evaluated considering the coordinate system created by the three planes which define the ABC datums. In total, the task-specific uncertainty of the 52 GD&T features is evaluated. Figure 8 shows the task-specific uncertainty assessment exercise based on the ZEISS UPMC CARAT 850 CMM measurements.

Name	Measured value	Nominal value	+Tol	-Tol	Deviation +/-
FLATNESS A	0.0404	0.0000	0.0508	0.0000	0.0404
FLATNESS B	0.0113	0.0000	0.0508	0.0000	0.0113
FLATNESS C	0.0186	0.0000	0.0508	0.0000	0.0186
Slot Width INF	12.7236	12.7000	0.1270	-0.1270	0.0236
Slot Length SUP	12.6970	12.7000	0.1270	-0.1270	-0.0030
Space Point Distance 1-E-...	12.7066	12.7000	0.1270	-0.1270	0.0066
Space Point Distance 2-E-...	12.7149	12.7000	0.1270	-0.1270	0.0149
Slot Width INF	40.6822	40.6721	0.1270	-0.1270	0.0101
Slot Length SUP	40.6567	40.6721	0.1270	-0.1270	-0.0153
Profile tolerance of PLANE INCL	0.1679	0.0000	0.0254	0.0000	0.1679 0.1425
PARALLELISM ESQ-A	0.0398	0.0000	0.0762	0.0000	0.0398
Diameter E CIL-1*Max.	50.8502	50.8000	0.5080	-0.5080	0.0502
Diameter E CIL-1*Min.	50.8373	50.8000	0.5080	-0.5080	0.0373

Figure 8. Task-specific uncertainty assessment based on the ZEISS CMM measurement and the ZEISS VCMM™ simulation. Example of measurement report including the VCMM outputs.

From these data, the standard uncertainty associated with the uncertainty of the MMC calibration of the dummy part (u_{cal}) is obtained.

2.3.3. Implementation of ISO 15530-3 Technical Specification

As previously stated, the uncertainty assessment method suggested in this article is based on the ISO 15530-3 technical specification application backed by the ISO 15530-4 technical specification, through which the task-specific uncertainty assessment for the calibrated values (u_{cal}) is realised.

According to the ISO 15530-3 technical specification, the uncertainty of the systematic error b (u_b) is assessed by the difference between the average value obtained during the measurement process variability (u_p) parameter assessment and the indicated value of the CMM. However, according to the GUM recommendation, the measurement results should be corrected by the amount of systematic effect. Thus, if the measurement result is not corrected by the systematic error, the error fully contributes to the uncertainty budget; thus, (u_b) = b .

The uncertainty budget presented here comprises the uncertainty contributors u_b , u_p and u_{cal} , whereas u_w is negligible because of the lack of variation between the calibrated and measured dummy. The same physical dummy part is used during the calibration and measurement processes.

Other potential uncertainty sources, such as the measuring system resolution or any divergence between the master and measured dummy parts, were discarded because of their negligible effect on the uncertainty budget.

Equation (4) shows the combined standard uncertainty, u , which is given by the quadrature sum of each uncertainty contributor. Equation (1) shows the expanded measurement uncertainty U determined with a coverage factor $k = 2$ for an approximated coverage probability of 95%.

$$u = \sqrt{u_p^2 + u_{cal}^2 + u_w^2 + u_b^2} \quad (4)$$

wherein:

u_w : standard uncertainty associated with material and manufacturing variations. This was negligible in this case.

u_{cal} : standard uncertainty associated with the uncertainty of the MMC calibration (task-specific uncertainty value of each GD&T estimated by ZEISS VCMM™)

u_b : standard uncertainty associated with systematic errors in the measurement process. It was assessed by the difference between the average value obtained during the measurement process variability (u_p) parameter assessment and the indicated value of the CMM.

u_p : to the standard uncertainty associated with the measurement process variability. The standard deviation of the ten-3D point cloud measurement repetitions was considered for each GD&T.

k : confidence interval: defines an interval with a level of confidence of approximately 95% ($k = 2$) with a normal distribution

U : expanded uncertainty for each GD&T comprising all the uncertainty error sources and their propagation with a confidence interval of 95% ($k = 2$).

Other uncertainty sources, such as thermal effects, measuring process drift, or the interaction between the light and the part surface, have not been considered separately within the uncertainty budget, as it is assumed that they contribute to (u_p) the standard uncertainty associated with the measurement process variability.

3. Results

This section describes the results obtained during the experimental implementation of the proposed uncertainty assessment method. For the sake of understanding, the experimental results are presented in such a way that every uncertainty contributor can be explained in detail. First, the standard uncertainty results associated with the measurement process variability (u_p) are presented. Then, the standard uncertainty results associated with the uncertainty of the MMC calibration (u_{cal}) are presented, along with the values indicated by the ZEISS CMM for each evaluated GD&T. Subsequently, the standard uncertainty results associated with the systematic error of the measurement process (u_b) are presented. Finally, the uncertainty budget for the 3D point cloud task-specific measurement uncertainty assessment is presented, where the expanded measurement uncertainty U is obtained for each GD&T element.

