European Metrology Network for Mathematics and Statistics

Strategic Research Agenda 2023 – 2033
Version 1.0 (07/2023)

EURAMET

MATHMET

Authorship and Imprint

This document was developed by the EURAMET e.V., European Metrology Network (EMN) for Mathematics and Statistics.

Authors: Sebastian Heidenreich, Markus Bär, Katy Klauenberg, Clemens Elster (PTB, Germany), Peter Harris, Keith Lines, Jean-Laurent Hippolyte, Indhu George, Louise Wright, Maurice Cox, Susan Rhodes (NPL, United Kingdom), Luca Zilberti, Alessandra Manzin, Francesca Pennecchi (INRiM, Italy), Gertjan Kok, Adriaan van der Veen (VSL, Netherlands), Nicolas Fischer (LNE, France), Joao Alves Sousa (IPQ, Portugal), Kristin Vogel (BAM, Germany), Alen Bosnjakovic (IMBiH, Bosnia and Herzegovina), Stephen Ellison (LGC, United Kingdom).

Part of this work has received funding from the EMPIR programme (18NET 05 'Support for a European Metrology Network for mathematics and statistics'). The EMPIR initiative is co-funded by the European Union's Horizon 2020 research and innovation programme and the EMPIR participating states.

EURAMET e.V.
Bundesallee 100
D-38116 Braunschweig
Germany

E-Mail: secretariat@euramet.org
Phone: +49 531 592 1960

Version

1.0 (07/2023)

Official language

The English language version of this document is the definitive version.

Copyright

ISBN 978-3-942992-78-7

Image on cover page: istockphoto.com/hh5800

Images on pages 20 and 29 via Pixabay

Further information

No representation is made, nor warranty given that this document or the information contained in it will be suitable for any particular purpose. In no event shall EURAMET, the authors or anyone else involved in the creation of the document be liable for any damages whatsoever arising out of the use of the information contained herein. The parties using the Strategic Research Agenda shall indemnify EURAMET accordingly.

For further information about this document, please contact your national contact person of the EURAMET European Metrology Network for Mathematics and Statistics (see https://www.euramet.org).

CONTENTS

Abstract

This document constitutes the Strategic Research Agenda (SRA) for the European Metrology Network for Mathematics and Statistics in Metrology (EMN Mathmet). The EMN Mathmet is an alliance of European National Metrology Institutes (NMIs), Designated Institutes (DIs) and an EMN Partner that aims to strengthen research and cooperation in the field. The SRA has been developed within a European project (EMPIR 18NET05 MATHMET) to promote and support the network. The SRA was developed based on a consultation process with stakeholders and the strategies of individual NMIs and DIs, and in alignment with the EURAMET 2030 strategy. As a key result, the SRA defines a long-term research goal: the EMN Mathmet will coordinate research to strengthen the trust in algorithms, software tools and data to underpin digital transformation. For this purpose, new emerging research topics where algorithms, software tools and data play a significant role were identified: (i) Artificial Intelligence and Machine Learning, and (ii) Computational Modelling and Virtual Metrology. The foundation for the development of these new topics is given by the traditional focus on (iii) Data Analysis and Uncertainty Evaluation. The SRA characterises the future needs and challenges in the field of mathematics and statistics in metrology and provides an outline of how the EMN Mathmet can meet these new emerging requirements.

1 INTRODUCTION

"Every industry and every organization will have to transform itself in the next few years. What is coming at us is bigger than the original internet, and you need to understand it, get on board with it, and figure out how to transform your business."

— Tim O'Reilly, Founder & CEO of O'Reilly Media

1.1 EMN Mathmet

EURAMET is the European Association of National Metrology Institutes (NMIs) with the mission to develop and disseminate an integrated, cost effective and internationally competitive measurement infrastructure for Europe that addresses the needs of industry, business and governments. To support this mission, since 2019 eleven European Metrology Networks (EMNs) have been established under the umbrella of EURAMET. They cover various themes including climate and ocean observation, quantum technologies, advanced manufacturing, traceability in laboratory medicine and food safety.

The European Metrology Network for Mathematics and Statistics (EMN Mathmet) was founded in June 2019 and its membership comprises 15 NMIs/Designated Institutes (DIs) and 1 EMN Partner, having signed a Memorandum of Understanding. Measurement science increasingly relies on new analytical and computational approaches, requiring expert knowledge of, and support from, the areas of applied mathematics, statistics, and numerical computation, and state-of-the-art computational tools. The mission of the EMN Mathmet is to create sustainable structures in mathematics and statistics for metrology to strengthen the European measurement infrastructure. To achieve the overall mission, Mathmet is

- creating and disseminating knowledge,
- gaining international leadership and recognition,
- building a coordinated infrastructure, and
- establishing stakeholder relations

in the areas of applied mathematics and statistics for metrology. The creation of a Strategic Research Agenda (SRA), which describes the future needs for research in mathematics and statistics for metrology, will help to inform and guide these activities. The research needs identified in this SRA are strongly driven by the digital transformation of industry, society and in turn metrology.

1.2 Need for a Strategic Research Agenda

Metrology and its implementation in NMIs, DIs, laboratories, academia, and industry increasingly depends on good quality models and computation. For example, challenges in modelling and understanding climate variables, the fiscal metering of renewable energy, and advanced manufacturing all rely on models and measurement results. In many of these areas, mathematics, statistics, and models play an essential role and can make the difference between a successful implementation of infrastructure and a poorly performing one.

This trend is strengthened by the increased use of intelligent systems as a result of the digital transformation. Everything is becoming "smarter" and "data-driven" at all levels of the manufacturing process. A key element of that development is, for example, the use of smart sensors or networks where measured data are directly processed.

This development not only unlocks possible new business fields, but also provides the opportunity to increase the efficiency of processes, reduce waste and enable the personalisation of products in many sectors, ranging from the environment to health and to manufacturing. In particular, the personalisation of healthcare products and medical therapies not only helps to improve quality of life but also contributes to saving limited resources.

The new role of metrology and the determination of uncertainty for complex and smart systems was already emphasised several years ago by M. Sené, I. Gilmore and J. T. Janssen:

“Measurement technology is becoming more powerful and complex … Tracking and quantifying the uncertainty of the final result can get lost amid all this data crunching …… An increasing number of research areas lack a metrological framework, however …. ***Quantifying uncertainty in complex problems is almost becoming a field in itself****. The metrology community needs to step up to this challenge, in particular by engaging more statisticians, data experts.”,* Nature 2017 [1].

Recently, the International Committee for Weights and Measures (CIPM) commissioned a task group to transform the SI into the digitised world. This task group on the “Digital-SI” formulated its *Grand vision of the SI Digital Framework* in which the measurement is described holistically including all accessible information. The following excerpt illustrates the importance and necessity of increasingly focusing on **models**, **data**, **software tools** and **algorithms**. Extract from BIMP-WS-digital SI/2021-Vision [2]:

“*The framework will allow more information to be represented digitally, not only measurement results, but also the system being measured, how the measurements were made, and the workflow (data, models, software) associated with establishing the measurement results …*

In the longer term:

- *Digital representation of measurement procedures, measurement workflows, analysis methods, provenance and traceability chains, that allows machines to access and act upon this information with little or no human intervention.*
- *Especially digital representation of key comparison and interlaboratory comparison.*
- *Embed the SI Digital Framework in cyber-physical systems, e.g. sensor networks, Internet of Things environments, autonomous systems, and establish traceability at point of measurement*.”

This emerging development is a unique challenge for metrology, which is built upon an established system comprising a quality infrastructure, accreditation services, conformity assessment, norms and standards, and market surveillance. Comprehensive changes are required to be able to continue to guarantee confidence in measurements and their digital equivalents, and the safety of systems that depend on measurement. In addition to technically advanced challenges, such as the deployment of digital calibration certificates, the development of a quality infrastructure for data and digitally supported testing and approval processes, there is a significant need for research in the field of mathematics and statistics for metrology.

Mathematicians and statisticians from NMIs and DIs across Europe have joined forces in the EMN Mathmet to identify and coordinate corresponding research needs. This SRA records those research needs, explains the motivation for them in terms of addressing requirements from the perspectives of the metrology community and other stakeholders, and proposes a roadmap for meeting those needs. It will be used to ensure that the EMN Mathmet has a sound basis to direct future research activities in the field of mathematics and statistics for metrology, and to increase the efficiency of the EMN in its task to support research developments in modern metrology at the European level. It will also enable the EMN to influence the structure and topics of future European research programmes to address emerging needs.

1.3 Scope

This SRA covers the research area of “Mathematics and Statistics for Metrology”, which is the focus of the EMN Mathmet. It is aligned with stakeholder needs, the EURAMET 2030 strategy and individual NMI strategies. The SRA aims to pool the resources of Mathmet members, associates, and partners to address new, emerging developments in mathematics and statistics for metrology. These new developments are primarily driven by the digital transformation.

The SRA was developed based on the results of surveys of Mathmet members and associates, stakeholder interviews, and workshops and consultations with the Stakeholder Advisory Committee of the EMN. The stakeholder consultation process identified two new developments in the field of mathematics and statistics for metrology as priorities for the SRA:

1. Artificial Intelligence and Machine Learning (AI&ML)
2. Computational Modelling and Virtual Metrology (CM&VM)

but recognises that the traditional research field of Data Analysis and Uncertainty Evaluation[1] (DA&UE) continues to be important, crosscutting all scientific applications and yet posing unresolved challenges.

The research topics constitute very large areas with their own research and user communities that go beyond those involved in, or dependent on, measurement. Therefore, this SRA focusses on the research required for AI&ML and CM&VM to be harnessed by the metrology community in ways that are principled and trustworthy, as well as to increase the reach of metrology into new applications. It points at the research needed to support the metrology community, principally NMIs and DIs such as members of the EMN Mathmet, other Euramet EMNs and TCs, but also more widely those industries and sectors that depend on measurement, such as advanced manufacturing, healthcare, environment, and energy. The intention is to look forward 10 years, but it is expected to review and update the material periodically.

This document is a recommendation for research activities within the EMN Mathmet for the next decade to underpin the vision that the

EMN Mathmet ensures quality and trust in algorithms, software tools and data for metrology, and in inferences made from such data, to foster the digital transformation.

The document is also addressed to stakeholders. It informs them about plans for future research and invites them to engage with the EMN Mathmet in activities which form the basis for long-term relationships. The SRA is a key document for the mathematics and statistics in metrology community and will become an integral part of the strategic agenda of the EMN Mathmet.

1.4 Identifying Stakeholder needs

The development of the Strategic Research Agenda for the EMN Mathmet was driven by an extensive and inclusive stakeholder consultation process. The goal of this process was to identify urgent stakeholder needs, new challenges, and opportunities in the areas of mathematics and statistics in metrology. Using a variety of consultation methods, EMN Mathmet ensured that the resulting SRA reflected a comprehensive and well-rounded understanding of stakeholder perspectives.

The first phase of the consultation process involved the mapping process of more than 60 stakeholders from a variety of sectors, including academia, industry, government, and research institutes. This mapping was critical to understanding the interests, concerns, and needs of the various stakeholder groups.

From the identified stakeholders, EMN Mathmet formed the Stakeholder Advisory Committee (SAC), which included nine key stakeholders with significant influence and diverse backgrounds. The SAC was essential in providing strategic guidance, insights, and recommendations for the SRA development process.

---

[1] In this document, we distinguish between uncertainties in the context of data analysis for physical measurements ('uncertainty evaluation') and generally in the context of emerging fields ('uncertainty quantification').

The SAC held regular discussion meetings to identify and assess new challenges and opportunities in mathematics and statistics in metrology. These meetings provided a platform for expert dialogue and fostered collaboration and knowledge sharing among stakeholders.

In addition to the SAC meetings, EMN Mathmet conducted interviews with key stakeholders to gain qualitative insights into their needs, expectations, and concerns. These interviews provided an in-depth understanding of stakeholder perspectives that informed the SRA's focus areas.

EMN Mathmet used questionnaires to collect quantitative data from stakeholders. These questionnaires were distributed during the joint ENBIS and EMN Mathmet workshop: Mathematical and Statistical Methods for Metrology (MSMM 2021). Survey responses provided valuable feedback on stakeholder priorities and needs.

Panel discussions at the MSMM 2021 and MATHMET 2022 conferences facilitated open dialogue among stakeholders and provided further insight into the emerging challenges of mathematics and statistics in metrology. These discussions also provided an opportunity for stakeholders to express their views on the SRA development process.

EMN Mathmet engaged with EURAMET, EURAMET EMNs, and EURAMET Technical Committees (TCs) to gather their feedback and recommendations. This collaboration ensured that the SRA was aligned with the broader metrology and research goals at the European level.

The needs of stakeholders involved in European joint research projects with strong EMN Mathmet participation were included to explore current challenges and opportunities in metrology. These discussions provided real-world insights and ensured that the SRA remained relevant and responsive to ongoing developments in the field.

Since October 2021, EMN Mathmet has started an initiative on measurement uncertainty training, for which a consortium of Mathmet and non-Mathmet members was formed with the aim of improving the quality, efficiency, and dissemination of measurement uncertainty training. A large variety of stakeholders are supporting this activity: 35 members encompassing NMIs, EURAMET associates, EMNs and TCs, metrological and accreditation bodies, companies and universities. Non-Mathmet members and stakeholders of the Mathmet “MU Training” activity are automatically members of the EMN stakeholder community and can be considered for inclusion in the SAC, if desired. They were informed about the first draft of the SRA and asked for feedback.

Through this extensive stakeholder consultation process, EMN Mathmet was able to define the challenges and priorities within the SRA, ultimately resulting in a strategic roadmap that addressed the most pressing stakeholder needs. By putting stakeholder needs and concerns at the top of the agenda, the SRA is able to drive significant advances and innovations in the field of mathematics and statistics in metrology. The consultation process will be carried out continuously, to update the needs of the stakeholder community and to identify new challenges: this will help to keep updating the SRA.

1.5 Outline of this SRA

Mathematics and statistics underpin the key competences in metrology in terms of (i) measurement traceability, (ii) conformity assessment, and (iii) key comparisons, through modelling and data analysis. Critical to these competencies is the evaluation and reporting of reliable measurement results, including measurement uncertainties, that can only be determined with suitable methods from mathematics and statistics.

Within the context of digitalisation, new approaches such as Artificial Intelligence and Virtual Metrology are increasingly used in industry for metrological applications. For these new approaches, there is still a lack of foundational methods in mathematics and statistics to ensure the metrological key competences for the next generation of metrology. This lack of methods implies that currently no metrological quality infrastructure can be built for such prospective applications, which places the EU at a competitive disadvantage. To explore what is needed and to address this grand challenge, the EMN Mathmet has focused its SRA on the topics of AI&ML and CM&VM.

In Section 2, the stakeholder consultation process that resulted in this SRA is described, roadmaps for the research topics considered in this SRA are collated, and recommendations and next steps for activities to be undertaken by the EMN Mathmet based on them are given. The material presented in this section is a distillation of the more detailed information provided in subsequent sections.

In Section 3, *Artificial Intelligence (AI)* and *Machine Learning (ML)* is introduced from the metrological perspective. After an introduction giving the background to the topic, the needs of stakeholders and the metrological infrastructure are summarised. Next, the key challenges and open research questions arising from the gap between the needs and the state of the art are discussed. The needs and challenges are supported by modern applications and use cases.

Section 4 deals with the research topic *Computational Modelling (CM)* and *Virtual Metrology (VM)*. After a brief introduction, the current state of the art from the metrological perspective and the needs of CM&VM based on NMI strategies and stakeholder consultations are presented. Next, based on the needs and the state of the art in CM&VM, the resulting challenges are described.

Section 5 is dedicated to the more classical topics of *Data Analysis (DA)* and *Uncertainty Evaluation (UE)*. Both topics are at the core of past and current research activities of the EMN Mathmet members and associates and will continue to be so. The expertise in these areas also enables the EMN Mathmet to meet the challenges of emerging research topics, such as AI&ML and CM&VM. Nevertheless, there are still open research questions not only of interest for the metrology community, but also for stakeholders from industry. This shorter section presents selected needs and challenges to be addressed in this area.

Finally, the conclusion (Section 6) discusses and summarises the key challenges for the identified research topics and gives an outline of how the overall vision that the "*EMN Mathmet ensures quality and trust in algorithms, software tools and data for metrology, and in inferences made from such data, to foster the digital transformation"* can be best achieved.

2 ROADMAPS AND RECOMMENDATIONS

2.1 Background to the roadmaps

For the two emerging research topics in the field of mathematics and statistics in metrology, the stakeholder consultation process was used to define key sub-topics for which research needs and challenges can be clearly identified.

For the first topic (i) on Artificial Intelligence and Machine Learning (AI & ML) these sub-topics are:

1. Uncertainty quantification
2. Generalisability and robustness
3. Interpretability
4. Quality framework.

For the second topic (ii) on Computational Modelling and Virtual metrology (CM & VM) they are:

1. Reliability and quality
2. Efficiency and real-time calculations
3. Uncertainty quantification.

To meet the research needs and challenges identified for these sub-topics within the next decade, the necessary steps are defined in the roadmaps presented in the following subsections. The steps are divided into those needed in the short-term (2023–2027) and those needed in the medium-term (2028–2033). A detailed description and justification for the roadmaps are given in the following sections.