3.1. Measurement Process Variability, u_p

Table 1 lists the results obtained for the (u_p) uncertainty contribution. The result comprises every uncertainty component that falls within the measurement process variability, such as the instrument repeatability itself, MBD data processing strategy, or any potential thermal drift, among other minor contributors. Table 1 also considers the obtained mean value for every measured GD&T for further (u_b) uncertainty contributor assessment.

Table 1. (u_p) uncertainty contributor assessment according to ISO 15530-3 technical specification.

Element ID	Quantity of Evaluated Features	Evaluated Propertie	ATOS Triple Scan (Mean Values in mm)	u_p _ Standard Deviation (mm)
Plane (ref. A)	1	Flatness	0.085	0.0090
Cylinder (E min)	1	Diameter	12.664	0.0090
Cylinder (E max)	1	Diameter	12.752	0.0090
Plane (ref. B)	1	Flatness	0.100	0.0080
Plane (ref. C)	1	Flatness	0.216	0.0080
Cylinder1	1	Positioning	0.044	0.0030
Cylinder group 1. Min	1	Diameter	50.792	0.0070
Cylinder group 1. Max	1	Diameter	50.854	0.0070
Cylinder group 2. Min	15	Diameter	6.350	0.0100
Cylinder group 2. Max	15	Diameter	6.575	0.0100
Cylinder group 3. Min	3	Diameter	12.684	0.0050
Cylinder group 4. Min	3	Diameter	12.756	0.0050
Cylinder (ref. D)	1	Diameter	38.109	0.0020
Cylinder (ref. D)	1	Positioning	0.048	0.0080
Surface "LARGE"	1	Profile error	0.055	0.0020
Surface "SHORT"	1	Profile error	0.119	0.0170
Group 1 of cylinders	1	Composed positioning	0.069	0.0090
Group 2 of cylinders	15	Composed positioning	0.045	0.0070
Group 3 of cylinders	3	Composed positioning	0.101	0.0070

For every GD&T element, the highest measurement process variability (u_p) results were below 20 μm whereas the average standard deviation was less than 10 μm . These results demonstrate that the measurement process, from the data acquisition process and 3D point cloud reconstruction to the MBD-based data processing procedure, is within the micrometre accuracy.

3.2. Uncertainty of the MMC Calibration, u_{cal}

Table 2 shows the results indicated by the CMM for every GD&T under study. Thus, it shows the reference result reported by the ZEISS UPMC CARAT CMM in addition to the task-specific uncertainty value (u_{cal}) estimated using the ZEISS VCMM™ tool. In this way, the CMM measurement uncertainty for every measured GD&T is assessed, and thus, a task-specific uncertainty budget can be finally accomplished.

For every GD&T element, the highest CMM calibration uncertainty (u_{cal}) values are up to 7 μm , whereas the average value is within 1 μm . These results demonstrate that (u_{cal}) uncertainty values estimated by the ZEISS VCMM™ tool are consistent with the ZEISS UPMC CARAT CMM Maximum Permissible Error (MPE) specification ($0.6 \pm L/1000$ in μm) although some specific and complex GD&T results are much worse due to the complexity of the evaluation. At this point, the authors made a special effort to understand these results. Thus, a second batch of calibration measurements is performed on the CMM, but the results did not vary significantly, indicating that the obtained results are properly evaluated by the VCMM tool. At this point, the author hypothesises that the source of these results possibly arises from the acquired raw data and employed measuring strategies.

Table 2. (u_{cal}) uncertainty contributor according to ISO 15530-4 technical specification.

Element ID	Quantity of Evaluated Features	Evaluated Properties	ZEISS (Ref. Values in mm)	u_{cal} From VCMM (in mm)
Plane (ref. A)	1	Flatness	0.0404	0.0004
Cylinder (E min)	1	Diameter	12.697	0.0005
Cylinder (E max)	1	Diameter	12.7236	0.0005
Plane (ref. B)	1	Flatness	0.1679	0.0058
Plane (ref. C)	1	Flatness	0.0398	0.0011
Cylinder1	1	Positioning	0.0014	0.0032
Cylinder group 1. Min	1	Diameter	50.8373	0.0008
Cylinder group 1. Max	1	Diameter	50.8502	0.0006
Cylinder group 2. Min	15	Diameter	6.5203	0.0004
Cylinder group 2. Max	15	Diameter	6.5339	0.0023
Cylinder group 3. Min	3	Diameter	12.7307	0.0006
Cylinder group 4. Min	3	Diameter	12.7506	0.0005
Cylinder (ref. D)	1	Diameter	38.1195	0.0005
Cylinder (ref. D)	1	Positioning	0.0306	0.0045
Surface "LARGE"	1	Profile error	0.1815	0.0055
Surface "SHORT"	1	Profile error	0.1631	0.0066
Group 1 of cylinders	1	Composed positioning	0.0279	0.0017
Group 2 of cylinders	15	Composed positioning	0.0448	0.0042
Group 3 of cylinders	3	Composed positioning	0.1045	0.0043

3.3. Uncertainty of the Systematic Error, u_b

Table 3 lists the standard uncertainty associated with the systematic error of the measurement process (u_b). It is assessed by the difference between the average value obtained during the measurement process variability (u_p) parameter assessment and the indicated value of the CMM during the dummy calibration process.