2.2 Roadmap – Artificial Intelligence and Machine Learning

		Time	
Topic	**Sub-topic**	**Short-term (2023 – 2027)**	**Medium-term (2028 – 2033)**
Artificial Intelligence and Machine Learning	Uncertainty Quantification (UQ)	• Information about general metrology requirements for UQ for ML models and based on requirements from specific application areas (advanced manufacturing, energy and environment, healthcare, food safety, etc.), information about the different sources of uncertainty and methods for quantifying the uncertainty for those sources. • Methods for UQ for ML models focussed on supervised learning (regression and classification tasks). • Methods for propagating uncertainty of test data through fixed (already trained) ML models. • Methods for propagating uncertainty of training data to quantify uncertainty of ML model parameters. • Methods for combining the effect of uncertainty from different sources: the test data, the training data, model misspecification, the training of ML models, etc. • Methods to address metrology requirements for UQ: non-Gaussian distributions, correlated and heterogeneous effects, large-scale, etc.	• Methods for UQ for ML models extended to transfer learning. • Methods for UQ for ML models extended to semi-supervised and unsupervised learning (e.g., clustering). • Methods for UQ for (data-driven) ML models augmented with domain knowledge (e.g., physics-inspired neural networks [PINNs]). • Information about metrology requirements for UQ for ML models in the context of reinforcement learning.
	Generalisability and Robustness	• Methods and metrics to compare the predictive performance of ML models on training data and test data that account for uncertainty in such data. • Methods to assess the extent to which ML predictions are resilient to random, systematic, adversarial, and out-of-distribution perturbations in input data.	• Methods to evaluate the global robustness of ML predictions to variations in the training data and in the hyper-parameters of the training algorithm. • Methods to improve the robustness of ML predictions, including those based on adversarial training, on modelling label error, and on

		• Methods to assess the sensitivity of ML predictions to new test data and its uncertainty. • Benchmarking datasets and design metrics to allow for quantitative evaluation of generalisability and robustness and comparisons of (current and new) methods.	automatically searching for deep neural network architectures that are inherently robust to noise and incorrectly labelled data. • Methods to assess and improve generalisability and robustness of ML models extended to transfer learning.
	Interpretability	• Methods designed for classification tasks adapted to apply to regression tasks, and those designed for images adapted for the types of input data of relevance in metrology applications (e.g., explanation methods that are gradient-, propagation- or perturbation-based or employ surrogate models to approximate the ML model). • Benchmarking datasets and design metrics to allow for quantitative evaluation of explanations and comparisons of (current and new) explanation methods.	• Methods for the design and training of inherently interpretable networks, such as regularisation, hybrid methods and architectural adjustments. • Methods based on incorporating domain knowledge to support interpretability (e.g., physics-inspired neural networks [PINNs]).
	Quality Framework	• Guidance on the specification and collection of training, validation and testing datasets to support trustworthy ML in metrology applications (e.g., addressing questions of quality, balance, bias, combining training datasets, augmenting training data, pre-processing, and cleaning). • Guidance on model choice (e.g., choice of neural network architecture, choice of kernel function for Gaussian Processes, etc.). • Guidance on model training (e.g., setting and optimising training parameters, adversarial training, etc.). • Framework to support verification and validation of ML models, including design of appropriate metrics.	• Methods for formal verification of ML models. • Framework to support reproducibility of results from data processing pipelines dependent on ML models (e.g., recognising probabilistic nature of process of model training and role of training, validation, and test data in that process). • Framework to support auditability of ML models in metrology applications. • Specification of a standard interface to support frameworks for benchmarking, validation, and certification of ML models.

2.3 Roadmap – Computational Modelling and Virtual Metrology

		Time	
Topic	**Sub-topic**	**Short-term (2023 – 2027)**	**Medium-term (2028 – 2033)**
Computational Modelling and Virtual Metrology	Reliability and Quality	• Validation framework for virtual experiments and digital twins. • Approaches to verify computational models. • Statistical procedures for the assessment of the discrepancy between standard measurements and the data from the virtual counterpart.	• Framework for key comparisons of virtual measurements devices. • Virtual test and reference standards, e.g., standardised virtual calibrations.
	Efficiency and Real-Time Calculations	• Real-time methods for computationally expensive systems. • Model error determinations, e.g., surrogate models.	• Methods to treat efficiently high dimensional parametric problems.
	Uncertainty Quantification (UQ)	• Methods for UQ for virtual measurements. • Methods for UQ for digital twins.	• Traceability chain to a virtual/real standard. • Digital twins for metrological applications.

2.4 Recommendations

The Strategic Research Agenda (SRA) is a key tool for achieving Mathmet's strategic goals. To achieve the goals effectively, several key steps are required. First, it is necessary to identify potential funding opportunities to ensure that research projects have the necessary resources. Second, a quality assurance framework (quality assurance tools) must be developed to ensure the scientific excellence and relevance of research results such as software, data and guidelines. Third, specific use cases should be identified to develop practical solutions to real-world metrological challenges and add value to the stakeholder communities. By combining these steps, Mathmet can successfully achieve its research goals while accelerating progress and innovation depicted on the roadmap. In the following, and in Appendix D, we illuminate these three important steps.

2.4.1 Quality Assurance tools

To achieve the goals of the roadmap, it is recommended to employ specified quality standards. The Quality Assurance Tools (QAT) were developed and designed by the JNP consortium to ensure that research outputs in the forms of data, software and guidelines are fit-for-purpose, achieve a sufficient level of quality, and are consistent with the aims of National Measurement Institutes to provide quality-assured and trusted outputs.

For research outputs in the form of data and software, the QAT is structured according to iterative life cycles for developing software and builds upon existing good practices for data management and software development, following the process-based approach of ISO 9001 and ISO 8000.

Components of the QAT

The QAT consists of separate components for data, software, and guidelines. For data and software an on-line interactive risk assessment tool guides the user in developing a quality management plan. For guidelines (future and existing), the QAT involves completing an interactive *checklist* comprising a set of questions, and making a recommendation based on the answers to those questions.

Figure 1: Sketch of the QAT. The QAT consists of 4 components: software, guidelines and data are assessed with a risk tool to ensure tailored quality assurance

The components are summarised below:

- A quality management plan is key to the QAT components for data and software. This plan lists the quality management activities needed for a particular dataset or piece of software. These activities follow a typical life cycle from requirements capture to design and development, verification, and validation through to release and maintenance. Review is an important activity that is carried out throughout the life cycle.
- Quality management plans are generated using an on-line, interactive risk assessment tool. Risk is quantified using a value called an integrity level. The integrity level is a number between 1 and 4, where 1 indicates the lowest level of risk (e.g., prototypes of software for internal use within an organisation) and 4 indicates the highest level (e.g., software that is safety critical). The integrity level is used to decide quality management activities, i.e., the activities listed on the plan, to be undertaken. Mathmet provides an on-line risk assessment tool to guide the user through the process of calculating an integrity level and generating a quality management plan.
- The concept of an integrity level is analogous to, but should not be confused with, the “safety integrity level” of IEC 61508 “Functional safety of electrical/electronic/ programmable electronic safety related systems”. This value is also a number between 1 and 4, where 1 specifies the lowest level of functional safety and 4 the highest.
- For software, the QAT can include established quality procedures and templates from the Mathmet members. However, the activities listed in the quality management plan must be carried out.
- Quality management of data is less well established than quality management of software. Accordingly, developing this component of the QAT was arguably a research activity to some extent. A key concept is that integrity levels are adapted to apply to data as well as software.
- As with software, an on-line risk assessment tool will assist with assigning integrity levels and generating quality management plans.
- The aim of the Guidelines component of the QAT is to ensure a sufficient level of quality in the development, assessment, and recommendation of existing and future guidelines for mathematics and statistics in metrology.
- The QAT is available on the Mathmet website and can be downloaded here.

The FAIR (Findable, Accessible, Interoperable, Reusable) principles are of particular interest for metrology because they provide a robust framework for managing, processing, and sharing measurement data. In terms of Artificial Intelligence and Virtual Metrology, compliance with these principles can ensure the quality and integrity of the data used, ultimately leading to more accurate and reliable predictions. When data conforms to FAIR principles, it is easier to identify and eliminate biases, errors, and inconsistencies, leading to a more trustworthy foundation for model development. This is especially important when data come from multiple sources that have different quality characteristics or are interdependent to varying degrees.

The quality assurance tools developed by the EMN Mathmet facilitate the implementation of FAIR principles for metrological applications by helping to support the integrity, origin, and traceability of data throughout the process of data acquisition, labelling, modeling, validation, and verification. A well-documented quality management system provides transparency and allows internal and external auditors to evaluate the processes and methods involved. This verifiability in turn contributes to user confidence. The development and adoption of generally applicable standards and guidance documents that incorporate best practices for data management can ensure that their models correspond to FAIR principles, ultimately leading to more accurate, reliable, and trustworthy forecasts in the metrology domain.

2.4.2 Applications and use cases

To achieve the goals depicted in the road maps it is recommended to define application and use cases tailored to the urgent needs of different stakeholder communities. By addressing the most important metrological issues of these stakeholders in the field of mathematics and statistics, we aim to pave the way for successful collaboration and to make a significant contribution to metrology in various fields.

By exploring these use cases, all stakeholders can gain valuable insights into potential solutions to the emerging challenges in mathematics and statistics in metrology, but also further develop specific applications.

Moreover, the identified use cases directly relate to the challenges defined in the SRA, ensuring that the initiatives taken are in line with the research and innovation objectives of the EMN Mathmet.

Electrical Properties Tomography (EPT) is a family of techniques that use the spatial distribution of the radiofrequency magnetic field acquired during a session of Magnetic Resonance Imaging (MRI) to obtain an indirect measurement of the dielectric properties (conductivity and/or permittivity) of biological tissues. The measurements obtained through standard EPT suffer from systematic errors at the boundary between different tissues, and first attempts to correct these errors using deep learning have appeared in the literature [14]. Quantifying uncertainty and assessing robustness are important considerations for these methods.

Magnetic Resonance Imaging (MRI) is a highly versatile and powerful medical imaging technology. The main challenges of MRI are high costs for purchase, infrastructure, and maintenance [3]. Currently MRI is only available at highly specialised hospitals and high-income countries. Portable and affordable open-source low-field MRI has the potential to overcome these challenges and revolutionise medical health care, which is fully in line with the current EU4Health Programme [4-6]. Cheaper production leads to imperfections in the MR system. To achieve high image quality and diagnostic accuracy these imperfections can be characterised metrologically and then taken into consideration during image reconstruction. This leads to a large scale non-linear inverse problem which can only be efficiently solved using **physics-informed neural networks** (PINNs). For a reliable diagnosis and hence optimal patient care it is essential to characterise the uncertainty and robustness of these complex image reconstruction algorithms.

Digital phantoms and in silico approaches are essential instruments for the validation and calibration of medical imaging techniques as well as for the development of novel AI-based diagnostic tools, as also stated by the Food and Drug Administration (FDA) in a recent report “Spotlight: Digital Health Regulatory Science Opportunities”. **Computer-generated virtual models**, with known anatomy and physiological functions, combined with in silico modelling, can be used as a gold standard to produce reference data for improving diagnostic devices (e.g., for **radiography**, **MRI**), as well as for optimising image processing and reconstruction techniques. Given the model of the physics of the imaging process, acquired data of the digital phantoms can be generated numerically, providing a tool for device calibration and for the generation of **synthetic data**, to be used as a ground truth for signal generation and image production process, and for the benchmarking of novel AI-based diagnostics tools.

The **Photoplethysmograph (PPG)** is a non-invasive device that measures the intensity of transmitted or refracted light through body tissue. PPG signals vary throughout the cardiac cycle and can be measured at various parts of the body, such as the finger, wrist, arm, ankle, heart, and neck. The signals provide valuable information about the cardiovascular, respiratory, and autonomic nervous systems. PPG devices are widely available and inexpensive, making them popular for clinical and wearable use. Additionally, contactless measurements by external cameras or smart devices allow for long-term monitoring without patient discomfort. Despite the vast amount of data collected through PPG signals, algorithmic evaluation of the signals to detect diseases or infer physiological parameters is almost never used in clinical settings. One reason for this is the lack of trust in the output of such algorithms. Machine learning methodologies are essential for the extraction and evaluation of key features used for diagnosis. However, confidence in the performance and predictions of machine learning algorithms is crucial in medical contexts, where diagnostic mistakes can be fatal or result in unnecessary anxiety and overtreatment. An analysis of the **uncertainty associated with machine learning algorithms** and their predictions is therefore indispensable for providing clinicians and users of wearable devices with critical information about the quality and trustworthiness of the produced results.

The increased reliance on renewables leads to challenges in ensuring the **stability of power grids**, and the European Network of Transmission System Operators for Electricity (ENTSO-E) has recognised a new need to monitor inertia as a critical system stability parameter [7]. Direct inertia measurement is challenging, but new approaches overcome these challenges by applying ML to grid measurement data to determine inertia. However, the acceptance of these solutions by grid operators through appropriate standardisation will require uncertainty evaluation. Uncertainty-aware inertia measurement in power grids will lay the foundation for standardisation of new grid monitoring systems and acceptance of the ML approach by risk-averse grid operators, ensuring stable and more economic operation of power grids dominated by renewable energy generation. In the UK alone, the cost of maintaining security margins due to unknown inertia has increased from less than €50 million in 2017 to €350 million in 2021 [8]. A further impact will be lower CO_2 emissions resulting from increased renewables uptake, essential to the transition to net-zero by 2050.

Inferring the **state-of-health of lithium batteries** from impedance measurements is of interest, for example, in the car industry, but the task is beyond the reach of traditional methods. While Gaussian Process regression for this task has been proposed recently [9], the exploration of more scalable ML approaches based on deep learning is still lacking. Uncertainty evaluation associated with these ML approaches is needed to take decisions reliably in conformity assessment. Forecasting the state-of-health of Li-ion batteries is needed in applications such as electric vehicles and other consumer electronics [9]. Uncertainty quantification of such forecasts helps to increase the reliability of a decision to exchange batteries, and it also fosters “the economical and sustainable re-use of the multitude of Li-ion batteries, contributing towards Europe’s renewable energy and climate change goals” [10].

Thermocouples are widely used in the aerospace industry, but they drift and need frequent recalibration in harsh conditions, motivating the recent development of self-calibrating thermocouples by the metrology community [11]. In the absence of complete physics-informed models for the way such thermocouples drift, a data-driven approach can be exploited to make decisions about when the drift warrants a recalibration of the sensor, and it is vital for the accreditation of any such calibration technique that it evaluates uncertainty. The ability to automate **self-calibrating thermocouples** and quantify the associated uncertainties will open the door to their certification within the regulatory frameworks to which end users must comply. This ability will reduce the need for expensive recalibration and replacement of thermocouples in high value manufacturing and other process control applications where process temperature uncertainty has a significant impact on product quality or process efficiency. The result will be more efficient processing, both in terms of energy use and in terms of product quality and yield, reduced wastage, and lower CO_2 emissions.

Microscopy-based techniques such as SEM (Scanning Electron Microscopy) are reference techniques for **measuring nanoparticle size**, which is important in the production and control of materials including titanium dioxide. Image analysis involving segmentation to determine the nanoparticle edges is a key step of the measuring process. The state-of-the-art methods available on the market are so-called `watershed algorithms', but this technique induces errors that negatively affect the reliability of the measurements, which has prompted the consideration of deep learning for segmentation. It is crucial that the methodology involving ML is quality assured, which requires methods for ensuring robustness and evaluating uncertainty [12, 13]. The development of nanoparticle characterisation algorithms with associated uncertainty statements will provide the nanomaterials industry with robust and reliable data, enabling the improvement in quality systems in response to regulatory requirements. For example, REACH (Registration, Evaluation, Authorisation and Restriction of Chemicals) is an EU regulation that addresses the production and use of chemical substances and their potential impacts on both human health and the environment. Moreover, characterisation of nanoparticle populations by electron microscopy is time-consuming and expensive: automating the image analysis process will significantly reduce costs.

Scatterometry is an optical measurement technique for determining the dimensions and parameters of thin film systems and of more complex periodic geometries with dimensions from a few nm to tens of $\mu\mathrm{m}$. To obtain the measurand an inverse problem has to be solved. The classical approach requires repeatedly solving the Maxwell equations for an idealised system, and when in addition Bayesian inference techniques are used to derive uncertainty information, computations may take several weeks or longer. In this context, neural networks can potentially help to solve the inversion much more quickly. Obtaining uncertainty information and interpretation of the output of the neural network is highly desirable to make this approach trustworthy.

Gas turbines represent a key component of both aeronautical engines and energy generation systems. The flow field inside these components is characterised by complex phenomena, and one promising strategy for prediction involves field inversion with respect to measurable physical variables followed by ML regression to find the correlation between the correction field and the physical variables [15]. The analysis of its robustness and sensitivity to the inputs represents a critical aspect for the application of this methodology on industrial cases, and interpretation of the models will also help for providing confidence in generalisability.

Mass spectrometry is a powerful analytical chemical technique for determining a sample's chemical composition by measuring the mass to charge ratio of vapourised and ionised molecules from the sample. Analytical chemical methods produce large volumes of high dimensional and complex data that are difficult to process, data mine or interpret. ML can be used to address the impact of these complexities in segmentation or classification tasks. However, it is often required that the specific features which differentiate the data (such as the location and relative intensity of peaks) are understood from a chemistry perspective and consequently interpreted in the context of the application area [16]. Synthetic datasets with known experimentally derived perturbations will be provided, along with ground truth labels.

In-situ calibration of (low-cost) sensor networks can be an effective alternative to traditional offsite calibration in laboratories, especially when many sensors are deployed in a dense network. Various in situ calibration strategies have been developed, depending on the architecture of the network, e.g., when reference sensors are deployed alongside the low-cost sensors, exploiting the mobility of instruments, by grouping sensors appropriately, and according to the nature of the involved quantities [17]. ML and DL techniques for in situ calibration play a key role in several fields of application, including environmental sensing [18, 19], but require, as every calibration procedure, a careful uncertainty quantification for the calibrated sensor output.

Other applications include optimisation of material design for concrete structures, the pressure traceability chain, medical diagnosis based on ECG and MRI data, applications in earth observation including landcover classification and altimetry, and other applications of nanoparticle characterisation including morphology of soot particles and clinical differential diagnostics. The development of principled uses of ML in these and other applications will enable metrologists to play a strategic role in supporting new technologies, informing standards, and providing validation services.

3 STRATEGIC TOPIC – ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING

The influence of AI will continue to develop in the future and impact all areas of science and society. Ursula von der Leyen, President of the European Commission, said:

"We all know that Artificial Intelligence can do amazing things. And I think we do not talk enough about what Artificial Intelligence is able to do to improve our daily lives. For example, if we look at the health care sector, we know that we use already now Artificial Intelligence for, for example, better diagnoses and earlier diagnoses. And better and earlier diagnoses are crucial when you treat for example cancer – or we use robots for precision surgery.

But Artificial Intelligence is also key for us when we want to reach our goal to be climate neutral in 2050 ..."

European Commission Speech - 20/294

3.1 Introduction

Artificial Intelligence (AI) describes the capability of a machine to acquire, process, create and apply knowledge, held in the form of a model, to conduct one or more given tasks. Here, a task comprises a set of activities (physical, perceptual and/or cognitive) to achieve a specific goal. AI covers a wide range of technologies that reflect different approaches to dealing with these general and complex problems. **Machine Learning** (ML) is a branch of AI that employs computational techniques to enable systems to learn specifically from data. There is a variety of ML algorithms, the three primary types being supervised learning, unsupervised learning, and reinforcement learning.

The goal of **supervised learning** is to learn a function or mapping that approximates the relationship between input and output variables using training data that comprises values for the input variables and corresponding values for the output variables. Supervised learning is applied to both **regression** tasks and **classification** tasks. A regression task uses a ML model to predict a continuous output variable for a given input, and a classification task is one that uses a ML model to predict a discrete output class or label for a given input. The goal of **unsupervised learning** is to learn the natural structure or patterns present within a dataset, for example, to separate the elements of the dataset into different groups. The goal of **reinforcement learning** (RL) is to learn an optimal policy through interactions with an environment that maximises a reward function or other user-provided reinforcement signal. Typical applications of RL are found in self-driving cars and in making recommendations of healthcare treatments. Within these types, there are many different ML models and ML algorithms.

To apply a ML algorithm often involves data pre-processing that can include preparing raw data in a format that can be accepted by the ML model, transforming the data to a different representation (for example, transforming a signal in the form of a time series to an image), and extracting features that are used as the input variables and/or to reduce the dimensionality of the data. **Deep Learning** (DL) is a subset of ML that makes use of neural network models to eliminate the need for some of these data pre-processing steps, particularly regarding the need to extract features, which might be done empirically or using domain knowledge. DL algorithms can ingest, process and analyse large quantities of unstructured data to learn with less human intervention, and to provide a model that is generally used to support a decision-making process undertaken by humans. DL models typically involve many free parameters and rely on the availability of large quantities of data to train them. **Transfer learning** involves using elements of a pre-trained model in a new model. If the two models are developed to perform similar tasks, then knowledge can be shared between them and training of the new model can be done using a smaller quantity of training data than would otherwise be needed.

A **prediction** is the output of a trained ML model when provided with new (unseen) input data. **Generalisability** is the ability of a trained ML model to return a reliable prediction for new input data. A **robust** ML model is one that performs well in the presence of perturbations of new data. **Adversarial training** of a (deep neural network) model is a brute force supervised learning method where the training of the model uses data created to deceive the model, e.g., to return a wrong classification, which is explicitly labeled as 'threatening' (or 'adversarial'). It can be used to improve the robustness of the model against 'attacks' involving such adversarial input data. **Interpretability** is the ease with which a human can comprehend the reasons for a model prediction.

In 2019, the EU's High-Level Expert group on AI published Guidelines [1] that describe **trustworthy AI** as AI that is 'lawful' (respecting all applicable laws and regulations), 'ethical' (respecting ethical principles and values), and 'robust' (both from a technical perspective and considering its social environment). The Guidelines put forward a set of seven key requirements that AI systems should meet to be deemed trustworthy, which include:

- Technical robustness and safety (including general safety, accuracy, reliability, and reproducibility),
- Transparency (including traceability, explainability and communication).

Key goals of metrology are to make measurements accurate, precise, comparable and reproducible. Such goals are achieved by ensuring measurement results are traceable to reference standards (such as primary measurement standards) through a chain of calibrations with the quality of measured values at each link in the chain expressed quantitatively using measurement uncertainty. By thinking about how these established metrology concepts of measurement traceability and measurement uncertainty might be applied to a ML prediction, considered as a measured value and developing tools to realise these concepts in the context of ML, contributes to establishing trust in the prediction and the ML model that generated it.

The primary focus of this SRA is to address the question of how ML can be harnessed in a principled, explainable, and transparent way to derive trusted information about physical, chemical, biological and environmental systems from measured data. Doing so involves helping the metrology community to make good use of ML without compromising established and accepted metrology principles. A secondary focus is to consider how the main concepts of metrology (such as measurement traceability, measurement uncertainty and calibration) can be used to inform the development of standards, regulation and policy to bring trust more generally to systems that use ML.

Within the context of these questions, this research topic is concerned with ML and DL, as opposed to the much broader topic of AI, and focusses on technical aspects of the use of ML and DL as opposed to legal, ethical and related aspects. Here, technical aspects cover mathematical and statistical issues that contribute to the trustworthiness of a prediction, including uncertainty quantification, generalisability and robustness, and interpretability, that are strongly aligned with the remit of the EMN Mathmet. However, consideration is also given to procedural issues, such as guiding the choice of ML model, the impact of the quality and provenance of training data, and the verification and validation of the ML algorithm and software used, that are also part of establishing such trustworthiness. Although currently such procedural issues are less well aligned with the remit of the EMN Mathmet, they are nevertheless important, and are issues that will be added to the remit of the EMN Mathmet in the future. These different aspects are explored in the following sections. There are computational considerations that have important practical implications for the use of ML, and particularly DL, and will be considered in future revisions of this SRA. Such considerations include the computing resources needed to train a highly parametrised DL model on a large dataset and the use of specialised hardware and cloud computing and data storage for that purpose.

The extent of the interaction between ML and metrology is expected to increase over time, as for many scientific endeavours, and so the scope of the research topic has been intentionally kept quite broad. In particular, the SRA is not limited to DL but considers ML in the broader sense. Indeed, it covers both supervised and unsupervised learning, and both regression and classification tasks as examples of supervised learning. It is proposed to keep a ‘watching-brief’ on the topic of RL with the possibility to include it within the scope of the SRA in the future as the role of RL within metrology becomes clearer.

Sources of information used for this research topic of the SRA are taken from:

- A survey of members of the EMN Mathmet
- Consultation with the members of the Stakeholder Advisory Committee of the EMN Mathmet, involving a questionnaire, interviews and a meeting with all members of the Stakeholder Advisory Committee

- Material provided by the members of the EMN Mathmet and by other European Metrology Networks, such as their own research agendas, roadmaps, fore-sighting documents, etc.
- The work and thinking undertaken by members of the EMN Mathmet to prepare Proposed Research Topics (PRTs), Joint Research Projects (JRPs), etc. as part of past and present EMRP, EMPIR and EMP research programmes
- Examples where ML models have been implemented at NMIs in specific metrology areas.

Such information is summarised in this document including in its appendices (Sections 9.1 – 9.3). Information has also been collected indirectly by individuals involved in the development of this SRA from study of the scientific literature, attendance at workshops, attendance at meetings of the technical committees of international and national standardisation bodies (ISO, IEC, CEN, CENELEC, etc.), and contact with industry.

The section is organised as follows. Sections 3.2 and 3.3 describe, respectively, the needs and challenges beyond the state of the art that the research seeks to address. These needs and challenges are presented at a high-level and expressed in generic terms. Use cases and applications are listed in Section 3.3 to provide instances of where the research can be expected to have impact by meeting specific needs and challenges. The list is not exhaustive and is expected to change with time as new applications of ML arise. Moreover, the list is not intended to imply an ordering or prioritisation of use cases and applications to be treated in research projects.

3.2 Needs

There have been huge advances in recent years in the capability of ML, and especially DL, to build accurate data-driven predictive models, thanks to the availability of increasing volumes of data and advances in computational processing power. ML is a key driver of digital transformation, which has the potential to revolutionise the way that we understand the world and make decisions across all sectors of society, including manufacturing, healthcare and the life sciences, energy and the environment, transportation, and smart cities. The use of ML in metrology offers the potential to extend the range of applications to include those in which the underlying physical model is not well understood or too complex to be computationally feasible. To support a ML-driven digital transformation, the metrology community must adopt ML and incorporate it into its frameworks. Doing so requires the principles and practices of the metrological community to be applied to ML systems to characterise the performance of such systems and the uncertainty of any results or predictions generated by them.

However, the widespread adoption of ML is hindered by the perceived untrustworthiness, in some quarters, of its outputs. The High-Level Expert Group on AI set up by the European Commission identifies trustworthy AI as their foundational ambition [1]. Furthermore, it is crucial, in a metrological context, that predictions based upon ML are quality assured and traceable to reference standards. Without such assurances, the decisions made based on ML models may be unreliable and the risks associated with those decisions cannot be duly assessed, which limits their value in the context of, for example, comparison of results, conformity assessment, safety and regulation.

The challenge of making ML trustworthy aligns closely with the vision of the European Partnership on Artificial Intelligence, Data and Robotics, which focusses on the development and deployment of trustworthy AI. The EU aims to invest €1.3 billion in this partnership, to be matched by €2.6 billion from industry by 2030. McKinsey has estimated that AI may deliver an additional output to the European economy of around €2.7 trillion, or 19 %, by 2030 [2], and trustworthy ML is crucial to realising this economic impact. For example, smart factories incorporating automated decision-making have the potential to bring a 90 % improvement in resource efficiency and a 30 % increase in productivity in advanced manufacturing [3].

Increased automation within manufacturing will also reduce energy consumption and consequently carbon emissions.

There is currently no good practice guide or standard that directly addresses uncertainty evaluation for ML. The AI sub-committee of ISO/IEC JTC 1 has released various technical reports on the wider issues of trustworthiness of ML and AI, for example ISO/IEC TR 20408 [4], but, with one exception (below), no standard on AI has yet been published. The Big Data Analytics Working Group of ISO/TC 69 “Applications of statistical methods” (ISO/TC 69/WG 12) is working on standards concerning Vocabulary and Symbols, Data Science Life Cycle and Model Validation. Furthermore, DIN and DKE, together with the German Federal Ministry of Economic Affairs and Energy, have developed the “Standardization Roadmap on Artificial Intelligence” [5], which shows which AI standards and specifications already exist and makes recommendations for areas in which there is still an urgent need for action.

The exception, just published (June 2022), is ISO/IEC 23053:2022 “Framework for Artificial Intelligence (AI) Systems Using Machine Learning (ML)”. The Standard gives concepts and definitions in AI, adherence to which (if agreed) would assist in the provision from a common stance of papers and tools in the area. It also provides a useful overview of the area.

While aspects of trustworthy ML are currently being explored in various metrology projects in isolation, there has been no systematic investigation into the requirements of trustworthy ML from a metrology point of view, and there is currently no systematic framework for addressing the challenges of trustworthy ML outlined above. Metrology has an important role to play in providing confidence in the trustworthy use of ML through validation against application-specific standards. Specifically, there is a need to address challenges in the following areas.

3.2.1 Uncertainty quantification

It is fundamental to their quality assurance that ML predictions are accompanied by reliable quantitative assessment of uncertainty for otherwise it is impossible to assess the reliability of those predictions. However, many of the most effective ML approaches, for example deep neural networks, are challenging to analyse mathematically, and this combined with their large-scale nature means that classical approaches, such as those proposed by the Guide to the Expression of Uncertainty in Measurement (GUM) framework [6, 7], are not directly applicable. There has been significant recent research devoted to uncertainty quantification of ML models, and especially deep neural networks, and various methods have been proposed. However, there has been no systematic investigation into the requirements of uncertainty quantification from a metrology point of view, and existing methods have deficiencies that need to be addressed before they meet the needs of metrology applications.

3.2.2 Generalisability and robustness

Without confidence in the ability of an ML model to generalise to new data, its predictions and associated uncertainties are unreliable. Rather than assuming knowledge of an underlying model, the ML approach is to learn a model from training data. This approach brings with it a crucial challenge: how can we be confident that a model learned on one dataset generalises to new data and is robust to perturbations in data? There are some available techniques to answer this question, including cross-validation, bootstrapping and design of experiments, but this challenge is especially pronounced for deep neural networks which are prone to overfit training data in a way that does not generalise and is not robust. Existing approaches to uncertainty evaluation in the metrology community assume the generalisability of a model. There is a need, therefore, for a systematic metrology framework for both evaluating and optimising the generalisability and robustness of ML models and deep neural networks. Since robustness is the extent to which a model is stable to perturbations in the input data, it follows that robustness is closely related to the propagation of uncertainty through ML models (see sections 3.2.1 and 3.3.1).

3.2.3 Interpretability

Confidence in the generalisability of an ML model cannot be established unless its predictions can be explained in their physical context. A vital aspect of trustworthiness of ML models in a metrology context is therefore their ability to explain the predictions they make through insights into the underlying physical process. ML algorithms typically prioritise prediction accuracy at the expense of such transparency, and deep neural networks are not inherently interpretable. Various methods for making deep neural networks interpretable have been proposed in the academic community, either by design or through post-hoc analysis, though the focus is largely upon image classification. However, there has been little systematic investigation into the requirements of interpretability from a metrology point of view, and existing methods need to be enhanced beyond the image classification context in which they were initially proposed to meet the needs of metrology applications.

3.2.4 Quality framework

A data-centric approach, as opposed to the model-centric approaches applied in more conventional analysis of measurement data, is at the heart of ML and its success. The choice and quality of the data used to train, validate, and test a ML model is critical to the performance of the model and, therefore, to the trust a user can have in a model prediction. Consideration should be given to whether the training data is appropriate for the learning task, and whether it is representative of the distribution of possible data including new (unseen) data. When data is gathered, it may contain biases, errors, and mistakes, which need to be identified and removed before training a model. In addition, the integrity and provenance of the data, and of the processes of gathering and labelling data, must be ensured. These issues are made more difficult when data are aggregated from various sources that may have different quality characteristics and when the sources are not independent of each other. The processes of training, validating and testing a ML model, as well as the algorithms, software and data used to undertake those processes, that together yield a ML prediction, should be documented to allow for traceability and transparency. Here, traceability is taken to mean the capability to keep track of the data, and model development and deployment processes, typically by means of documented identification. Furthermore, traceability facilitates auditability, enabling the assessment of data, algorithms, software, and design processes. Assessment by internal and external auditors, and the availability of their evaluation results, can contribute to the confidence that users have in the deployed solution. There is a need for guidance and tool support for these procedural issues, such as in the form of Quality Management Systems covering the data, algorithms and software on which ML models and predictions depend. Moreover, there is a need to capture good practice regarding such procedural issues in Standards and guidance documents in a widely applicable manner.

3.3 Challenges beyond the state of the art

3.3.1 Uncertainty quantification

ML is currently employed within a range of European metrology projects, including medical imaging, analysis of ECG signals, digital pathology, free-form surface reconstruction, mass spectrometry, critical dimension determination, nanoparticle image segmentation and reconstruction, and energy systems modelling. While aspects of uncertainty quantification for ML have been explored in certain metrology projects in isolation, there has been no systematic investigation into its requirements from a metrology perspective. The ‘bottom-up’ evaluation of uncertainty using a measurement model was standardised for the metrology community in the influential ‘Guide to the expression of uncertainty in measurement’ (GUM) and its supporting documents [6, 7]. However, the techniques described in these guidelines concern measurement models that are well understood and commonly capture domain knowledge and are therefore not directly applicable in the context of ML and data-driven models.

Indeed, most of the academic research into ML has been focused upon the goal of optimising predictive accuracy [8], and less attention has been directed towards evaluating the uncertainty associated with the output of ML algorithms. Bayesian uncertainty evaluation has long been established for algorithms based on classical statistical modelling such as linear regression and Gaussian processes [9]. However, uncertainty quantification is much more challenging for deep neural networks, which have recently demonstrated impressive performance in prediction accuracy [10]. Existing methods for uncertainty quantification for deep neural networks fall into two categories: methods that capture model (or epistemic) uncertainty due to insufficient training data, and methods that capture intrinsic (or aleatoric) uncertainty, which is independent of the choice of model. There is an opportunity here for metrology to help 'migrate' these perspectives on uncertainty towards the more classical approach to uncertainty quantification that involves identifying, evaluating, and propagating the various sources of uncertainty that can arise to understand how they contribute to the overall, combined uncertainty. Here, uncertainty can arise from the input and output values constituting the training data, the coverage and structure of the training data, the structure of a deep neural network including the number and depth of the layers, the hyperparameters controlling the training of a deep neural network (such as learning rate, batch size, number of epochs, etc.), the weights defining the network, etc. Doing so can be expected to provide insights about the ML and DL models deployed in applications.

Bayesian inference can in theory be used to characterise model uncertainty, but the resulting methods are computationally demanding and do not scale well to large volumes of data. Several recent approaches perform approximate Bayesian inference at reduced computational complexity, including variational inference, probabilistic back propagation, and Monte Carlo Dropout [11, 12]. Existing methods typically make assumptions that are not valid in metrology applications, for example that target distributions are independent Gaussians, and that model misfit is uniform (homoscedastic) over the input space. Meanwhile, methods for capturing intrinsic uncertainty are predominantly non-Bayesian and include quantile regression and probabilistic neural networks. More recent approaches attempt to capture model and intrinsic uncertainty simultaneously and differentiate the sources of uncertainty [13]. Bayesian neural networks can capture both epistemic uncertainty (representing ignorance about the model) and aleatoric uncertainty (arising from randomness inherent in data). The difficulties in using such models lie in the calibration of posterior uncertainties and the choice of prior distributions due to the unclear physical meaning of model parameters and hyperparameters.

Most existing methods for ML uncertainty quantification do not directly make use of knowledge about data uncertainties and do not propagate them through the model. Uncertainties associated with training data will lead to uncertainty associated with the model parameters, which is a component of model uncertainty, and is one source of the uncertainty associated with a prediction. Another source is the uncertainty associated with new input data, and the uncertainties derived from the two sources need to be correctly combined. There is a small body of literature addressing this problem for various ML models, see for example [14, 15, 16], but these methods need to be adapted and enhanced to satisfy metrology requirements.

A systematic investigation is needed to gather the metrology requirements for uncertainty quantification for ML and to understand the deficiencies in existing methods to meet those requirements. The results of that investigation can be used as the basis for identifying, selecting, and enhancing existing methods to reflect the requirements. Selected methods need to be adapted so that they can distinguish between the different sources of uncertainty that are relevant in metrology applications. New methods need to be developed that make less restrictive assumptions, for example, allowing for correlations and systematic errors (as treated in the GUM suite of documents) and more realistic (non-Gaussian) target distributions.

3.3.2 Generalisability and robustness

There are two main ways in which generalisability and robustness of DL models is currently assessed. The first is to examine empirically how the predictive performance using training data and test data compares. The second is to assess the extent to which the predictions of the model are resilient to either random, systematic, or adversarial perturbations in the inputs or to out-of-distribution inputs. Research in the ML community to date has focused almost entirely upon classification networks, as opposed to the regression networks that are of more interest in metrology applications. There is a large body of work proposing methods for improving the generalisability and robustness of neural networks, including regularisation/smoothing, data augmentation, adversarial training, and label smoothing [17, 18, 19, 20]. However, there is a need to understand how these methods compare, which factors affect the choice of method, and how to apply them in the context of the rapidly developing area of transfer learning.

Methods for evaluating the generalisability and robustness of regression models in a metrology setting need to be developed, with a focus upon methods for evaluating robustness to small changes in the training set and associated hyperparameters, including in the context of transfer learning. Existing methods for improving the generalisability and robustness of deep neural networks such as adversarial training need to be enhanced according to the requirements identified for regression models in metrology. New methods for uncertainty propagation and sensitivity analysis that exploit the known structure of deep neural networks and that scale better to large-scale problems are also needed.

3.3.3 Interpretability

Methods have been studied for extracting various types of interpretable output from DL algorithms. Some methods, such as Local Interpretable Model-Agnostic Explanations (LIME), approximate a DL model with a more transparent surrogate model [21], while other methods, such as Layerwise Relevance Propagation (LRP), seek to identify which input variables or intermediate nodes of the network are most relevant in each classification [19, 22]. Many of these methods have been designed with image classification problems in mind. An alternative approach is to train deep neural networks in such a way that they are inherently interpretable, and approaches that have been investigated include hybrid methods, architectural adjustments, and regularisation [23, 24, 25]. In this direction, physics-informed neural networks (PINNs) encode physical models in the form of partial differential equations as a component of the neural network itself, yielding more interpretable results [26].

For both approaches, what is less clear is to what extent the currently available methods satisfy the requirements for interpretability and transparency in metrology applications. A systematic investigation is needed to capture the metrology requirements for interpretability for ML, which are likely to vary significantly with application, and to understand the deficiencies in existing methods to meet those requirements. Methods designed for classification tasks need to be adapted so that they are suitable for regression models, and methods designed for images need to be adapted for the types of input data of relevance in metrology applications.