The systematic error (u_b) results show a wide range of values. While some results are within a few microns, others are between 0.1 ÷ 0.2 mm (absolute values). The average value for all the GD/Ts is within 50 µm. At this point, it is not easy to understand and justify the wide range of values obtained for the (u_b) uncertainty contributor, but it is possibly explained by the higher point quantity and more homogeneous point distribution of the measurements obtained by the 3D optical system compared to the CMM measurements.

At this point, it makes sense to highlight, as in the introduction, that 3D optical systems are a relatively new technology, and their measurement error sources are still being researched. They are affected by many potential uncertainty error sources, such as the light condition, measurement and material properties, system orientation and resolution, ambient temperature, measurement volume, and chromatic effects, which complicate the measurement uncertainty assessment process to a large extent. Nonetheless, as stated previously within the introduction, this study aims to present a point cloud measurement task-specific uncertainty assessment method and its experimental implementation. The obtained expanded measurement uncertainty results are not as important as those of the method presented by the authors.

Table 3. (u_b) uncertainty contributor according to ISO 15530-3 technical specification.

Element ID	Quantity of Evaluated Features	Evaluated Properties	ATOS Triple Scan (Mean Values in mm)	ZEISS (Ref. Values in mm)	$u_{b_Systematic}$ Error (mm)
Plane (ref. A)	1	Flatness	0.085	0.0404	−0.0446
Cylinder (E min)	1	Diameter	12.664	12.697	0.033
Cylinder (E max)	1	Diameter	12.752	12.7236	−0.0284
Plane (ref. B)	1	Flatness	0.1	0.1679	0.0679
Plane (ref. C)	1	Flatness	0.216	0.0398	−0.1762
Cylinder1	1	Positioning	0.044	0.0014	−0.0426
Cylinder group 1. Min	1	Diameter	50.792	50.8373	0.0453
Cylinder group 1. Max	1	Diameter	50.854	50.8502	−0.0035
Cylinder group 2. Min	15	Diameter	6.35	6.5203	0.1703
Cylinder group 2. Max	15	Diameter	6.575	6.5339	−0.0407
Cylinder group 3. Min	3	Diameter	12.684	12.7307	0.0467
Cylinder group 4. Min	3	Diameter	12.756	12.7506	−0.0056
Cylinder (ref. D)	1	Diameter	38.109	38.1195	0.0105
Cylinder (ref. D)	1	Positioning	0.048	0.0306	−0.0174
Surface "LARGE"	1	Profile error	0.055	0.1815	0.1265
Surface "SHORT"	1	Profile error	0.119	0.1631	0.0441
Group 1 of cylinders	1	Composed positioning	0.069	0.0279	−0.0411
Group 2 of cylinders	15	Composed positioning	0.045	0.0448	−0.0002
Group 3 of cylinders	3	Composed positioning	0.101	0.1045	0.0035

3.4. Expanded Measurement Uncertainty, U

Table 4 summarises the uncertainty budget for the experimental implementation of the proposed method. It shows the three major uncertainty contributors and the expanded measurement uncertainty result U obtained from Equation (2). It should be noted that the measurement results are not corrected by the amount of systematic effects; therefore, the (u_b) contributor is considered within the final uncertainty budget.

The uncertainty budget shows that the systematic error contributor (u_b) is the main contributor to the final result. While the CMM calibration uncertainty (u_{cal}) contributor average value falls within 1 μm and the measurement process variability (u_p) average value is less than 10 μm , the systematic error (u_b) average value falls within 50 μm . Thus, the CMM calibration uncertainty (u_{cal}) becomes negligible, which means that the main contributors to the task-specific uncertainty budget are the measurement process variability (u_p) and systematic error contributor (u_b). As stated before, the measurement result is not corrected by the number of systematic effects which, in this case, are the main uncertainty sources for the measurement with 3D optical systems.

Table 4. (U) Expanded measurement uncertainty assessment results. Uncertainty budget.