3.3.4 Quality framework

The purpose of a quality framework for software and data management is to ensure that researchers can find, access and re-use software and data, thus maximising the effectiveness and reproducibility of the research undertaken. A current activity of the EMN Mathmet is to develop and promote such a quality framework to ensure that research outputs in the forms of software and data are fit for purpose, achieve a sufficient level of quality, and are consistent with the aims of NMIs to provide quality-assured and trusted outputs. There are also other projects under way, such as [27], with the aim to support research data management. However, there is no focus on ML specifically. In the context of ML, there is a close dependence between software and data (e.g., the model architecture and the training algorithm, as well as the training data, impacts the model training and model prediction). Consequently, there is a challenge to manage together (the quality of) the software and the data. One of the issues highlighted at the online workshop “The International System of Units (SI) in FAIR digital data” [28, 29], organised by BIPM in 2021, was the importance of the quality of data, especially for “big data” applications.

The interconnection between software and data is also a challenge to the verification and the validation of ML methods, algorithms, and software. For example, understanding the impact of using different, albeit similar, training sets for repeating or reproducing a ML model is challenging due to the probabilistic nature of model training. Moreover, a general testing procedure and unique definitions of correctness criteria remain unclear as the ground truth of a ML model output often does not exist (i.e., there is a lack of an oracle) except in simple cases.

Verification and validation should also ensure that the model behaviour is reasonable when evaluated at unexpected inputs, or at an input loosely related to the training set. When used in these ways, the model may have an unpredictable behaviour that is not quantified or explained by an appropriate uncertainty. Adversarial learning is a branch of ML for tackling these issues by studying models that improve their predictive behaviour using competing neural networks.

In April 2021, the European Commission submitted its proposal for a European Union regulatory framework on artificial intelligence. The “Artificial Intelligence Act” represents the first attempt globally to regulate AI horizontally [30, 31]. The AI Act promises a “proportionate” risk-based approach that “imposes regulatory burdens only when an AI system is likely to pose high risks to fundamental rights and safety”. Targeting specific sectors and applications, the AI Act classifies risk into four levels: unacceptable, high, limited, and minimal. At least conceptually, in terms of linking required actions and interventions to decisions about risk, the approach mirrors that taken in the Quality Management Systems for research outputs in the forms of software and data, which are being promoted by the EMN Mathmet.

A discussion paper [32] (in German) emphasises three features of AI that fundamentally challenge and reshape the approach of standardisation and certification to AI. One challenge is that technical standards are quickly outdated due to the fast development of AI. The technical standards require constant updating that can be hindered by the lengthy processes involved in standardisation and certification. Another challenge is that the definition and verification of technical requirements for AI is greatly complicated by the probabilistic nature of AI systems.

Various methods of current research have been identified as contributing to the certification of AI systems, e.g., in the context of “Assurance of Machine Learning for use in Autonomous Systems” (AMLAS) [33]. These include explainable AI, formal verification, statistical validation, uncertainty quantification, and online monitoring with boundary conditions, and a number of these are identified as technical challenges within this SRA. Finally, the French metrology institute LNE has chosen a process-orientated approach for the certification of AI systems. The approach is not based on certifying the functionality of the AI system itself but, instead, focusses on the steps in the process to design, develop, evaluate and operate the AI system and assesses those against requirements set by LNE in a certification standard [34].

4 STRATEGIC TOPIC – COMPUTATIONAL MODELLING AND VIRTUAL METROLOGY

Digital representations of processes in the real world are becoming increasingly important for science and society, as impressively demonstrated by the Destination Earth – a highly accurate digital model of the Earth.

Margrethe Vestager, Executive Vice-President for a Europe fit for the Digital Age, said: "Destination Earth will improve our understanding of climate change and enable solutions at global, regional and local level. This initiative is a clear example that we cannot fight climate change without digital technologies. For example, the digital modelling of the Earth will help to predict major environmental degradation with unprecedented reliability."

European Commission Press release 20 March 2022

4.1 Introduction

Virtual metrology (VM) is a generic term for all virtual tools used in metrology, but it is not a direct measurement. In a broad sense VM can be considered as a copy of metrology in the real world. The term VM was first shaped by the semiconductor industry. According to the Semiconductor Equipment and Materials International (SEMI), virtual metrology is defined by the SEMI E133 [29] standard as

"*… the technology of prediction of post process metrology variables (either measurable or non-measurable) using process and wafer state information that could include upstream metrology and/or sensor data*".

In the semiconductor industry, metrology plays a significant role in the manufacturing process and VM is seen as an emerging technique to reduce capital expenditure and cycle time [30]. However, virtual metrology is not limited to the manufacturing processes in the semiconductor industry. Recently the significance of virtual metrology for Industry 4.0 was pointed out by Dreyfus et. al. [31],

"Virtual metrology (VM) involves estimating a product's quality directly from production process data without physically measuring it. This enables the product quality of each unit of production to be monitored in real time, while preserving the process efficiency. "

According to that development, NMIs started to consider what is needed such that VM is on the same quality level as metrology. In this perspective, the focus of VM is initially on the virtual experiment and the virtual measuring device. The development of virtual measuring devices is closely linked to real measurements.

At a later stage, when various virtual measurement devices have been developed, further metrological procedures can be approached, e.g.

- Calibration of virtual devices
- Comparison of virtual devices
- Traceability chain to a virtual/real standard

To this end, a similar quality infrastructure should be provided for VM as for real metrology. Fig. 2 schematically shows the structure of VM with its subclasses.

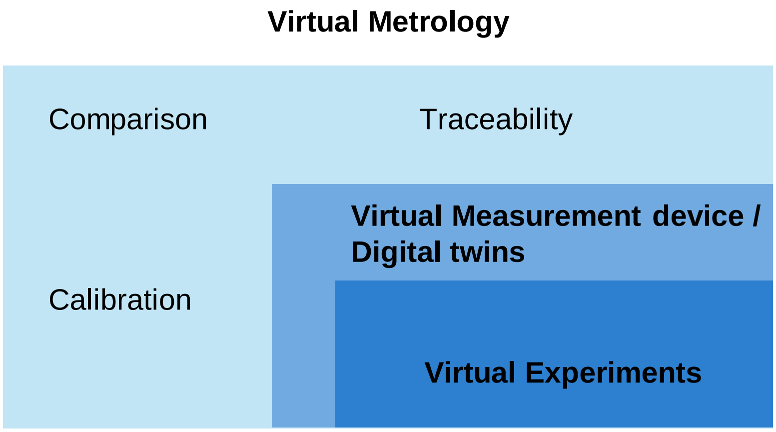

Figure 2: Schematic illustration of the structure of Virtual Metrology.

In the development of virtual experiments and virtual measurement devices, physically-driven models (e.g., via simulations) and data-driven models (e.g., using machine learning) are used for the mapping of input parameters to the output quantities of interest. To this end, artificial intelligence and computational modelling are the building blocks to develop virtual experiments and virtual measurement devices/digital twins.

Computational modelling (CM) is an approach that utilises computers to simulate and study complex systems using mathematical algorithms. From fundamental research to applications, computational modelling provides valuable information about a wide range of systems. In recent decades, simulations have been established in science as a third pillar alongside theory and experiment. In metrology, numerical simulations provide a deep insight into the measurement process, allow improvement and development of measurement methods, design novel measurement devices, facilitate the evaluation of uncertainties and are necessary to evaluate indirect measurements. The application areas range from healthcare and manufacturing to the energy sector and the environmental. With increasing computing power, it has been possible to deploy complex algorithms opening the door to new applications. New areas like virtual experiments and digital twins are developed, implemented, and applied in various industrial applications, such as advanced manufacturing processes or personalised medicine [1-2].

4.2 Needs

A **Virtual experiment** represents a digital copy of a real experiment or measurement process. It produces virtual data with properties corresponding to those of the data that would be observed in the corresponding real experiment. The use of virtual experiments becomes increasingly important in modern metrology and industrial applications [4-9]. They are utilised to specify machine tolerances or to evaluate accuracies of measurement devices [10]. In combination with a sensitivity analysis, virtual experiments are useful to identify significant sources of uncertainty [11-12].

In the refinement and optimisation of measurement processes, virtual experiments can play an essential role, enabling many experimental scenarios to be explored which is infeasible using real experiments. Virtual experiments have also been proposed for the evaluation of measurement uncertainties [8, 13-15].

Despite their previous use in the evaluation of measurement uncertainties, there are still issues to be resolved in this context, especially when ensuring compliance with current standards for uncertainty evaluation in metrology such as the GUM [16-18]. One issue is that the quantity of interest, or measurand, is often different from the observation quantity of an experiment and the model underlying a virtual experiment is therefore conceptually different from the model used for uncertainty evaluation according to GUM. For specific model structures, a GUM-compliant uncertainty evaluation for real measurements using virtual experiments has been developed recently [19]. However, treatment of the general case is still lacking, and this gap needs to be filled. Another issue is that often different ways of employing a virtual experiment for uncertainty quantification are possible (and in use) which challenges a harmonised treatment of uncertainty evaluation based on virtual experiments.

As models involved in science and engineering become very complex, their analytical solution is often compromised [22]. In many applications, simplified models must be used to obtain results within a reasonable time frame. Model order reduction techniques [23] allow physics-based models to be solved in almost real-time without affecting too much the solution accuracy. Hence, a framework is needed that allows the evaluation of model accuracy and that helps to judge whether approximate models are fit for purpose. Another important issue, in the uncertainty quantification, is to adequately account for errors in the model utilising virtual experiments and numerical modelling. To summarise, the challenges with respect to virtual experiments are:

- uncertainty evaluation as far as possible compliant with the GUM
- faithfully modelling the measuring system

- validation of the model
- solution of the model numerically correct (e.g., solutions of partial differential equations)
- speeding up the solution process
- building and validating surrogate models for uncertainty evaluation

Industry and academia define a **Digital twin** in several ways [3] and a metrological definition is still lacking. According to the white paper of Grieves [32], a digital twin in general contains three parts:

a. Physical object in Real Space
b. Virtual object in Virtual Space
c. A connection via information and data between (a) and (b).

Digital twins have an enormous growth potential. According to a research report "Digital Twin Market by Technology, Type, Application, Industry and Geography - Global Forecast to 2026" published by Markets and Markets, the digital twin market was valued at USD 3.1 billion in 2020 and is expected to reach USD 48.2 billion by 2026. Increasing demand for digital twins in the healthcare and pharmaceutical industries due to the outbreak of the COVID-19 pandemic, the changing face of maintenance, and growing adoption of digital twin solutions to cope with the COVID-19 pandemic are the key factors driving the growth of the digital twin market [28].

A digital twin enables companies to have a digital footprint for their products, from design to development, ensuring quality control throughout the entire product life cycle [3]. These twins can be used for simulating the measurement accuracy of measuring systems, for predicting the manufacturing and measurement capability of machine tools as well as for the life cycle management of parts. An intriguing example comes from the autonomous driving community. There, digital twins are used to assess autonomous driving cars in simulated driving situations. In many applications they are essential for reliable uncertainty quantification.

The basis for virtual experiments and digital twins is AI-based algorithms (see Sec. 3) and **computational modelling** (CM). The latter allows for simulation of physical models for virtual experiments or digital twins. Furthermore, CM can reduce costs and save resources by

- simplifications of technically complicated procedures
- improved uncertainty quantification and model error reduction
- replacing expensive laboratory tests
- optimisation and design of materials
- determination of material parameters from combined measurement methodologies
- quantifying physical quantities that are not directly measurable
- control of manufacturing process
- design for manufacturability as well as performance
- through-life assessment and smart maintenance of assets.

In fact, in addition to increasing the efficiency of processes, CM is also used to improve measurement modalities by:

- giving insights to the measurement (e.g., sources of uncertainty)
- decreasing measurement uncertainties
- optimising and designing measurement modalities
- designing novel measurement principles.

Ultimately, CM in metrology is pushed forward by novel trends in the industry and needs of society. The networking of systems, their independent communication and autonomous systems are playing an increasingly important role in modern applications and production processes. From autonomous driving to fully automated production lines, new demands are imposed on metrology, i.e. (i) real time metrology, (ii) enhanced metrology by predictions (iii) investigation of unexpected results and on CM, i.e., (i) sensitivity of CM, (ii) robustness of CM.

The remainder of this section will discuss how computational modelling can support the development of metrology solutions that meet these new demands for speed, robustness and reliability of results.

4.2.1 Examples of needs

Quantitative magnetic resonance imaging

Introduction and state-of-the art

(Key words: virtual experiments, inverse problems, uncertainty quantification)

Medical imaging is a branch of medical diagnostics that exploits tomographic images of the body tissues to take clinical decisions. In this sector, a special role is played by Magnetic Resonance Imaging (MRI), an imaging modality that does not involve ionising radiation and provides three-dimensional images with high spatial resolution. Every year, more than 40 million MRI scans are performed in the EU [35], with numbers increasing. The success of MRI is mainly due to its adjustable contrast capabilities (particularly with respect to soft tissue), which are unmatched and cannot be provided by any other imaging modality. Despite this success, standard MRI results mostly have a qualitative nature (i.e., they display contrast between different tissues, which must be interpreted by visual inspection and requires the presence of healthy reference tissues) that limits their objectivity and comparability. In addition, conventional MRI does not provide direct information about the nature of the pathology, nor does it quantify biomarkers. To address these issues, quantitative MRI approaches are being developed. These techniques produce images where each pixel represents the measurement of a given physical parameter; hence, they have the potential to eliminate interobserver variability and reduce the need for invasive quantitative procedures (e.g., biopsies). Quantitative MRI (qMRI) can enable new biomarkers to be identified for a plethora of pathologies that cannot be physically diagnosed, boosting early disease detection. The data provided by qMRI are independent of any acquisition or hardware-related features, leading to improved comparability of diagnostic results. Thus, they can be used to optimise the clinical path, to improve the quality of life of patients and to reduce the associated economic burden.

One example of qMRI is Electric Properties Tomography (EPT). EPT analyses the spatial distribution of the radiofrequency magnetic field that takes place during the scan, in the patient's body, to deduce the spatial distribution of the dielectric properties (conductivity and permittivity) of the tissues in the body itself. An EPT problem can be tackled in two ways [36]. One possibility is to adopt local approaches, which data coming from the pixel intensity and its neighbourhood. The other possibility is to use global approaches, which solve the EPT problem for the entire domain at once. Typically, local approaches are based on a direct elaboration of the spatial distribution of the magnetic field; they are quite efficient and straightforward to implement, but they are based on approximate models, which make them prone to errors. Global EPT approaches exploit more general (and more complicated) models, which retrieve the dielectric properties through an inverse minimisation process and imply a heavier computational burden. Recently, data-driven approaches, which exploits deep learning, have been also proposed. The development and characterisation of EPT methods requires the availability of simulated data, where the ground truth is perfectly known, produced through virtual experiments involving the numerical solution of Maxwell's equations.

Future challenges

(Key words: validation, model uncertainties, Bayesian approaches)

One of the main challenges for the full development of qMRI is the evaluation of the uncertainty at the level of each single pixel, and importantly the correlation across pixels, so that the quantitative parameters can be considered as measurements results in the strict metrological sense. This element is fundamental in order to know the reliability of the estimated data and represents the first step towards the characterisation of the qMRI techniques in terms of sensitivity and specificity in their clinical application. One possible way to achieve this goal is to set the qMRI techniques according to a Bayesian approach, which takes the statistical properties of the raw MRI data into account and provides, along with estimates of the target parameters, the joint probability distribution of the parameters themselves. It must be also noted that, in some qMRI techniques (e.g., local EPT methods based on approximate models), significant reconstruction errors may occur in some portions of the image where the effect of the applied approximations is greater. These errors have a systematic nature, but they may arise due to a number of reasons and exhibit an erratic behaviour, so that, from a practical viewpoint, they seem to be random and therefore must be accounted for in the uncertainty evaluation. Another important challenge is related to qMRI techniques based on the solution of inverse problems. Typically, these methods minimise the difference between measured data and corresponding predicted data, via simulation, based on the estimated parameters. This process involves at least three delicate aspects. First, the simulations must describe the real apparatus in a realistic way. Then, the solution of the problem must be accurate and therefore the validation of the tool adopted to perform the simulations is fundamental. Finally, because of the computational burden (which may be significant), an efficient implementation is required to make the process suitable to be adopted in clinical practice. Solutions to these problems would be also beneficial to similar non-invasive tomographic imaging techniques like Electrical Impedance Tomography (EIT) and Electrical Resistance Tomography (ERT), which respectively fall within the fields of clinical and materials science [84,85].

Scatterometry

Introduction and state-of-the art

(Key words: inverse problems, Bayesian inference, surrogate modelling, uncertainty quantification)

Scatterometry is an optical scattering technique frequently used for the characterisation of periodic nanostructures on surfaces in semiconductor industry (determination of critical dimensions) [37-40]. In contrast to other techniques like electron microscopy, optical microscopy or atomic force microscopy, scatterometry is a non-destructive and indirect method. Nanostructure geometry and associated uncertainties can be determined from diffraction patterns by solving a statistical inverse problem [41]. The inverse problem of scatterometry is in general ill-posed and regularization techniques must be applied [75]. The geometry is typically parametrised and required parameters are often obtained by weighted least squares minimisation [42], with weights derived directly from uncertainties in the measurements. However, the quality of these weights depends highly on the measurements used and itself influences the reconstruction results of the geometry parameters [43]. An alternative approach is to apply a maximum likelihood estimate, which introduces a likelihood function based on an error model and optimises weighting terms as hyper parameters instead of using predefined values. Based on the same principle but additionally including some prior knowledge is the maximum posterior approach, which is a state-of-the-art method in parameter reconstruction for optical measurements [44]. In the frameworks presented in the paper, uncertainties are typically obtained from the Fisher information or covariance matrix, which relies on an assumed shape of the posterior. However, this shape is generally unknown and hence can lead to significant errors in the evaluation of uncertainties if the actual posterior shape differs from the assumed one.

The Bayesian approach allows the integration of prior knowledge [45] and approximates the probability density function of the geometry parameters independently of any shape assumptions. Uncertainties obtained by employing posterior sampling techniques in the context of Bayesian inference are thus much more robust. On the other hand, (random) sampling schemes such as MCMC simulations require a large number of evaluations of the scattering forward model, which is not feasible for demanding computations on complicated domains. One way to reduce the computational time is to construct functional representations of the stochastic forward model that are fast to evaluate. The most popular strategies to obtain a functional surrogate for parametric differential equations are (generalised) polynomial chaos expansions [46-48].

Future challenges

(Key words: approximation errors, model uncertainties, high-dimensional inverse problems, real-time metrology)

New technologies and applications of scatterometry are leading to novel challenges in modelling. The Bayesian approach has the advantage of determining reliable uncertainties and is able to incorporate prior knowledge in a consistent manner. However, due to the statistical nature of the approach, it is in many cases computationally expensive leading to long evaluation times. For the requirements of real-time metrology, novel algorithms must be developed to approximate the Bayesian posterior distribution. A similar challenge is the solution of high-dimensional inverse problems. When the number of parameters sought is very large, the problem becomes highly dimensional and numerically intractable due to the curse of dimensionality. Some model reduction methods based on tensor decomposition [48] provide promising preliminary results but need to be further developed with regard to metrological requirements. Both approaches, tensor decompositions and Bayesian approximations, introduce additional uncertainties. While measurement uncertainties are often adequately addressed, model and approximation uncertainties are often neglected. However, these latter uncertainties are necessary for a reliable expression of the uncertainties.

Laser flash thermal diffusivity

Introduction and state of the art

(Key words: inverse problems, Bayesian inference, uncertainty quantification)

The laser flash thermal diffusivity experiment is used to determine the thermal diffusivity of homogeneous isotropic materials from measurements of surface temperature of a sample as it undergoes heating by a short pulse of laser light. The thermal conductivity of a material can be calculated from the thermal diffusivity if the specific heat capacity and density of the material are known. Thermal conductivity is a key property for thermal design and assessment of thermal performance of systems. In particular, knowledge of the thermal conductivity is essential for design of the thermal barrier coating systems that are commonly used to protect high value components such as turbine blades from the high temperatures and corrosive atmospheres created during typical operating conditions. These coating systems generally consist of three layers: a substrate, a bond coat and a topcoat. The topcoat and the bond coat are typically applied using spraying or particle deposition techniques and are generally tens of microns thick.

The experiment uses a cylindrical sample of radius around 25 mm and thickness typically around 2 mm. For many of the materials used in thermal barrier coating systems, it is not possible to create a sample with the required dimensions without significantly affecting the microstructure of the material and hence the thermal conductivity of the material.

A computational model of the laser flash thermal diffusivity experiment on a layered sample has been constructed and wrapped in an optimisation routine [49]. This model has made it possible to determine the thermal conductivity of layers within the sample by using the optimisation routine to minimise the difference between measured temperatures and model predictions. This approach has been used in conjunction with Latin hypercube sampling to evaluate associated uncertainties and has been implemented using probabilistic finite elements in a Bayesian framework to examine the effects of some parameters [50].

The model has been further extended to provide a quantified description of thermal bond quality by introducing a conductance parameter at the interfaces between layers [51]. This extension allows the system as a whole to be characterised, meaning that the effects of poor thermal bonds, which may evolve during operation, can be incorporated into performance models. The extended models also make it possible to include thermal imaging temperature measurements into the objective function to determine the extent of the region of poor bond quality.

Future challenges

The CM technique as it currently exists has some limitations and shortcomings. The heat transfer mechanisms implemented within the model may not reflect the mechanisms occurring within some materials. In particular, the microstructure of some materials (for instance those with a strongly columnar structure) may mean that the assumption of an isotropic material is invalid, and that consideration of the microstructure needs to be implemented in more detail.

The boundary conditions applied within the model would also benefit from further investigation. At present it is assumed that the sample is not in conductive contact with any part of the measuring system, but some measurements have suggested that conductive losses may be present. A more general spatially varying boundary condition would allow this possibility to be considered.

Energy performance in the building sector

Introduction and state of the art

(Key words: inverse problems, Bayesian inference, uncertainty quantification)

In the context of energy and environmental renovation, major advances are expected and necessary in the building sector. For existing buildings, the reduction of energy consumption requires a better assessment of the energy performance of buildings and their improvement through rehabilitation actions. Particular attention must be paid to in situ evaluation and control of the thermal performance of the buildings before and after a rehabilitation action in order to prevent any defect and thus to obtain a building with the expected performances. The construction must also be "sustainable", e.g., with a low environmental impact by using local materials with less effect on natural resources. In this respect, bio-based walls using hemp concrete, wood fibre, cellulose wadding and raw earth are particularly promising materials to build highly insulated walls.

Due to complex thermo-hydric transfers arising in bio-based materials, the indirect measurement of the thermal performance of a wall requires the inversion of a computationally expensive forward model (in 2D or 3D) producing costly virtual measurements (e.g., of the temperature at the surface of the wall). Inversion-based indirect measurements usually involve the estimation of so-called calibration (tuning) parameters of the forward model. Also, the inversion process relies on physical (in-situ) measurements obtained in pre-determined experimental conditions.

Uncertainty propagation in inversion problems is generally addressed by neglecting input uncertainty sources or considering them small enough to be aggregated with uncertainties associated with the physical measurements [52], which is not compliant with the uncertainty propagation principles of the GUM. Indeed, the difficulty that arises for uncertainty propagation in inversion models is that there is usually no close functional form to describe the relation between the input uncertainty sources and the estimate of the calibration parameters [53].

In building applications, this issue has been addressed recently for cheap 1D forward models where many approaches have been developed to create approximate relationships through which uncertainties are propagated [54-57]. Recent work [58] combines Metropolis–Hastings [76-77] with Monte Carlo sampling from the uncertainty sources in an attempt of full Bayesian inversion. More recently [59] proposed a full Bayesian approach for input uncertainty propagation using an excess variance approach.

In those studies, the Bayesian framework is usually preferred for the convenient representation of uncertainty that it conveys and the convenient use of prior distributions acting like a regularisation technique. The downside is that Bayesian approaches rely on computationally expensive estimation algorithms.

Future challenges

(Key words: cost reduction techniques, approximation errors, uncertainty quantification, time series)

Reducing the computational cost of both the forward model and the Bayesian procedure is a major challenge for the estimation of thermal performance of walls using inversion models. Solving this issue requires smart and fast approximations of both the forward model and the posterior distributions of the parameters.

Such approaches could combine model order reduction techniques [60-61] and multi-fidelity approaches [52, 62-65] in the Bayesian framework. In addition, the propagation of the input uncertainty associated with time series measurements must be carefully addressed.

For a comprehensive uncertainty analysis, the uncertainty propagation should also consider model or approximation errors.

Design of miniaturised magnetic sensors

Introduction and state-of-the art

(Key words: modelling, optimisation, sensors)

In recent years, advances in nanostructure fabrication have driven the design of miniaturised magnetic sensors for innovative applications in industrial, automotive, aerospace, ICT, green technology and medical fields. In the biomedical area, they can be integrated within lab-on-a-chip systems for point-of-care uses, like the detection of functionalised magnetic labels, e.g., magnetic nanoparticles or microbeads, employed for molecular targeting, biomolecule quantification, sample purification or cell manipulation. Magnetoresistance and Hall effect-based devices have optimal properties, offering very high sensitivity and magnetic field resolution, as well as easy miniaturisation and lab-on-a-chip integration. Another application in the biomedical field regards the measurement of biomagnetic signals from human organs, which remains a challenging area of research, due to the very low magnitude of the magnetic field to be detected, in the range of picotesla for magnetocardiography and magnetomyography down to femtotesla for magnetoencephalography [66]. Magnetic tunnel junctions are currently considered as the most competitive sensors that could achieve such extremely high sensitivity at room temperature and low-frequency domain, opening the way also towards the development of sensor networks for remote healthcare monitoring.

The design and optimisation of high-sensitivity magnetic sensors can be performed by means of computational models able to simulate the physical phenomena at the basis of the mechanism of magnetic field detection [67]. As an example, micromagnetic numerical tools have been extensively used to study magnetoresistive devices with different geometrical, structural and material properties of the sensing elements. Special attention has been focused on giant magnetoresistance multi-layers, spin-valves, magnetic tunnel junctions, anisotropic magnetoresistance and planar Hall effect-based devices. At the same time, several studies have been orientated to the numerical implementation of classical models of the Hall effect in the diffusive transport regime for the evaluation of the magnetic field resolution of micron-sized semiconductor or graphene Hall probes [68].

Special attention has been devoted to the implementation of numerical codes, e.g., based on finite element or finite difference methods, able to emulate the mechanism of magnetic field detection, with the final aim of supporting the interpretation of experimental results and providing inputs for sensor engineering and optimisation.

Future challenges

(Key words: parallel computing, high performance computing)

One of the challenges of the computational modelling of miniaturised magnetic field sensors is the need for developing highly efficient solvers, able to describe the multi-physics and multi-scale nature of the involved physical phenomena with sufficient accuracy. To this aim, non-standard numerical techniques (e.g., fast multiple methods, FFT-based approaches) can represent a valid approach to reduce computational time, enabling the treatment of large-scale problems. At the same time, parallel computing with the aid of graphics processing units (GPUs) is becoming a diffused strategy but requires a good knowledge of the computational architecture to exploit as much as possible its potential and develop computationally efficient solvers.

Another challenge is the numerical validation process, which needs the availability of experimental data for mutual comparison; to this aim, the combination of different characterisation techniques of sensor geometry, material properties and performances can be fundamental to provide feedbacks for the modelling and uncertainty evaluation.

Also, the engineering and array integration of magnetic field sensors for wearable medical devices and biosensing applications (e.g., magnetocardiography and magnetoencephalography) is a very challenging task. To this aim, ad hoc machine learning techniques can be used to solve the arising optimisation problem, training the algorithms with both synthetic datasets (e.g., from extended micromagnetic modelling in the case of magnetoresistive sensors) and experimental datasets.

Utility networks

Introduction and state-of-the art

(Key words: state estimation, model validation, uncertainty quantification, real-time metrology)

For successfully operating utility networks like the electrical grid [69] or the gas distribution network [70] it is necessary to know the internal state of the network. For example, for the electrical grid, it is necessary to monitor current flow through a line since excessive current might damage the infrastructure. Also, the frequency of the voltage signal should not fall too much below its nominal value (50 Hz in Europe). In the case of detecting an event, it is highly desirable to be able to locate its root cause in place and time. For the gas grid, it is similarly desirable to be able to locate causes of problems. Also, the gas composition may not be the same in different locations in the network, due to local generation of biogas and/or hydrogen, which might be fed into the grid. So, for both types of network, state estimation and reconstruction are of high interest. In both cases the number of measurements is limited, and thus optimal usage of the measurement data should be made. It is thereby crucial that the models have been validated as far as possible and that the limitations and uncertainties of model predications are clear.

Future challenges

(Key words: validation, model uncertainties, uncertainty quantification)

In both the electrical and the gas grid there is a large increase of local generation (solar panels and wind turbines for the electrical grid, bio-gas production for the gas grid). At the same time, the consumption by end-users can be monitored at a much finer scale by means of reading the data of smart energy meters (if permission is granted). The models are becoming more dynamic instead of some static. The models are becoming more complex, and the amount of available data is increasing, and their mathematical solution and validation is thus becoming more challenging [71, 72] There is also a tendency to use data-driven methods from machine learning to address specific questions. For example, in [73] a neural network is used to estimate the grid inertia. These models need validation as well, and the uncertainty of all model outputs need to be quantified.

4.3 State of the art

Computational modelling has become an integral part in many fields of modern metrology [8,15] and a key driver in many industrial applications [24-25]. In metrology, CM is used to understand measurement processes, optimise measurement setups, evaluate uncertainties and indirect measurements. In recent years, significant progress has been made in the accurate modelling of measurement processes (virtual experiments) and its inverse modelling (statistical inverse problems) including uncertainty quantification.

4.3.1 Virtual experiments and digital twins

Exploring the accuracy of measurement devices, specifying machine tolerances, and identifying significant sources of uncertainty are examples in which virtual experiments are employed nowadays [24].

In combination with Monte Carlo methods, virtual experiments are used for evaluating uncertainties [8,15,25]. However, the results of these approaches generally differ from a GUM-compliant uncertainty evaluation, which marks the *de facto* standard for uncertainty evaluation in metrology. The basis of the GUM is an evaluation model relating the measurand to all meaningful influencing quantities, including the observations obtained in an experiment. The evaluation model typically forms a partially inverse model to the forward model of the involved experiment. Since virtual experiments emulate the latter, Monte Carlo runs of a virtual experiment and the Monte Carlo methods for uncertainty evaluation [17, 18] are conceptually different approaches. To facilitate the use of results from virtual experiments in traceable measurement chains, a GUM-compliant uncertainty evaluation based on virtual measurements is needed. Whilst this has been achieved for linear models recently [19], a methodology for the general case is still lacking. Filling this gap is a key requirement to ensure traceability for the routine employment of virtual experiments in metrology and industrial applications.

Digital twins are used in a variety of applications, such as advanced manufacturing [26], healthcare [27], and smart city environments [28]. According to [29], a digital twin is "a set of virtual information constructs that mimics the structure, context and behaviour of an individual/unique physical asset, or a group of physical assets, is dynamically updated with data from its physical twin throughout its life cycle and informs decisions that realise value". The essential elements of a digital twin are a virtual representation (model), a physical realisation (asset), and a transfer of data/information (connected) between the two [29]. The virtual representation can consist of purely data-based models, physics-based models, or a combination of both. In [22], the latter is called a "hybrid twin" emphasising that the digital counterpart consists of both parts. In such a setting, the physics-based model is often not exact (either because the complex process cannot be modelled with all details or because the system is so computationally expensive that a surrogate model needs to be used). The error of the model can then be reduced by adapting it to actual measurement data (data-driven modelling). However, in such a setting, it is not clear how the uncertainty of the whole system can be determined. Furthermore, the question arises of how a model that has been calibrated with measurement data can be validated. Rules for validation, see [21], need to be adapted for such a framework.

The key value of digital twins is the closed loop with manufacturing and design, early warning, continuous prediction as well as optimising measurement activities via communication among different devices. However, for metrology, the uncertainty evaluation is usually only available for the "static" case, where it is assumed that the digital model does not change over time [8,24]. To account for time-dependent influences, such as mechanical deformations, thermal drifts, or vibrations, the dynamical case needs to be considered. Hence, the uncertainty quantification for digital twin needs to be developed using data from actual measurements collected in real time, i.e., use of Kalman filtering.

It is useful to note that what "real time" means will vary from application to application. A digital twin is typically used to monitor a specific aspect of a system that is known to change over time and to take some action or make some decision based on the state of that aspect. "Real time" therefore needs to reflect the time scale over which the system is changing, so that the change in the system can be identified and is independent of how often new data is available. For a metrological system, this time scale will be affected by a combination of the variability of the local conditions (temperature, vibration, etc.) and the inherent stability of the artefact being measured.

4.3.2 Inverse problems and uncertainty quantification

A central feature of metrology is that an uncertainty statement must be included in all measurement results. The concept of traceability depends fundamentally on uncertainty quantification. The treatment of measurement models and the evaluation of associated measurement uncertainties are summarised in the GUM suite of documents and its supplements that are published by the Joint Committee for Guides in Metrology (JCGM) [16]. Evaluating the measurement uncertainty associated with the outputs of the models arising in important applications is often difficult and challenging. The simple analytical treatment outlined in the GUM and the Monte Carlo approach of GUMS1 have been successfully applied to many cases where the measurand is a single output quantity and the measurement model is either given in an explicit form or solvable with small computational effort [17]. The GUM approach essentially involves uncertainty propagation through a known direct model whereas multivariate and implicit modes are addressed in GUMS2, by means of both the law of propagation of uncertainty and the Monte Carlo approach [18]. However, for most inverse problems Bayesian inference methods are more appropriate. The successful European project on "Novel mathematical and statistical methods for uncertainty evaluation" developed tools for uncertainty quantification for a bigger range of models, e.g., by developing a Bayesian approach to regression and inverse problems, and by using smart sampling schemes to enable uncertainty quantification in computationally expensive physical models [23].

4.4 Challenges and trends

Significant progress has been made in CM and VM, although fundamental questions still remain unanswered. While many systems are qualitatively well described by complex computational models, there are still problems in accurately modelling measurement processes. Therefore, the power of CM for metrology has not yet been fully exploited for many applications. For a broader acceptance and application, it is necessary that computational models address the following challenges: *Reliability and Efficiency of numerical algorithms.*

4.4.1 Reliability

Reliable simulations are a key for the development of trust in CM and VM. The reproducibility of results is particularly crucial for reliability, similar to experiments. In a recent survey, numerous computational studies were considered, and the efforts carried out to reproduce the results are measured. Within some reasonable time limits (one week) for the authors it was not possible to reproduce the results from any of the considered papers [33]. This impressively shows how urgently a framework is needed to make CM reliable. For reliability in CM, five conditions can be defined [34]:

- Validation:
  - confirmation that the results agree with experiments, the ultimate test for credibility of a simulation
  - updating the model according to environmental changes
- Verification:
  - the software accords with its specification
  - there are no implementation errors or bugs
  - the algorithm is numerically stable [74]
- Reproducibility:
  - ensure that results are repeatable and reproducible within a pre-defined quality range
- Comprehensibility
  - understanding the algorithms
  - documentation of algorithms and input data
  - documentation of parameter values and associated uncertainties
  - training of experts
- Uncertainty Quantification:
  - Identification of the provenance of errors within the computational model
  - sources of uncertainty, e.g., approximation errors, parameter uncertainties, imperfection of models
  - propagation of uncertainty.

4.4.2 Efficiency

In metrology, problems often arise where statistical methods and computationally expensive models are applied. These problems often lead to situations where the computing time, and, in the near future, energy consumption becomes a limitation for the application. To overcome the latter, two paths can be taken. Efficient implementation of algorithms and optimisation of computer code are able to decrease computational time and energy consumption. Another option to increase efficiency is to use advanced mathematical models and algorithms, i.e.:

- model reduction techniques
- surrogate modelling (e.g., approximation of the forward model)
- modelling with sparse grids
- model selection methods
- real-time simulations.

Agile, efficient algorithms are of particular interest when performance is to be provided on edge devices (e.g., device that connects the local network to an external network) rather than in the cloud, where memory and computing power are large. A particular example that underpins these challenges are large-scale problems. Large-scale problems are problems that involve large scale (big) data or high dimensional parameter spaces. To treat large scale problems efficiently, a substantial extension of novel computational methods is required, see e.g. [14-17].

In order to make large scale problems computationally tractable, approximate or surrogate models are used instead. However, the uncertainty contribution due to the use of an approximate model could be significant and difficult to quantify.

For the purpose of making large scale inverse problems computationally tractable and well posed, prior information (e.g., smoothness, sparsity) can be included. Imaging is a case in hand where appropriate prior information has dramatically improved the performance of these systems. The effect of the choice of prior information on the estimated parameters can be significant but is difficult to quantify.

Markov chain Monte Carlo (MCMC) sampling methods [76-77] have enabled inverse problems to be analysed correctly using Bayesian inference. However, the convergence rates are quite slow, so that the computational effort can become excessive for complex models. New approaches (sequential Monte Carlo, multi-level Monte Carlo, approximate Bayesian computation) [81-83] potentially can address this issue but the performance of these newer algorithms needs to be assessed.