Element ID	Quantity of Evaluated Features	Evaluated Properties	ATOS Triple Scan (Mean Values in mm)	ZEISS (Ref. Values in mm)	$u_{b_Systematic}$ Error (mm)	U_Expanded Uncertainty ($k = 2$ in mm)
Plane (ref. A)	1	Flatness	0.085	0.0404	−0.0446	0.091
Cylinder (E min)	1	Diameter	12.664	12.697	0.033	0.0684
Cylinder (E max)	1	Diameter	12.752	12.7236	−0.0284	0.0596
Plane (ref. B)	1	Flatness	0.1	0.1679	0.0679	0.1372
Plane (ref. C)	1	Flatness	0.216	0.0398	−0.1762	0.3528
Cylinder1	1	Positioning	0.044	0.0014	−0.0426	0.0856
Cylinder group 1. Min	1	Diameter	50.792	50.8373	0.0453	0.0916
Cylinder group 1. Max	1	Diameter	50.854	50.8502	−0.0035	0.0156
Cylinder group 2. Min	15	Diameter	6.35	6.5203	0.1703	0.3412
Cylinder group 2. Max	15	Diameter	6.575	6.5339	−0.0407	0.084
Cylinder group 3. Min	3	Diameter	12.684	12.7307	0.0467	0.094
Cylinder group 4. Min	3	Diameter	12.756	12.7506	−0.0056	0.015
Cylinder (ref. D)	1	Diameter	38.109	38.1195	0.0105	0.0214
Cylinder (ref. D)	1	Positioning	0.048	0.0306	−0.0174	0.0394
Surface “LARGE”	1	Profile error	0.055	0.1815	0.1265	0.2532
Surface “SHORT”	1	Profile error	0.119	0.1631	0.0441	0.0954
Group 1 of cylinders	1	Composed positioning	0.069	0.0279	−0.0411	0.0842
Group 2 of cylinders	15	Composed positioning	0.045	0.0448	−0.0002	0.0164
Group 3 of cylinders	3	Composed positioning	0.101	0.1045	0.0035	0.0178

4. Discussion

This article presents a methodology for task-specific uncertainty assessment of 3D point clouds based on ISO 15530-3 and ISO 15530-4 technical specifications and the application of MBD-based post-processing for the automatic processing of point clouds.

It presents an uncertainty budget comprising three main uncertainty contributors according to ISO 15530-3 technical specifications. The three major uncertainty contributors are (a) measurement process variability (u_p), (b) uncertainty of the CMM calibration (u_{cal}), and (c) uncertainty of the systematic error (u_b). The uncertainty associated with the material and manufacturing variations, u_w , is considered negligible.

The methodology presented here suggests an automatic 3D point cloud measurement and evaluation process, where the statistical analysis of multiple GD&T results is based on an MBD-based approach. From these data, the standard uncertainty associated with the measurement process variability (u_p) is automatically obtained. The standard uncertainty associated with the uncertainty of the MMC calibration (u_{cal}) is obtained using the ZEISS VCMM™ tool, which assesses a task-specific uncertainty value for every calibrated feature according to ISO 15530-4 technical specifications. Finally, the standard uncertainty associated with the systematic error of the measurement process (u_b) is obtained from the difference between the average value obtained during the measurement process variability and the indicated value of the CMM during the dummy part calibration process.

The experimental results show that the systematic error contribution (u_b) is the main contributor to the uncertainty budget. While the CMM calibration uncertainty (u_{cal}) contributor average value falls within 1 μm and the measurement process variability (u_p) average

value is less than 10 μm , the systematic error (u_b) average value falls within 50 μm . Thus, the CMM calibration uncertainty (u_{cal}) becomes negligible, which means that the main contributors to the task-specific uncertainty budget are the measurement process variability (u_p) and systematic error contributor (u_b).

In summary, a reliable task-specific uncertainty method is developed and successfully implemented. In the absence of numerical simulation models for optical systems, which are not currently available, this methodology allows for the establishment of an uncertainty budget to understand the order of magnitude of the measurement uncertainty of 3D optical systems.

One of the limitations of the presented methodology is scalability to large components, where CMM reference values are hardly achievable by calibrating the existing dummy part. Hence, this methodology could be applied to the scanning of geometric parts approximately up to 1.5 ÷ 2 m and manufactured in serial production, since this is the most common working range for many applications such as automotive or the manufacturing of metallic components.

Finally, concerning the MBD-based metrology data processing strategy, the experimental approach presented in this article demonstrates that the nominal PMI-based method is appropriate for converting dense point cloud data into desired dimensional metrology results (GD&Ts). It enables an effective data processing approach in terms of accuracy, speed, and robustness which in turn allows a fully automatic geometric point cloud evaluation process to avoid errors during the result interpretation and procurement processes.

Further work will focus on analysing the deviations between the results of the 3D optical system and the reference values obtained with CMM. Because these measuring technologies differ considerably in terms of accuracy, number, and point distribution, deviations will remain, but they would help to understand the complex intrinsic performance of 3D scanning systems. These preliminary results and accuracy assessment methods could support the development of AI-based numerical methods that describe the optical performance of 3D scanners.