A clone of true-life

A true-life clone of an object or measurement device in the physical world – a digital twin or a virtual measurement device – gives new insights to the real-world performance of products, processes and metrology. These virtual clones have a huge potential for efficient engineering, better understanding of the real object throughout its life cycle, avoiding potential problems, cost and resources saving. The difference between computational modelling and the recent development of digital twins and virtual experiments is not that a specific model for the application is used, but rather all digital knowledge ranging from models to data during the complete lifetime cycle is used. Thus, model-based approaches on classical simulations and approaches based on data analytics are integrated to form a virtual clone of the true-life object. A particular challenge here is that various research fields must be linked together. From data analysis, statistics, modelling to numerics, many factors are involved.

Virtual experiments

As for CM in general, the key challenge of modelling virtual experiments is to ensure the reliability and trustworthiness of virtual experiments in metrology. Such reliability and trustworthiness require verified and validated models, uncertainty evaluation in accordance with current standards, traceable and reliable virtual measurements, smart data selection as well as robust and reliable algorithms that combine model-based and data-driven methods. Current challenges require development of:

- advanced mathematical modelling techniques, such as multiscale and multiphysics modelling, and their use.
- methods for evaluating the uncertainty associated with real measurements by using the results from corresponding virtual experiments. Method development should be in line with current standards for uncertainty evaluation in metrology.
- methods for uncertainty quantification for virtual measurement devices representing complex measurement processes and mechanisms. Such activity requires the inclusion of dynamical influences (such as thermal drift or vibrations) in the model in addition to the continual updating (calibration and enrichment) of the underlying models based on fresh data, e.g., for life cycle management.
- approaches for the validation of virtual experiments. These approaches include methods for the verification of the models as well as statistical procedures for the assessment of differences between measurements of calibrated standards and corresponding data from their virtual counterpart.

- methods to account for errors in the model, specifically for computationally expensive systems, where surrogate models (which contain, e.g., approximation errors) often need to be used. These methods include those involving the treatment of models that have been calibrated with the help of experimental data.

A metrologically sound implementation of virtual experiments is a crucial element for digital twins. Up to now, uncertainty evaluation according to current standards by means of virtual measurement systems has only been realised for particular “static” cases, i.e., for the case where the underlying model is assumed to be independent of time. By extending the existing frameworks to systems with, e.g., time-dependent influences, many more metrological applications can be covered. First attempts to include dynamic influences have been made [20]. However, the number of contributors addressed as well as the methods used for their inclusion in digital twin models are currently very limited.

From the point of view of metrology, rigorous methods to include the uncertainty of known influence factors as well as approaches to identify additional contributions to measurement uncertainty are still lacking (but see [86]). These methods and approaches are, however, essential to optimise performance, allow validation, and ensure traceability. Each numerical model must be validated before it is routinely used, e.g., to support conformity assessment (see [86]). or take safety-relevant decisions. A transfer and adaptation of procedures such as those in [21] to virtual measurement processes will increase the trust in the results derived by virtual experiments. One important aspect in the validation of a virtual experiment is a framework to compare and statistically assess virtual and real measurement values.

Digital twins

Even though the application of digital twins in metrology is limited at present, a larger use is expected in the future. However, the building blocks of a digital twin are related to AI, CM and classical data analysis. In general, digital twins can be considered as consisting of

- a real-time virtual representation of a real-world physical system or process by a model;
- a physical object/system that is changing over time (either progressively as in the case of wear or in response to its environment);
- a stream of measurement data taken from the physical object; and
- a method of updating the model based on the data stream to reflect the change in the physical object.

The updating process produces an estimate of the parameter associated with the aspect of the physical system that is changing over time. This parameter often cannot be measured directly. A digital twin is therefore often seeking to solve an inverse problem.

The twin is typically used either to make a decision about the physical system, or to provide a quantified estimate of some aspect of the physical system. In both cases the output of a digital twin, and particularly a metrological digital twin, should have an associated uncertainty.

The presence of a model within a digital twin means that many of the challenges in digital twins are the same as the challenges in computational modelling more broadly, but it is useful to consider the drivers of these challenges. Many of the applications for which digital twins are of interest are safety-critical, infrastructure monitoring being a common example, and so reliable validation of models is particularly important.

Selection of the best method for updating digital twins is not a solved problem. For industrial digital twins, the data stream is likely to be noisy and potentially have large associated uncertainties and significant correlation, which may mean that traditional optimisation methods may struggle to converge. Many digital twins use computationally expensive models created using proprietary “black box” software, which further limits the range of optimisation methods that can be applied effectively to the updating problem. The requirement to produce a value and an associated uncertainty makes methods such as data assimilation [78-80] attractive, but these seem not to have been investigated for digital twinning as yet.

The components listed above mean that the digital twin has four different timescales that need to be consistent so that the end goal can be achieved in a timely manner:

- The time required to update and evaluate the model;
- The time scale over which the physical object is changing;
- The time required to gather the data (i.e., the time period between updates of the data stream);
- The time by which a decision must be made or a quantified estimate obtained.

There are several ways in which these timescales can come into conflict. The data time scale needs to be sufficiently short compared to the physical object's time scale so that the change in the physical object is captured in enough detail within the data. In many cases the time scale of data gathering can be controlled to match the other time scales so this is often not a problem, but for cases with multiple data streams feeding into a single model there can be challenges associated with synchronisation and with the effects of problems such as jitter on the uncertainties associated with the data.

In some cases where high-fidelity physics-based models are required the time required to solve the model can be hours or in some cases days. For model updating processes that require repeated evaluation of the model (as would be expected for processes that produce an uncertainty), there can be conflicts between this time scale and the decision-making time scale. For digital twins to be useful in time-pressured environments, demonstrably reliable surrogate models will need to be developed. Again, this requirement is not unique to digital twins.

At present, digital twins are usually descriptions of single objects such as a road or railway bridge or a co-ordinate measuring machine. Many organisations would like to be able to link digital twins that describe different aspects of a system or different components of that system. For instance, a car manufacturer may want to link a digital twin that estimates brake wear from acoustic measurements to a full car digital twin that uses weather data and vehicle condition to estimate safe stopping distances for a driver. These linked aspects require development of methods to map model results, and associated uncertainties, between different types of model and potentially between different length and time scales.

The remarks above largely focus on industrial digital twins where the physical twin is in an uncontrolled environment and the ability to make quick decisions is often important, rather than metrological digital twins within a controlled laboratory.

Metrological digital twins will face the same challenges, but it is likely that the main challenge for metrological digital twins is the reliable evaluation of the uncertainty associated with their outputs. As is discussed in the virtual measurement section of this document, for the output of a digital twin to be used within an NMI, the uncertainties it generates must be calculated using a rigorous approach that is consistent with the GUM framework. This requirement is likely to mean that new methods of model updating that are consistent with the framework must be developed.

Furthermore, digital twins must not only fulfil the reliability requirements mentioned above (e.g., reproducibility, documentation in a machine-readable format) but additionally:

- data quality framework
- description of the data processing pipeline to the results in the data base (e.g., ontologies)

5 DATA ANALYSIS AND UNCERTAINTY EVALUATION

5.1 Introduction

Alongside the two research topics described in chapters 3 and 4 of this SRA, it is recognised that an increased use of mathematics and statistics is required to face new measurement challenges. In many areas of metrology, mathematics and statistics play an essential role, for example in imaging by tomography, in the purity analysis of nominally pure materials used for producing certified reference materials, and in the use of measurement data for fiscal metering of energy. In each of these areas, the challenges are different, but they have in common that with a simplistic approach to the modelling and uncertainty evaluation, unsatisfactory results are obtained which are detrimental to the overall performance of the application (e.g., tumor size measurement using tomography). In this chapter, an overview of challenges related to the evaluation of measurement data and the evaluation of measurement uncertainty is presented.

5.2 State-of-the-art and beyond

5.2.1 Uncertainty evaluation for small sample sizes

Small sample sizes occur frequently in metrology. Obtaining larger sets of data can be prohibitive (e.g., in destructive testing), prohibitively expensive, or impossible due to other practical constraints (e.g., time in a dynamic setting). Such small data sets, with often no more than two or three observations, traditionally pose a challenge for statistics and uncertainty evaluation. Various approaches for dealing with small sample sizes have been proposed, e.g., using large student *t*-factors, pooled standard uncertainties, and uncertainties evaluated based on other larger samples that may or may not be representative. This area is still underdeveloped. Bayesian methods could be employed, but these require a proper inclusion of prior knowledge about, e.g., the performance of the measurement procedure used. Another obstacle is that often these Bayesian methods need to be appropriately simplified for use in routine measurement. Guidance for practical cases is still lacking.

5.2.2 Uncertainty evaluation for large sample sizes

Traditional uncertainty evaluation methods generally work best for moderate to large sample sizes. When the amount of measurement data and/or the number of model inputs becomes large, problems can occur (see also model reduction techniques in section 4). With large numbers of repetitions, the assumption that the data are mutually independent and identically distributed is usually problematic and leads to understating measurement uncertainty. Time series analysis is relatively new in many areas of metrology, yet it is one of the ways to properly evaluate large datasets [1-2]. In the case of many inputs to a model, the Monte Carlo approach may be no longer feasible, and smart sampling methods such as those presented in [3] and references therein may be needed. All these issues require a proper mathematical and statistical treatment for producing valid estimates and uncertainties. Like the case of small sample sizes, also a translation from a sophisticated model to a practically implementable model may be required. Guidance for practical cases is largely lacking.

5.2.3 Bayesian statistics

Bayesian statistical methods offer a general approach for evaluating measurement data and its uncertainty. In the last two decades it has attracted substantial attention from experts working at NMIs and it has been applied to several use cases, e.g., data reduction in (key) comparisons, evaluation of homogeneity and stability in reference material production, type A evaluation of standard uncertainty. Open research questions are related to its implementation and effectiveness in still other use cases, its efficient implementation in computational problems (e.g., scatterometry, see example needs in section 4) and how to use it in combination with errors-in-variables models. Other questions relate to the selection of prior distributions. Also, sometimes counterintuitive results are obtained, like the fact that the calculated uncertainty for a specific measurand depends on the number of other measurands being inferred at the same time using the data.

5.2.4 Analysing key comparison data

Key comparisons play an underpinning role to the trust in metrology on a worldwide scale, based on the Mutual Recognition Agreement (MRA) [9]. An essential part herein is played by a mathematically sound evaluation of key comparison data. Various methods and alternatives have been presented and new methods are derived covering yet other cases. In view of the vast number of metrology domains, comparison schemes and uncertainty structures, the domain of analysing (key) comparison data cannot be considered as solved.

5.2.5 Statistical tests

Hypothesis testing and statistical tests can be performed, e.g., for conformity assessment [4-5], and more widely in the comparison of measurement results. Many issues can be at play here, such as measurement data may not be distributed normally or may be correlated. In such circumstances, appropriate statistical tests need to be developed to address the particulars of the measurement data. The development of statistical tests for conformity assessment and comparing measurement results is a rapidly developing area, and, as in many cases, each problem may ask for another treatment. As the number of applications and variations is large, new method development will always remain needed.

5.2.6 Model design, model selection, model validation, accounting for model errors

The topics of model design, selection, model validation and accounting for model errors in uncertainty evaluation are main themes in all mathematical model building in metrology. Current trends are that models become more data-driven, i.e., they describe the features of a data set. Data models and measurement models are at the heart of any application of metrology, and every application requires its own model. Also, modelling measurements that involve a complex mix of systematic and random effects and correlations can be challenging. The same applies to models that need to address constraints (such as that an absolute temperature cannot be negative.) For existing models, there can still be issues with the evaluation of the measurement uncertainty of the model outputs, which require further developing these models, or combining them with the law of propagation of uncertainty, Monte Carlo method or another method to quantify measurement uncertainty. Mathematical and statistical support and new method development are essential for many novel applications in metrology.

Some guidance exists for evaluating the uncertainty using observation models that involve solving an inverse problem, and for analysing regression problems [6]. However, also in the domain of data analysis and uncertainty evaluation, new model types with more complex structures, interdependencies and correlations are required as the applications in which they are used demand a proper treatment of the measurement data. Well-founded mathematical and statistical analysis concerning the uncertainty evaluation of the model outputs are required to support advances in many metrological areas.

5.2.7 The GUM suite of documents

As illustrated in the sections above, data analysis and uncertainty evaluation for ‘classical’ metrology problems cannot be considered as a ‘solved research domain’ from a mathematical and statistical point of view. This also means that the GUM suite of documents [7] is not finished; on the contrary, it is currently under development by the Joint Committee for Guides in Metrology — JCGM WG1 (in which several Mathmet members have actively participated for many years) and will need continuous enrichment and update. Some new documents are intended to be developed, on topics such as statistical models and data analysis for interlaboratory studies, least squares approaches and Bayesian methods [8]. Some of the documents already published could take advantage from future extensions, revisions or ancillary documents, such as a collection of examples on measurement uncertainty evaluation, partly to support JCGM 100:2008, JCGM 101:2008 and JCGM 102:2011, and the generalisation of JCGM 106:2012 conformity assessment approach to multivariate models and to a sample of items. Making a guidance document that is sufficiently general and balances the views and requirements from the entire metrological community is an art in itself.

6 SUMMARY

This document presented the Strategic Research Agenda (SRA) prepared by the EMN Mathmet within the framework of the EMPIR Joint Network Project “Support for the European Metrology Network on Mathematics and Statistics” based on a stakeholder consultation process and in line with the strategies of EURAMET and the participating NMIs.

The SRA identified an overall vision for the EMN Mathmet to *ensure quality and trust in algorithms, software tools and data for metrology.*

Two emerging research topics of great importance for the metrology community were illuminated: *Artificial Intelligence and Machine Learning* and *Computational Modelling and Virtual Metrology*. Both research topics were presented and key needs and challenges from the metrological perspective were identified:

- Artificial Intelligence and Machine Learning:
  - Uncertainty quantification
  - Generalisability and robustness
  - Interpretability
  - Quality framework
- Computational Modelling and Virtual Metrology:
  - Reliability and quality
  - Efficiency and real-time calculations
  - Uncertainty quantification

To demonstrate that behind the challenges are metrological applications, current joint research (EMPIR) projects and use cases related to both research topics are presented.

To address the key challenges in the research topics, the EMN Mathmet uses expertise and capabilities in modelling and data analysis. Some of the known concepts can be further elaborated and applied to the new research topics, while other concepts need to be newly developed. However, in the classical research topics of data analysis and uncertainty evaluation are still open questions that should be addressed. The needs and challenges regarding these classical areas were identified and discussed.

7 MANY THANKS TO THE STAKEHOLDER ADVISORY COMMITTEE

Stakeholder	Interests
Paolo Moscatti, President of Eurolab	• Exchange of information on experience, calibration and measurement services of experienced • Measurement uncertainty
Antonio Pievatolo, past President of ENBIS	• Measurement uncertainty • Design of experiments • Process modeling and control • Reliability and safety • Artificial intelligence and machine learning • Predictive analytics • Virtual sensors
Pavel Klenovský, Chair of WELMEC	• Measurement uncertainty
Horst Lewitschnig, Lead Principal Engineer at Infineon Technologies Austria AG	• Autonomous driving • Automotive quality • Propagation of uncertainties • Sensors
Tjerk Timan, Senior Scientist at TNO	• Big data • Artificial intelligence and machine learning • Modelling • Standardisation
John Greenwood, Assessment Manager at UKAS	• Measurement uncertainty
Emma Woolliams, Chair of the EMN COO	• Measurement uncertainty • Artificial intelligence and machine learning • Modelling
Steve Ellison, Chair of Eurachem MU & Traceability WG	• Measurement uncertainty • Standardisation
Jérôme Lopez, Director of CFM	• Measurement uncertainty

8 REFERENCES

8.1 References for Introduction

1. Sené, M., Gilmore, I. and Janssen, JT. Metrology is key to reproducing results. Nature 547, 397–399 (2017). https://doi.org/10.1038/547397a
2. CIPM Task Group on the digital SI, *BIPM-WS-digital-SI/2021-Vision* (2021), https://www.bipm.org/documents/20126/46590079/WIP+Grand_Vision_v3.4.pdf/aaeccfe3-0abf-1aaf-ea05-25bf1fb2819f
3. Moritz M, Redlich T, Günyar S, Winter L, Wulfsberg JP. 2019 On the Economic Value of Open Source Hardware – Case Study of an Open Source Magnetic Resonance Imaging Scanner. J. Open Hardw. 3. (doi:10.5334/joh.14)
4. Marques JP, Simonis FFJ, Webb AG. 2019 Low-field MRI: An MR physics perspective. J. Magn. Reson. Imaging 49, 1528–1542. (doi:10.1002/jmri.26637)
5. https://www.opensourceimaging.org/2023/01/09/first-open-source-mri-scanner-presented-the-osii-one/
6. European Commission. 2021 EU4Health Programme 2021-2027.
7. ENTSO-E. Completing the map – Power system needs in 2030 and 2040, 2021: Completing the map – Power system needs in 2030 and 2040 (windows.net).
8. National Grid ESO. System Inertia Monitoring (2021): PowerPoint Presentation (naspi.org)
9. Zhang, Y., Tang, Q., Zhang, Y., Wang, J., Stimming, U., Lee, A. Identifying degradation patterns of lithium-ion batteries from impedance spectroscopy using machine learning. Nature communications 11(1): 1-6, 2020.
10. Publishable Summary - Quality assessment of electric vehicle Li-ion batteries for second use applications (17IND10): Details - EURAMET.
11. CCPI Europe. CCPI Europe signs license with UK National Physical Laboratory for self-validating thermocouples: CCPI Europe signs license agreement with NPL for INSEVA thermocouple (ccpi-europe.com)
12. Coquelin, L., Fischer, N., Feltin, N., Devoille, L., Felhi, G. Towards the use of deep generative models for the characterization in size of aggregated TiO2 nanoparticles measured by Scanning Electron Microscopy (SEM). Materials Research Express 6:8 (2019): 085001.
13. Crouzier, L., Delvallée, A., Ducourtieux, S., Devoille, L., Tromas, C., Feltin, N. A new method for measuring nanoparticle diameter from a set of SEM images using a remarkable point. UltraMicroscopy 207 (2019): 112847.
14. Mandija, S., Meliadò, E.F., Huttinga, N.R.F. et al. Opening a new window on MR-based Electrical Properties Tomography with deep learning. Sci Rep 9, 8895 (2019).
15. Ferrero, A., Iollo, A., Larocca, F. Field inversion for data-augmented RANS modelling in turbomachinery flows. Computers & Fluids 201 (2020): 104474.
16. Thomas, S., Race, A., Steven, R., Gilmore, I., Bunch, J. Dimensionality reduction of mass spectrometry imaging data using autoencoders. IEEE Symposium Series on Computational Intelligence (2016).
17. Delaine, F.; Lebental, B.; Rivano, H. In Situ Calibration Algorithms for Environmental Sensor Networks: A Review. IEEE Sens. J. 2019, 19, 5968–5978.
18. Jose M. Barcelo-Ordinas, Messaud Doudou, Jorge Garcia-Vidala, Nadjib Badache, Self-Calibration Methods for Uncontrolled Environments in Sensor Networks: A Reference Survey, arXiv:1905.11060v1 [cs.NI] 27 May 2019.

19. Esposito, E., De Vito, S., Salvato, M., Bright, V., Jones, R. L., and Popoola, O.: Dynamic neural network architectures for on field stochastic calibration of indicative low cost air quality sensing systems, Sensor. Actuat. B-Chem., 231, 701–713, 2016.

8.2 References for Artificial Intelligence and Machine Learning section

1. Independent High-Level Expert Group on Artificial Intelligence. Ethics guidelines for trustworthy AI. Brussels: European Commission; 2019: Ethics guidelines for trustworthy AI | Shaping Europe's digital future (europa.eu).
2. Bughin, J., Seong, J., Manyika, J., Hämäläinen, L., Windhagen, E., Hazan, E. Notes from the AI Frontier: Tackling Europe's Gap in Digital and AI. McKinsey & Company: New York, NY, USA. 2019.
3. Daugherty, P., Berthon, B. Winning with the Industrial Internet of Things: How to Accelerate the Journey to Productivity and Growth. Accenture, 2015.
4. ISO/IEC TR 24028: 2020 Information technology — Artificial intelligence — Overview of trustworthiness in artificial intelligence.
5. Standardization Roadmap AI (din.de)
6. BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP, OIML. Evaluation of measurement data – Guide to the expression of uncertainty in measurement. Joint Committee for Guides in Metrology, JCGM 100:2008.
7. BIPM. Guides in Metrology: publications - BIPM.
8. Donoho, D. 50 years of data science. Journal of Computational and Graphical Statistics 26(4):745-66, 2017.
9. Friedman, J., Hastie, T., Tibshirani, R. The elements of statistical learning. New York: Springer series in statistics; 2001.
10. Goodfellow, I., Bengio, Y., Courville, A. Deep learning. MIT press; 2016 Nov 10.
11. Gal, Y. Uncertainty in deep learning. University of Cambridge, 2016.
12. Kingma, D., Salimans, T., Welling, M. Variational dropout and the local reparameterization trick. Advances in neural information processing systems pp. 2575-2583, 2015.
13. Kendall, A., Gal Y. What uncertainties do we need in Bayesian deep learning for computer vision? Advances in neural information processing systems pp. 5574-5584, 2017.
14. Girard, A., Rasmussen, C., Quiñonero-Candela, J., Murray-Smith, R. Gaussian Process Priors with Uncertain Inputs: Application to Multiple-Step Ahead Time Series Forecasting. Advances in Neural Information Processing Systems 15, 2003.
15. Loquercio, A., Segu, M., Scaramuzza, D. A general framework for uncertainty estimation in deep learning. IEEE Robotics and Automation Letters 5(2):3153–3160, 2020.
16. Martin, J., Elster, Aleatoric Uncertainty for Errors-in-Variables Models in Deep Regression, Neural Processing Letters, 2022.
17. Goodfellow, I., Shlens, J., Szegedy, C. Explaining and Harnessing Adversarial Examples. International Conference on Learning Representations, 2015.
18. Muller, R., Kornblith, S., Hinton, G. When does label smoothing help? Advances in neural Information processing systems pp 4694-4703, 2019.
19. Martin, J., Elster, C. Inspecting adversarial examples using the Fisher information. Neurocomputing 382: 80-86, 2020.

20. Martin, J., Elster, C. Detecting unusual input to neural networks. Applied Intelligence 51.4: 2198-2209, 2021.
21. Ribeiro, M., Singh, S., Guestrin, C. “Why should I trust you?”: Explaining the predictions of any classifier. ACM Conference on Knowledge Discovery and Data Mining (KDD), 2016.
22. Bach, S., Binder, A., Montavon, G., Klauschen, F., Müller, K.R., Samek, W. On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation. PloS one 10(7), 2015.
23. Melis, D., Jaakkola, T. Towards robust interpretability with self-explaining neural networks. Advances in Neural Information Processing Systems pp. 7775-7784, 2018.
24. Zhang, Q., Wu, Y., Zhu, C. Interpretable convolutional neural networks. IEEE Conference on Computer Vision and Pattern Recognition pp. 8827-8836, 2018.
25. Wu, M., Hughes, M., Parbhoo, S., Zazzi, M., Roth, V., Doshi-Velez, F. Beyond sparsity: Tree regularization of deep models for interpretability. arXiv 1711.06178, 2017.
26. Cuomo, S., Di Cola, V. S., Giampaolo, F., Rozza, G., Raissi, M., and Piccialli, F. Scientific Machine Learning through Physics-Informed Neural Networks: Where we are and what's next. *arXiv preprint arXiv:2201.05624, 2022*.
27. BAM - Projects - BAM Data Store - Research data management
28. The SI in Digital Communication - BIPM
29. Wilkinson, Mark D., et al. The FAIR Guiding Principles for scientific data management and stewardship, Scientific Data 3.1: 1-9, 2016.
30. EUR-Lex - 52021PC0206 - EN - EUR-Lex (europa.eu)
31. FCAI-Policy-Brief_Final_060122.pdf (brookings.edu)
32. L. Beining. “Vertrauenswürdige KI durch Standards?”, Stiftung Neue Verantwortung, 2020.
33. Fraunhofer IPA, “White Paper: Zuverlässige KI”, 2020.
34. Certification of processes for AI | LNE, Laboratoire national de métrologie et d'essais

8.3 References for Computational Modelling and Virtual Metrology section

1. Fuller, Z. Fan, C. Day and C. Barlow, “Digital Twin: Enabling Technologies, Challenges and Open Research,” IEEE. Access, vol. 8, pp. 108952-108971, 2020.
2. L. Wright and S. Davidson, “How to tell the difference between a model and a digital twin,” Adv. Model. And Simul. in Eng. Sci, vol. 7, p. 13, 2020.
3. A. Parrot and L. Warshaw, Industry 4.0 and the digital twin. Manufacturing meets its match, 2017. https://www2.deloitte.com/us/en/insights/focus/industry-4-0/digital-twin-technology-smart-factory.html.
4. Irikura, K.K.; Johnson, R.D., III; Kacker, R.N. “Uncertainty associated with virtual measurements from computational quantumchemistry models.” Metrologia 2004, 41, 369.
5. Trenk, M.; Franke, M.; Schwenke, H. “The ‘Virtual CMM’ a software tool for uncertainty evaluation—Practical application in an accredited calibration lab.” Proc. ASPE Uncertain. Anal. Meas. Des. 2004, 9, 68–75.
6. Liu, G.; Xu, Q.; Gao, F.; Guan, Q.; Fang, Q. “Analysis of key technologies for virtual instruments metrology”. In Fourth International Symposium on Precision Mechanical Measurements; International Society for Optics and Photonics: Bellingham,WA, USA, 2008; Volume 7130, p. 71305B.

7. Doytchinov, I.; Tonnellier, X.; Shore, P.; Nicquevert, B.; Modena, M.; Durand, H.M. "Application of probabilistic modelling for the uncertainty evaluation of alignment measurements of large accelerator magnets assemblies." Meas. Sci. Technol. 2018, 29, 054001.
8. Heißelmann, D.; Franke, M.; Rost, K.; Wendt, K.; Kistner, T.; Schwehn, C. "Determination of measurement uncertainty by Monte Carlo simulation," in Advanced Mathematical and Computational Tools in Metrology and Testing XI, 2018, pp. 192-202.
9. Sepahi-Boroujeni, S.; Mayer, J.; Khameneifar, F. Efficient uncertainty estimation of indirectly measured geometric errors of five-axis machine tools via Monte-Carlo validated GUM framework. Precis. Eng. 2021, 67, 160–171.
10. Wiegmann, A.; Stavridis, M.; Walzel, M.; Siewert, F.; Zeschke, T.; Schulz, M.; Elster, C. "Accuracy evaluation for sub-aperture interferometry measurements of a synchrotron mirror using virtual experiments." Precis. Eng. 2011, 35, 183–190.
11. Balzani, D.; Brands, D.; Schröder, J.; Carstensen, C. "Sensitivity analysis of statistical measures for the reconstruction of microstructures based on the minimization of generalized least-square functionals." Tech. Mech. 2010, 30, 297–315.
12. Fortmeier, I.; Stavridis, M.; Wiegmann, A.; Schulz, M.; Baer, G.; Pruss, C.; Osten, W.; Elster, C. "Sensitivity analysis of tilted-wave interferometer asphere measurements using virtual experiments." In Modeling Aspects in Optical Metrology IV; International Society for Optics and Photonics: Bellingham, WA, USA, 2013; Volume 8789, p. 878907.
13. Trapet, E.; Wäldele, F. "The Virtual CMM Concept." In Advanced Mathematical Tools in Metrology; World Scientific: Singapore, 1996; Volume 2, pp. 238–247.
14. Balsamo, A.; Di Ciommo, M.; Mugno, R.; Rebaglia, B.; Ricci, E.; Grella, R." Evaluation of CMM uncertainty through Monte Carlo simulations." CIRP Ann. 1999, 48, 425–428.
15. F. Aggogeri et al., "Measurement uncertainty assessment of Coordinate Measuring Machines by simulation and planned experimentation," CIRP Journal of Manufacturing Science and Technology, vol. 4, pp. 51-56, 2011.
16. BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP, and OIML, Evaluation of measurement data — Guide to the expression of uncertainty in measurement, Joint Committee for Guides in Metrology, JCGM 100 (2008).
17. BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP, and OIML, Evaluation of measurement data — Supplement 1 to the Guide to the expression of uncertainty in measurement: Propagation of distributions using a Monte Carlo method, Joint Committee for Guides in Metrology JCGM 101 (2008).
18. Joint Committee for Guides in Metrology, Evaluation of measurement data - Supplement 2 to the "Guide to the expression of uncertainty in measurement" - Extension to any number of output quantities, Sèvres, France: BIPM, 2011.
19. G. Wübbeler et al., "GUM-compliant uncertainty evaluation using virtual experiments," in International Workshop on Metrology for Virtual Measuring Instruments and Digital Twins, Berlin, Germany, 2021.
20. A. Gąska et al., "Virtual CMM-based model for uncertainty estimation of coordinate measurements performed in industrial conditions," Measurement, vol. 98, pp. 361-371, 2017.
21. B. H. Thacker et al., "Concepts of Model Verification and Validation," Los Alamos National Laboratory, 2004. https://www.osti.gov/servlets/purl/835920-Yl8Ow3/native/
22. F. Chinesta et al., "Virtual, Digital and Hybrid Twins: A New Paradigm in Data-Based Engineering and Engineered Data," Archives of Computational Methods in Engineering, vol. 27, p. 105–134, 2020.

23. EMRP Project NEW04 “Novel mathematical and statistical approaches to uncertainty evaluation”, https://www.ptb.de/emrp/new04-home.html (see best practice guide on Bayesian regression, on computationally expensive systems, and on conformity assessment and decision making).
24. H. N. Najm, “Uncertainty quantification and polynomial chaos techniques in computational fluid dynamics”, Ann. Rev. Fluid. Mech. 41: 35-52 (2009).
25. J. P. C. Kleijnen, “Regression and Kriging metamodels with their experimental designs in simulation: A review”, European Journal of Operational Research 1: 1 – 16 (2017).
26. A Beskos, D Crisan, A Jasra, On the stability of sequential Monte Carlo methods in high dimensions, The Annals of Applied Probability 24 (4), 1396-1445 (2014).
27. J. N. Kutz, “Data-driven modeling & scientific computation: methods for complex systems and big data”, Oxford University Press (0xford, 2013).
28. https://www.marketsandmarkets.com/PressReleases/digital-twin.asp
29. SEMI, SEMI E133 – 1014- SEMI, Standard specification for automated process control system ingerface, Milpitas, CA (Semiconductor Euqipment and Materials).
30. Orji, N. , Obeng, Y. , Beitia, C. , Mashiro, S. and Moyne, J. (2018), Virtual Metrology White Paper - INTERNATIONAL ROADMAP FOR DEVICES AND SYSTEMS(IRDS), IEEE-INTERNATIONAL ROADMAP FOR DEVICES AND SYSTEMS (IRDS), https://tsapps.nist.gov/publication/get_pdf.cfm?pub_id=924090 (Accessed June 24, 2022).
31. P.-A. Dreyfus, F. Psarommatis, G. May and D. Kiritsis, Virtual metrology as an approach for product quality in Industry 4.0: A systematic review an integrative conceptual framework, Int. J. Production Research 60 (2), 742, Taylor & Francis (2022).
32. M. Grieves, Digital twin: Manufacturing excellence through virtual factory replication, White paper, 10.5281/zenodo.1493930
33. M. Krafczyk, A. Shi, A. Bhaskar, D. Marinov and V. Stodden, Learning from reproducing computational results: three reproducibility principles and the reproduction package. Phil. Trans. R. Soc. A 378 (2021) 20200069.
34. P. V. Coveney, D. Groen and A. G. Hoekstra, Reliability and reproducibility in computational science: implementing validation, verification and uncertainty quantification, Phil. Trans. R. Soc. A 379 (2022) 20200409.
35. “Healthcare resource statistics - technical resources and medical technology.” https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Healthcare_resource_statistics_-_technical_resources_and_medical_technology (accessed Sep. 30, 2021).
36. A. Arduino, „EPTlib: An Open-Source Extensible Collection of Electric Properties Tomography Techniques“, Appl. Sci. 11, 3237, 2021.
37. Hsu, S. and Terry, F., “Spectroscopic ellipsometry and reflectometry from gratings (scatterometry) for critical dimension measurement and in situ, real-time process monitoring,” Thin Solid Films 455, 828–836 (2004).
38. Mack, C., [Fundamental principles of optical lithography: the science of microfabrication], John Wiley & Sons (nov 2008).
39. Scholze, F., Soltwisch, V., Dai, G., Henn, M.-A., and Gross, H., “Comparison of CD measurements of an EUV photomask by EUV scatterometry and CD-AFM,” in [Photomask Technology 2013], 8880, 88800O, International Society for Optics and Photonics (2013).

40. Henn, M.-A., Gross, H., Heidenreich, S., Scholze, F., Elster, C., and Bär, M., “Improved reconstruction of critical dimensions in extreme ultraviolet scatterometry by modeling systematic errors,” Measurement Science and Technology 25(4), 044003 (2014).
41. Germer, T. A., Patrick, H. J., Silver, R. M., and Bunday, B., “Developing an uncertainty analysis for optical scatterometry,” in Metrology, Inspection, and Process Control for Microlithography XXIII], 7272, 72720T, International Society for Optics and Photonics (2009).
42. O. El Gawhary, N. Kumar, S. Pereira, W. Coene, and H. Urbach. “Performance analysis of coherent optical scatterometry”. In: Applied Physics B 105.4 (2011)
43. M.-A. Henn, H. Gross, F. Scholze, M. Wurm, C. Elster, and M. Bär. “A maximum likelihood approach to the inverse problem of scatterometry”. In: Opt. Express 20.12 (2012).
44. M. Hammerschmidt, M. Weiser, X. G. Santiago, L. Zschiedrich, B. Bodermann, and S. Burger. “Quantifying parameter uncertainties in optical scatterometry using Bayesian inversion”. In: Modeling Aspects in Optical Metrology VI. Edited by B. Bodermann, K. Frenner, and R. M. Silver. Volume 10330. International Society for Optics and Photonics. SPIE, 2017, pages 8–17 (cited on pages 86, 87, 91, 102–104, 106).
45. S. Heidenreich, H. Gross, M. Wurm, B. Bodermann, and M. Bär. “The statistical inverse problem of scatterometry: Bayesian inference and the effect of different priors”. In: Modeling Aspects in Optical Metrology V. Volume 9526. International Society for Optics and Photonics. 2015, 95260U (cited on pages 86, 97).
46. R. Ghanem and P.-T. Spanos. “Polynomial Chaos in Stochastic Finite Elements”. In: Journal of Applied Mechanics-transactions of The Asme - J APPL MECH 57 (1990), pages 197–202 (cited on page 86).
47. N. Wiener. “The Homogeneous Chaos”. In: Amer. J. Math. 60.4 (1938), pages 897– 936 (cited on pages 5, 86).
48. D. Xiu. “Fast Numerical Methods for Stochastic Computations: A Review”. In: Commun. Comput. Phys 5 (2009), pages 242–272. doi:10.1007/s11538-008-9369-3 (cited on page 86).
49. “Parameter estimation from laser flash experiment data.” Wright, L, Yang, X-S, Matthews, C, Chapman, L, Roberts, S. Computational Optimization and Applications in Engineering and Industry., 2011, Springer, 205-220.
50. “Surrogate accelerated Bayesian inversion for the determination of the thermal diffusivity of a material”, James A J Rynn, Simon L Cotter, Catherine E Powell and Louise Wright, Metrologia 56(1), 015018, 2019.
51. “Development of High Temperature Multi-Layer Laser Flash Artefacts”, Ateeb Farooqui, Roger Morrell, Jiyu Wu, Louise Wright, Bruno Hay, Michael Pekris, Mark J. Whiting & Theo Saunders. Int J Thermophys 43, 13 (2022). https://doi.org/10.1007/s10765-021-02928-4.
52. G. Perrin, C. Durantin, Taking into account input uncertainties in the Bayesian calibration of time-consuming simulators, Journal de la Société Française de Statistique 160 (2019) 24–46.
53. J. Wang, N. Zabaras, Hierarchical bayesian models for inverse problems in heat conduction, Inverse Problems 21 (1) (2004) 183–206.
54. A. François, L. Ibos, V. Feuillet, J. Meulemans, Estimation of the thermal resistance of a building wall with inverse techniques based on rapid active in situ measurements and white-box or ARX blackbox models, Energy & Buildings 226 (2020).