Regarding the VCMM approach, this study demonstrates that simulation-based metrology should be applied for task-specific assessment of reference values. This shows the applicability of digital twins within the metrology field in terms of a priori uncertainty estimation and a posteriori uncertainty assessment. Thus, the measurement procedure can be optimised based on those digital twin simulation results. Another interesting future research line within the VCMM field is to employ the simulation-based metrology concept to create nominal dense reference point clouds with known uncertainty values. Therefore, fast uncertainty assessment procedures should be developed for dense point cloud data.

Author Contributions: U.M. contributed to the state-of-the-art and uncertainty budget realisation. G.K. contributed to the development of the methodology for uncertainty budget realisation. B.A. contributed to the information related to technical specifications and uncertainty budget realisation. S.G. contributed to experimental realisation. All authors have read and agreed to the published version of the manuscript.

Funding: This work was carried out under a research program called ELKARTEK 2021–2022. The project called PRECITEK (grant number KK-2021/00039) is supported by The Basque Business Development Agency.

Informed Consent Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Barbosa, C.R.H.; Sousa, M.C.; Almeida, M.F.L.; Calili, R.F. Smart Manufacturing and Digitalization of Metrology: A Systematic Literature Review and a Research Agenda. *Sensors* **2022**, *22*, 6114. [[CrossRef](#)]
2. Imkamp, D.; Berthold, J.; Heizmann, M.; Kniel, K.; Peterek, M.; Schmitt, R.; Seidler, J.; Sommer, K.-D. Challenges and trends in manufacturing measurement technology—the “Industrie 4.0” concept. *J. Sens. Sens. Syst.* **2016**, *83*, 325–335. [[CrossRef](#)]

3. Lazzari, A.; Pou, J.-M.; Dubois, C.; Leblond, L. Smart metrology: The importance of metrology of decisions in the big data era. *IEEE Instrum. Meas. Mag.* **2017**, *20*, 22–29. [CrossRef]
4. Peter de Groot. Metrology–Fuelling Industry 4.0 Data Driven Production. Metrology News. 28 June 2021. Available online: <https://metrology.news/metrology-fuelling-industry-4-0-data-driven-production/> (accessed on 21 July 2022).
5. Carmignato, S.; De Chiffre, L.; Bosse, H.; Leach, R.K.; Balsamo, A.; Estler, W.T. Dimensional artefacts to achieve metrological traceability in advanced manufacturing. *CIRP Ann.* **2020**, *69*, 693–716. [CrossRef]
6. Gohari, H.; Berry, C.; Barari, A. A Digital Twin for Integrated Inspection System in Digital Manufacturing. *IFAC-PapersOnLine* **2019**, *52*, 182–187. [CrossRef]
7. Kunzmann, H.; Pfeifer, T.; Schmitt, R.; Schwenke, H.; Weckenmann, A. Productive metrology-adding value to manufacture. *CIRP Ann.-Manuf. Technol.* **2005**, *54*, 155–168. [CrossRef]
8. Rubel, R.I.; Ali, M.H.; Akram, M.W. Role of in-process metrology in industry 4.0 smart manufacturing. *Acad. J. Manuf. Eng.* **2022**, *20*, 1–7.
9. Yandayan, T.; Burdekin, M. In-process dimensional measurement and control of workpiece accuracy. *Int. J. Mach. Tools Manuf.* **1997**, *37*, 1423–1439. [CrossRef]
10. Kuş, A. Implementation of 3D optical scanning technology for automotive applications. *Sensors* **2009**, *9*, 1967–1979. [CrossRef]
11. Dury, M.R.; Woodward, S.D.; Brown, S.B.; McCarthy, M.B. Characterising 3D optical scanner measurement performance for precision engineering. In Proceedings of the ASPE 2016 Annual Meeting, Portland, OR, USA, 23–28 October 2016; pp. 167–172.
12. Joint Committee for Guides in Metrology (JCGM). *International Vocabulary of Metrology*, 4th ed. Committee Draft. 11 January 2021.
13. Joint Committee for Guides in Metrology (JCGM). JCGM 100:2008 GUM 1995 with Minor Corrections—Evaluation of Measurement Data—Guide to the Expression of Uncertainty in Measurement. *September 2008*. Available online: <https://www.iso.org/sites/JCGM/GUM/JCGM100/C045315e.html?csnumber=50461> (accessed on 21 July 2022).
14. Dury, M.R.; Brown, S.B.; McCarthy, M.B.; Woodward, S.D. 3D optical scanner dimensional verification facility at the NPL’s “national FreeForm centre”. In Proceedings of the Laser Metrology and Machine Performance 11th International Conference and Exhibition on Laser Metrology, Machine Tool, C Robot Performance, LAMDAMAP 2015, Huddersfield, UK, 17–18 March 2015; pp. 187–197.
15. Dury, M.; Brown, S.; McCarthy, M.; Woodward, S. Blowing hot and cold: Temperature sensitivities of 3D optical scanners. In Proceedings of the 15th International Conference of the European Society for Precision Engineering and Nanotechnology, EUSPEN 2015, Leuven, Belgium, 1–5 June 2015; pp. 161–162.
16. Dury, M.; Woodward, S.; Brown, S.; McCarthy, M. Assessing fringe projector volumetric error sources using the NPL tetrahedral artefact. In Proceedings of the 16th International Conference of the European Society for Precision Engineering and Nanotechnology, EUSPEN 2016, Nottingham, UK, 30 May–3 June 2016; pp. 1–3.
17. Verein Deutscher Ingenieure (VDI). VDI/VDE 2634 Part 2: Optical 3-D measuring systems—Optical systems based on area scanning. Engl. VDI/VDE-Gesellschaft Mess- und Automatisierungstechnik. 2012, Düsseldorf. Available online: <https://standards.globalspec.com/std/9914533/vdi-vde-2634-blatt-2> (accessed on 21 July 2022).
18. Verein Deutscher Ingenieure (VDI). VDI 2634 Part 3: Optical 3D-measuring systems (Multiple view systems based on area scanning). Engl. VDI/VDE-Gesellschaft Mess- und Automatisierungstechnik 2008, Düsseldorf. Available online: <https://standards.globalspec.com/std/9914423/vdi-vde-2634-blatt-3> (accessed on 21 July 2022).
19. Digital Metrology Standards Consortium (ANSI). Quality Information Framework (QIF)—An Integrated Model for Manufacturing Quality Information. Engl. VDI/VDE-Gesellschaft Mess- und Automatisierungstechnik 2020, Düsseldorf. Available online: <https://qifstandards.org/> (accessed on 21 July 2022).
20. Goher, K.; Shehab, E.; Al-Ashaab, A. Model-Based Definition and Enterprise: State-of-the-art and future trends. *Proc. Inst. Mech. Eng. Part B J. Eng. Manuf.* **2020**, *235*, 2288–2299. [CrossRef]
21. Ram, P.S.; Lawrence, K.D. *Implementation of Quality Information Framework (QIF): Towards Automatic Generation of Inspection Plan from Model-Based Definition (MBD) of Parts*; 29 October 2020; Springer: Singapore, 2020. [CrossRef]
22. Ruemler, S.P.; Zimmerman, K.E.; Hartman, N.W.; Hedberg, T.; Feeney, A.B. Promoting model-based definition to establish a complete product definition. In Proceedings of the ASME 2016 11th International Manufacturing Science and Engineering Conference, MSEC 2016, Blacksburg, VA, USA, 27 June–1 July 2016; p. 2. [CrossRef]
23. Lu, Y.; Xu, X.; Wang, L. Smart manufacturing process and system automation—A critical review of the standards and envisioned scenarios. *J. Manuf. Syst.* **2020**, *56*, 312–325. [CrossRef]
24. Bruce, M. Reduce Cost, Increase Speed with Model-Based Definition. SME Media 2020. 2020. Available online: <https://www.sme.org/technologies/articles/2020/december/reduce-cost-increase-speed-with-model-based-definition/?ite=2522&ito=2433&itc=e7e28286-9020-4706-8ec4-21d7883bf037&itx%5Bidio%5D=1144843> (accessed on 21 July 2022).
25. Heysiattalab, S.; Morse, E.P. From STEP to QIF: Product and Manufacturing Information. In Proceedings of the 31st ASPE Annual Meeting, 2016, Portland, OR, USA, 23–28 October 2016. [CrossRef]
26. Corallo, A.; Del Vecchio, V.; Lezzi, M.; Luperto, A. Model-Based Enterprise Approach in the Product Lifecycle Management: State-of-the-Art and Future Research Directions. *Sustainability* **2022**, *14*, 1370. [CrossRef]
27. Shehab, E.; Schuler, M.; Bamforth, P.; Goher, K. Model-Based Enterprise Framework for Aerospace Manufacturing Engineering. *Adv. Transdiscipl. Eng.* **2019**, *9*, 207–212. [CrossRef]

28. Frechette, S.P. Model Based Enterprise for Manufacturing. In Proceedings of the 44th CIRP International Conference on Manufacturing Systems, Madison WI, USA, 1–3 June 2011.
29. Ding, L.; Dai, S.; Mu, P. Point cloud measurements-uncertainty calculation on spatial-feature based registration. *Sens. Rev.* **2019**, *39*, 129–136. [CrossRef]
30. Senin, N.; Catalucci, S.; Moretti, M.; Leach, R.K. Statistical point cloud model to investigate measurement uncertainty in coordinate metrology. *Precis. Eng.* **2021**, *70*, 44–62. [CrossRef]
31. Yang, Y.; Jin, Y.J.; Price, M.P.; Abdelal, G.; Colm, H.; Maropoulos, P. Investigation of point cloud registration uncertainty for gap measurement of aircraft wing assembly. In Proceedings of the 2021 IEEE 8th International Workshop on Metrology for AeroSpace (MetroAeroSpace), Virtual Conference, Italy, 23–25 June 2021.
32. Zhang, K.; Bi, W.; Fu, X. A New Method of Point-Clouds Accurate Measurement and Reconstruction. *Int. J. Database Theory Appl.* **2014**, *7*, 81–94. [CrossRef]
33. Forbes, A. Uncertainties associated with position, size and shape for point cloud data. *J. Phys. Conf. Ser.* **2018**, *1065*, 142023. [CrossRef]
34. Mohammadikaji, M.; Bergmann, S.; Irgenfried, S.; Beyerer, J.; Dachsbacher, C.; Worn, H. A framework for uncertainty propagation in 3D shape measurement using laser triangulation. In Proceedings of the 2016 IEEE International Instrumentation and Measurement Technology Conference Proceedings, Taipei, China, 23–26 May 2016. [CrossRef]
35. Zhao, Y.; Cheng, Y.; Xu, Q. Uncertainty Modeling and Evaluation of Contour Measurement by Structured Light Scanner. *Meas. Sci. Technol.* **2022**, *33*, 095018. [CrossRef]
36. Mahmud, M.; Joannic, D.; Roy, M.; Isheil, A.; Fontaine, J.-F. 3D part inspection path planning of a laser scanner with control on the uncertainty. *Comput. Aided Des.* **2011**, *43*, 345–355. [CrossRef]
37. Feng, H.-Y.; Liu, Y.; Xi, F. Analysis of digitizing errors of a laser scanning system. *Precis. Eng.* **2001**, *25*, 185–191. [CrossRef]
38. Xi, F.; Liu, Y.; Feng, H.-Y. Error compensation for three-dimensional line laser scanning data. *Int. J. Adv. Manuf. Technol.* **2001**, *18*, 211–216. [CrossRef]
39. Dupuis, J.; Kuhlmann, H. High-precision surface inspection: Uncertainty evaluation within an accuracy range of 15µm with triangulation-based laser line scanners. *J. Appl. Geod.* **2014**, *8*, 109–118. [CrossRef]
40. Mendricky, R. Determination of measurement accuracy of optical 3D scanners. *MM Sci. J.* **2016**, *2016*, 1565–1572. [CrossRef]
41. Mendricky, R. Analysis of measurement accuracy of contactless 3D optical scanners. *MM Sci. J.* **2015**, *2015*, 711–716. [CrossRef]
42. Mendricky, R. Aspects affecting accuracy of optical 3d digitization. *MM Sci. J.* **2018**, *2018*, 2267–2275. [CrossRef]
43. Acko, B.; Klobucar, R. *Metrological Approach for Testing Performance of Optical 3D Measurements Systems*; Springer International Publishing: Berlin/Heidelberg, Germany, 2019. [CrossRef]
44. Guidi, G.; Russo, M.; Magrassi, G.; Bordegoni, M. Performance evaluation of triangulation based range sensors. *Sensors* **2010**, *10*, 7192–7215. [CrossRef]
45. Van Gestel, N.; Cuyper, S.; Bleys, P.; Kruth, J.-P. A performance evaluation test for laser line scanners on CMMs. *Opt. Lasers Eng.* **2009**, *47*, 336–342. [CrossRef]
46. ISO/TS 15530-3:2004; Geometrical Product Specifications (GPS)—Coordinate Measuring Machines (CMM): Technique for Determining the Uncertainty of Measurement—Part 3: Use of Calibrated Workpieces or Standards. ISO: Geneva, Switzerland, 2004. Available online: <https://www.iso.org/standard/38695.html> (accessed on 21 July 2022).
47. ISO/TS 15530-4:2008; Geometrical Product Specifications (GPS)—Coordinate Measuring Machines (CMM): Technique for Determining the Uncertainty of Measurement—Part 4: Evaluating Task-Specific Measurement Uncertainty Using Simulation. ISO: Geneva, Switzerland, 2008. Available online: <https://www.iso.org/standard/38696.html> (accessed on 21 July 2022).
48. Heißelmann, D.; Franke, M.; Rost, K.; Wendt, K.; Kistner, T.; Schwahn, C. Determination of measurement uncertainty by Monte Carlo simulation. In *Advanced Mathematical and Computational Tools in Metrology and Testing XI*; Series on Advances in Mathematics for Applied Sciences; World Scientific Publishing: Singapore, 2018; Volume 89, pp. 192–202. [CrossRef]
49. Trenk, M.; Franke, M.; Schwenke, H.; KG, F. The “Virtual CMM” a software tool for uncertainty evaluation—practical application in an accredited calibration lab. *Proc. ASPE Uncertain. Anal. Meas. Des.* **2004**, *6*, 68–75.
50. Wilhelm, R.; Hocken, R.; Schwenke, H. Task specific uncertainty in coordinate measurement. *CIRP Ann.-Manuf. Technol.* **2001**, *50*, 553–563. [CrossRef]
51. Vlaeyen, M.; Haitjema, H.; Dewulf, W. Digital twin of an optical measurement system. *Sensors* **2021**, *21*, 6638. [CrossRef]
52. Stojadinovic, S.M.; Zivanovic, S.; Slavkovic, N.; Durakbasa, N.M. Digital measurement twin for CMM inspection based on step-NC. *Int. J. Comput. Integr. Manuf.* **2021**, *34*, 1327–1347. [CrossRef]
53. Ríos, J.; Staudter, G.; Weber, M.; Anderl, R. Enabling the digital twin: A review of the modelling of measurement uncertainty on data transfer standards and its relationship with data from tests. *Int. J. Prod. Lifecycle Manag.* **2020**, *12*, 226–249. [CrossRef]
54. Zuo, C.; Qian, J.; Feng, S.; Yin, W.; Li, Y.; Fan, P.; Han, J.; Qian, K.; Chen, Q. Deep learning in optical metrology: A review. *Light Sci. Appl.* **2022**, *11*, 39. [CrossRef]
55. Vallejo, M.; de la Espriella, C.; Gómez-Santamaría, J.A.; Ramírez-Barrera, A.F.; Delgado-Trejos, E. Soft metrology based on machine learning: A review. *Meas. Sci. Technol.* **2019**, *31*, 032001. [CrossRef]
56. Choi, J.E.; Hong, S.J. Machine learning-based virtual metrology on film thickness in amorphous carbon layer deposition process. *Meas. Sens.* **2021**, *16*, 100046. [CrossRef]

57. ISO 10303-242; Industrial Automation Systems and Integration-Product Data Representation and Exchange-Part 242: Application Protocol: Managed Model-Based 3D Engineering. ISO: Geneva, Switzerland, 2014. Available online: <https://www.iso.org/standard/57620.html> (accessed on 21 July 2022).
58. ASME Y14.41-2012; Digital Product Definition Data Practices. 2012. An American National Standard. The American Society of Mechanical Engineers: Fairfield, NJ, USA, 2012. Available online: <http://www.asme.org> (accessed on 21 July 2022).
59. ISO 16792; Technical Product Documentation—Digital Product Definition Data Practices. ISO: Geneva, Switzerland, 2006. Available online: <https://www.iso.org/standard/31065.html> (accessed on 21 July 2022).
60. Camba, J.D.; Contero, M. Assessing the impact of geometric design intent annotations on parametric model alteration activities. *Comput. Ind.* **2015**, *71*, 35–45. [[CrossRef](#)]
61. Nyffenegger, F.; Ríos, J.; Rivest, L.; Bouras, A. (Eds.) *Proceedings of the 17th IFIP WG 5.1 International Conference, PLM 2020, Rapperswil, Switzerland, 5–8 July 2020*; Springer: Berlin/Heidelberg, Germany, 2020; p. 594. [[CrossRef](#)]
62. Liu, R.; Duan, G.-J.; Liu, J. A framework for model-based integrated inspection. *Int. J. Adv. Manuf. Technol.* **2019**, *103*, 3643–3665. [[CrossRef](#)]
63. Albuquerque, V.A.; Liou, F.W.; Mitchell, O.R. Inspection point placement and path planning algorithms for automatic CMM inspection. *Int. J. Comput. Integr. Manuf.* **2000**, *13*, 107–120. [[CrossRef](#)]
64. Ainsworth, I.; Ristic, M.; Brujic, D. CAD-based measurement path planning for free-form shapes using contact probes. *Int. J. Adv. Manuf. Technol.* **2000**, *16*, 23–31. [[CrossRef](#)]
65. Nguyen, J. MBD (Model-Based Definition). 8 February 2021. Available online: <https://www.capvidia.com/blog/mbd-model-based-definition-in-the-21st-century> (accessed on 21 July 2022).
66. Jamshidi, J.; Maropoulos, P.G. Design of an Information System for Metrology Contents. In *Proceedings of the 6th CIRP-Sponsored International Conference on Digital Enterprise Technology, HongKong, China, 14–16 December 2009*; Volume 18, pp. 1701–1719. [[CrossRef](#)]
67. Steck, R. Model-Based Product Definition Links Design and Metrology. *Metrology News*, 14 February 2022. Available online: <https://metrology.news/model-based-product-definition-links-design-and-metrology/> (accessed on 21 July 2022).
68. ISO 1101:2017; Geometrical Product Specifications (GPS)—Geometrical Tolerancing—Tolerances of Form, Orientation, Location and Run-Out. ISO: Geneva, Switzerland, 2017. Available online: <https://www.iso.org/standard/66777.html> (accessed on 21 July 2022).

MDPI
St. Alban-Anlage 66
4052 Basel
Switzerland
Tel. +41 61 683 77 34
Fax +41 61 302 89 18
www.mdpi.com

Metrology Editorial Office
E-mail: metrology@mdpi.com
www.mdpi.com/journal/metrology



MDPI
St. Alban-Anlage 66
4052 Basel
Switzerland

Tel: +41 61 683 77 34

www.mdpi.com



ISBN 978-3-0365-6609-2