55. M. Iglesias, Z. Sawlan, M. Scavino, R. Tempone, C. Wood, Bayesian inferences of the thermal properties of a wall using temperature and heat flux measurements, International Journal of Heat and Mass Transfer 116 (2018) 417–431.
56. S. Thébault, R. Bouchié, Refinement of the isabele method regarding uncertainty quantification and thermal dynamics modelling, Energy and Buildings 178 (2018) 182–205.
57. V. Gori, C.A. Elwell, Estimation of thermophysical properties from in-situ measurements in all seasons: Quantifying and reducing errors using dynamic grey-box methods, Energy & Buildings 167 (2018) 290–300.
58. A. Rodler, S. Guernouti, M. Musy, Bayesian inference method for in situ thermal conductivity and heat capacity identification: Comparison to ISO standard, Construction and Building Materials 196 (2019) 574–593.
59. S. Demeyer et al, Bayesian uncertainty analysis of inversion models applied to the inference of thermal properties of walls Metrologia, 58, 014001, 2021.
60. R. Chakir, Y. Maday et al, A non-intrusive reduced basis approach for parametrized heat transfer problems, J. of Comput. Physics, 376, 617-633, 2019.
61. A. Quarteroni et al, Reduced Basis Methods for Partial Differential Equations, Springer book, 2016.
62. M.C. Kennedy et al, Bayesian calibration of computer models, Journal of the royal statistical society, 425-464, 63, 2001.
63. R. Stroh, S. Demeyer et al, Assessing fire safety using complex numerical models with a Bayesian multi-fidelity approach, Fire Safety Journal, 91, 1016-1025, 2017.
64. D. Lebel et al, Statistical inverse identification for nonlinear train dynamics using a surrogate model in a Bayesian framework, J. of Sound & Vibration, 158-176, 458, 2019.
65. G. Perrin, Adaptive calibration of a computer code with time-series output, Reliability Engineering & System Safety, 196, 2020.
66. C. Zheng et al., Advances in Magnetics - Magnetoresistive Sensor Development Roadmap (Non-Recording Applications), IEEE Trans. Mag. 55, 0800130 (2019).
67. A. Manzin, Modeling of Nanostructured Magnetic Field Sensors, Compendium on Electromagnetic Analysis, pp. 181-210 (2020).
68. A. Manzin, V. Nabaei, O. Kazakova, Modelling and optimization of submicron Hall sensors for the detection of superparamagnetic beads, J. Appl. Phys. 111, 07E513 (2012).
69. F. F. Wu, “Power system state estimation: A survey,” Int. J. Electr. Power Energy Syst., vol. 12, no. 2, pp. 80–87, 1990.
70. Hesam Ahmadian Behrooz, R. Bozorgmehry Boozarjomehry, Modeling and state estimation for gas transmission networks, Journal of Natural Gas Science and Engineering,Volume 22, 2015, Pages 551-570,ISSN 1875-5100, https://doi.org/10.1016/j.jngse.2015.01.002.
71. Y. Huang, S. Werner, J. Huang, N. Kashyap and V. Gupta, "State Estimation in Electric Power Grids: Meeting New Challenges Presented by the Requirements of the Future Grid," in IEEE Signal Processing Magazine, vol. 29, no. 5, pp. 33-43, Sept. 2012, doi: 10.1109/MSP.2012.2187037.
72. Zihang Zhang, Isam Saedi, Sleiman Mhanna, Kai Wu, Pierluigi Mancarella, Modelling of gas network transient flows with multiple hydrogen injections and gas composition tracking, International Journal of Hydrogen Energy, Volume 47, Issue 4, 2022, Pages 2220-2233, ISSN 0360-3199, https://doi.org/10.1016/j.ijhydene.2021.10.165.

73. Poudyal, Abodh et al. “Convolutional Neural Network-based Inertia Estimation using Local Frequency Measurements.” 2020 52nd North American Power Symposium (NAPS) (2021): 1-6.
74. Nicholas J. Higham (1996). Accuracy and Stability of Numerical Algorithms. Philadelphia: Society of Industrial and Applied Mathematics. ISBN 0-89871-355-2.
75. A. Tarantola, Inverse Problem Theory, Elsevier Amsterdam etc., 1987.
76. N. Metropolis, A. W. Rosenbluth, A. H. Teller, and E. Teller, “Equation of state calculations by fast computing machines.” J. Chem. Phys. 21:1087-1092,1953.
77. W.K. Hastings, “Monte Carlo sampling methods using Markov chains and their applications”, Biometrika 57:97-109, 1970.
78. A. Carrassi, M. Bocquet, L. Bertio, and G. Eversen, *“Data assimilation in the geoscience: An overview of methods, issues and perspectives. “* Wiley Interdisciplinary Reviews: Climate Change 9, (5), e535, 2018.
79. W. A. Lahoz and P. Schneider, “Data assimilation: making sense of Earth Obervation“, Frontiers in Environmental Science 2, 16, 2014.
80. P. L. Houtekamer and F. Zhang, ”Review of the ensample Kalman filter for atmospheric data assimilation.” Monthly Weather Review 144 (12), 4489, 2016.
81. M. B. Giles, “Multilevel monte carlo methods”, Acta numerica 14, 259, 2015.
82. O. Cappe, S. J. Godsill, and E. Moulines, “An overview of existing methods and recent advances in sequential Monte Carlo”, Proceedings of IEEE 95 (5), 899, 2007.
83. K. Csilléry, M. GB. Blum, O. E. Gaggiotti, and O. Francois, “Approximate Bayesian computation (ABC) in practice”, Trends in ecology & evolution 25 (7), 410, 2010.
84. Rubin J, Berra L. Electrical impedance tomography in the adult intensive care unit: clinical applications and future directions. Current Opinion in Critical Care. 2022 Jun 1;28(3):292-301.
85. Cultrera A, et al., Mapping the conductivity of graphene with Electrical Resistance Tomography. Scientific Reports. 2019 Jul 23;9(1):1-9.
86. https://www.bipm.org/en/committees/jc/jcgm/publications.

8.4 References for Data Analysis and Uncertainty Evaluation

1. N. F. Zhang, “Calculating of the uncertainty of the mean of autocorrelated measurements”, Metrologia 43, 276, 2006.
2. N. F. Zhang, “Allan variance of time series models for measurement data”, Metrologie 45, 549, 2008.
3. EMPIR NEW04 JRP, “Best practice guide to uncertainty evaluation for computationally expensive models”, https://www.ptb.de/emrp/fileadmin/documents/nmasatue/NEW04/Papers/BPG-WP2-NEW04.pdf.
4. BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP, and OIML, “Evaluation of measurement data – The role of measurement uncertainty in conformity assessment ”, JCGM 106 (2012).
5. ISO 10576-1:2013: “Statistical methods – Guidelines for the evaluation of conformity with specified requirements – part1: General principles, (2002).
6. EMPIR NEW04 JRP, “A guide to Bayesian inference for regression problems”, https://www.ptb.de/emrp/fileadmin/documents/nmasatue/NEW04/Papers/BPGWP1.pdf

7. BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP, and OIML, “JCGM Publications: Guides in metrology”, https://www.bipm.org/en/committees/jc/jcgm/publications.
8. BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP, and OIML, GUM Newsletters, https://www.bipm.org/en/committees/jc/jcgm/wg/jcgm-wg1-gum/newsletters.
9. Bureau International des Poids et Mesures Mutual Recognition of National Measurement Standards, of Calibration and Measurement Certificates issued by National Metrology Institutes CIPM revision 2003 (http://www.bipm.org/utils/en/pdf/CIPM-MRA-2003.pdf).

9 APPENDICES

9.1 APPENDIX A: Survey of members of the EMN Mathmet about AI&ML

The following information was collected through a survey of the members of the EMN Mathmet. It comprises a selection of the questions included in the survey and a summary and aggregation of the responses presented in a way to avoid duplication.

Does your NMI have a strategy for AI/ML in metrology? If so, what are the main objectives?

- To ensure confidence in, and trustworthiness of, AI outputs
- To be aware of the main developments
- To look for opportunities for the metrological community
- To support applications in different fields, including health, climate, and environmental monitoring, IoT, smart grids, neuromorphic engineering, condensed matter
- To enable machine learning based data analysis
- To improve decision making
- To aid scientific discovery
- To apply AI tools in the laboratory
- To develop AI based services

What potential do you expect from the application of AI for metrology?

- Driving future inline metrology
- Improving data analysis
- Supporting new measurement modalities (e.g., in imaging and sensor networks) that are increasingly data-driven rather than model-driven
- Supporting the digital transformation of metrology and applications of metrology (e.g., in autonomous vehicles, environmental monitoring, etc.)
- Reducing post-measurement costs
- Helping to refine modelling
- Supporting data intensive applications involving complex or unknown physical models
- Replacing model-based simulation techniques with data-driven approaches
- Coupling with in silico models to realize digital twins able to mimic virtual experiments
- Providing insight into multiparameter modelling of large volumes of data coming from sensor networks (e.g., IoT)
- Handling of big data

- Accelerating development cycles
- Extracting knowledge from complex data
- Enabling efficient data analysis, and predictive maintenance of experiments
- Establishing new services to build trust in AI applications
- Benefits unclear

What challenges do you expect to face if AI is to be used for metrology?

- Confidence of AI results (robustness and reliability)
- Transparency of AI (explainability and interpretability)
- Uncertainty quantification (enabling traceability)
- Verification and validation of algorithms and software
- Computational resources
- Understanding data quality (and ensuring quality of training data)
- Maintainability of AI models
- Loss of deep understanding (compared to iterative physical model building and validation)
- Combining experimental and simulation data

What kind of input could metrology provide for AI or ML?

- Standardisation of AI
- Providing reference data to train AI
- Understanding impact of measurement uncertainty on robustness of AI systems
- Providing a principled framework for using AI (exploiting well-established concepts of traceability, uncertainty, and calibration, etc.)
- Formulating the requirements of AI and setting specific research goals
- Making AI more reliable and transparent
- Providing metrics for data quality
- Assessing confidence in AI and ML results and certifying applications
- Not clear yet: Measurement problems are likely to arise on a case-by-case basis. For example, facial recognition may rely on feature identification followed by a measurement step (e.g., distance between features). In practice, measurement and (even less) measurement uncertainty do not seem to be limiting AI development and one feature of AI is that it can be trained for tolerance to a degree of uncertainty (in the wider sense).

What do you think are important research priorities for AI in metrology?

- Uncertainty
- Interpretability and explainability (transparency)
- Robustness and reliability of AI
- Traceability of AI
- Trustworthiness of AI
- Secure AI

- Data quality
- Verification and validation of algorithms and software
- Finding relatively simple examples in areas (close to) classical metrology that demonstrate the power of these methods and the need for the metrological community to pay more attention to AI
- Good understanding of the fundamentals (allowing correct interpretation of results)
- Application to small data sets
- Regularisation
- Not clear yet

Extracts taken from 'INRIM Metrology towards 2030: Developing relevant measurement science and technology'

- "Our industry and society, however, is rapidly changing. The recent progress in neural networks has allowed for efficient implementation of machine learning algorithms which will change, for instance, the way we drive cars, perform medical diagnoses, control industrial processes, and take strategic decisions in general. The full impact that the current digital revolution will have on our lives in the near future is difficult to foresee."
- "Major changes are taking place in industry, including a pervasive digital transformation which aims to use large sensor networks with wireless connectivity for taking better decisions as part of the Industry 4.0 concept. This enhanced metrological capability goes together with automation and with the extremely rapid development of neural networks and machine learning algorithms. Hence metrology tools will have to be developed that can deal with large amounts of data and virtual entities, and that can provide trust in decisions taken by artificial intelligence. Information technology security has to be assured in this process."
- "Measurement technologies – which can be classified by their technology readiness levels (TRLs) – range from classical technologies (high TRL) to quantum (medium-low TRL), and further down to those technologies for chemical and digital metrology (typically low TRL) that mostly have to be developed."
- "Finally, during the next decade, we cannot underestimate the role of digital transformation. The use of novel digital technology leads to a highly interconnected economy, industry, and society. Measurement results, data, algorithms, mathematical and statistical procedures, as well as communication and security architectures form the basis of digital expansion and transformation. In this context, it is essential that metrology embraces this paradigm shift. In particular, among many others, there are two significant challenges, namely improving the underlying mathematical, computational, and statistical sciences used in testing and measuring, and developing standardization and calibration protocols for a digital metrological infrastructure."
- "Supporting this significant effort within the context of a growing National Calibration System requires the development of a digital metrological infrastructure – thereby taking maximum advantage of novel digital technologies in a world where the economy, industry, and society will be increasingly interconnected. In particular, this digital transformation requires developing standardization and calibration protocols which are specific for digital technologies."
- "Imaging and sensing on biological neurons and neural networks will provide inspiration for the activity on new hardware platforms for artificial neural networks discussed in the next section."
- "Nowadays industry is facing a new transition based on (large amounts of) data and partially autonomous machines – a concept often referred to as Industry 4.0."

- “Digital transformation is much more than collecting data: it is a way to redesign the industrial production (even dynamically) in order to match a wider set of requirements by the customers. INRiM will contribute on three levels: sensor technology, data sensor networks, and innovative neural networks.”
- “To process vast amounts of data and implement – partially automatic – decision making processes, neural networks and artificial intelligence will play an increasingly important role. The current bottleneck in this respect lies in the traditional hardware platforms (computers/electronics) which are inefficient for this task and extremely energy consuming compared to biological neural networks. INRiM will dedicate effort in developing new materials and technologies to find better hardware implementations of neural networks, which, if successful, will constitute one of the major breakthroughs in the field of large-scale data processing and machine decision making.”
- “In addition to the experimental capabilities and know-how, INRiM will also be able to provide sophisticated computational post-processing techniques for improved data interpretation which are typical for the metrology community. Examples include the use of digital twins for precision measurements and data analysis, uncertainty quantification and data comparability of measurement methods, suitable metrics for machine learning algorithms, and computational imaging for life sciences.”

Extracts taken from “Metrology and AI: PTB’s AI Strategy, March 2022”

- “It is not only in Industry 4.0 where significant resources can be saved through the predictive maintenance of machines and systems using AI. New fields of application for AI are also constantly opening up in the intelligent control of supply systems in smart homes and smart cities, in self-learning diagnostic tools for personalised medicine, and in autonomous vehicles.”
- “PTB thus sees it as its duty to conduct fundamental research on the data quality and reliability of AI procedures. Building on this research PTB seeks to advance the development of the legal framework for AI approval and regulation in cooperation with other QI stakeholders. In addition, PTB would also like to further the opportunities arising from the use of AI methods in the research and development environment and make them safely usable.”
- “These algorithms are characterised by a high complexity and a high-dimensional parameter space. Another feature is the very high adaptability of AI methods. However, this can lead to undesired features of the training data unintentionally being built into the algorithm. Therefore, in contrast to other software, a check of the algorithm based on the source code alone often no longer feasible.”
- “PTB sets itself the goal of developing suitable metrics for the evaluation of AI and data in its metrological research mission, adapting existing measurement and testing processes to the use of AI and, at the same time, testing and expanding the safe application of AI for metrological research and services.”
- “In data-based methods, the uncertainty is made up of three components … Uncertainty due to inherent limitation in the model fit of the learning system; Uncertainty due to data quality; Uncertainty due to divergent training, testing and application contexts.”
- “One of PTB’s goals in this area is ultimately to establish a standardised measure for the quantification of explainability.”
- “All sources regarding evaluation, certification and conformity assessment of AI applications or products with AI components mention the need for reference data as well as generally accepted criteria for data quality and data handling.”
- “Use Case: Metrology for Autonomous Driving – Trust in AI”

- “Use Case: Quality control for explainable AI in clinical diagnostics”
- “Use Case: AI for optical metrology – shape measurements and nanometrology”
- “Use Case: Soot particle characterisation with the help of AI”
- “PTB aims to establish well-organised and harmonised machine-usable data and AI methods as trust anchors for future technologies in metrology, to develop and provide digital standards (e. g., reference data sets) for metrology and to set up the necessary infrastructures.”
- “Since the previous approach of optimising the model or the code to achieve better performance of the AI method often no longer achieves drastic improvements in many Use Cases, a data-centric approach has recently garnered a lot of attention. With this method, the training data are specifically selected according to strict guidelines (e. g., consistent labelling, discarding noisy data, coordinated handling and labelling of data sets that are difficult to evaluate) and can thus improve the performance of AI systems significantly, according to initial pilot studies [49]. In this context, too, the influence of the quality of the data and fundamental data requirements is becoming clearer than ever.”

9.2 APPENDIX B: Survey of members of the Stakeholder Advisory Committee about AI&ML

The following table summarizes the needs and interests of the stakeholder advisory committee (SAC) members which were collected through questionnaires, interviews, and from SAC meetings.

ENBIS	• Virtual sensors process data originally gathered by physical sensors. The data they deliver is then typically embedded into more complex functions or software applications that merge this input with data from other sources and execute analytics algorithms on the combined set of data. In case these analytics algorithm embed ML, for example, a proper UQ method is needed.
Infineon	• Health Status of a device. • PHM (prognostics and health management). • Aleatoric and epistemic uncertainty. • Accuracy and uncertainty of AI output variables.
Eurolab	• How to validate AI model and results? Is it possible to create a validation procedure?
EMN COO	• How to propagate uncertainties in observations to products derived from those observations using neural networks. • How to validate models developed using ML methods when applied to new datasets. • Uncertainties in classification (e.g., land type maps, cloud masking) and how those become other uncertainties when used in quantitative processing. • Some scientists use neural networks when standard regression would work and be easier to propagate uncertainties through. • Inverse problems from a neural network derived forward model.
TNO	• Model robustness in AI/ML. • Contestability of AI-based decision-making systems. • Oversight bodies and standards: what and how to audit AI/ML? • Testing facilities and reference datasets to train AI on.
UKAS	• Defining the scope: what is actually being measured, what is inferred. • Traceable measurements for remote inspection. • Records - reproducibility conditions?
Eurachem	• Relatively few clear needs here for most analytical chemistry and biology laboratories; most measurements still essentially univariate with few regulatory measurements multivariate/ML based.
CFM	• Uncertainty for ML algorithms. • Validation understanding of AI. • Confidence in AI results. • Training of experts in uncertainty evaluation.

9.3 APPENDIX C: Needs from other European Metrology Networks

EMN Climate and Ocean Observation

Extracts taken from the Stakeholder Needs Review Report, European Metrology Network for Climate and Ocean Observation Version 1.0 (01/2021).

“Metrology challenges for remote sensing

Metrological assessment of uncertainty of models and algorithms particularly those required to transform top-of-atmosphere measurands to bottom-of-atmosphere parameters, including support to developers, guidance on assessment, such as the challenges of ‘machine learning’ methods. Means to establish uncertainty characterisation and representation for ‘classification’ systems e.g., land cover type, cloud masks etc. is also needed.”

“General cross-cutting metrological support for climate and ocean observation

Finally, machine-learning techniques, including neural networks, are increasingly being used in the generation of ECV products, particularly for more complex variables that rely on proxy measures or to associate satellite observations with an in-situ metric. Neural networks are also used in applications of ECVs beyond direct climate modelling, and particularly in “climate services”, for example to establish risk or to quantify “embedded carbon”.

Metrologists, and particularly data scientists working in metrology institutes, can support the analysis of uncertainty through modelling and the interplay between measurements, observations, and models. Metrological research is needed to understand how model uncertainty is evaluated and validated, to develop standardised ways of reporting results of model performance evaluation, and to provide methods for uncertainty propagation through non-linear models. Furthermore, error covariance information is very important, but with huge data sets (87 billion observations) new methods are needed to process such information in a computationally affordable manner, and methods such as machine learning will themselves need to be robustly evaluated to assess the uncertainties they might introduce into the analysis through their use.”

EMN Quantum Technologies

Extract taken from European Metrology Network on Quantum Technologies: Strategic Research Agenda (Draft 4 May 2022).

“Subfield: Quantum computing

Applications in chemistry, Fintech, Machine learning

The first real-world applications of NISQ processors are expected to be in chemistry but there are also promising signs that e.g., important optimisation problems in finance can be addressed with processors of moderate size. Machine learning is likely to be another important application area.”

EMN Smart Electricity Grids

Extracts taken from EMN Smart Electricity Grids Strategic Research Agenda Draft version 2.0 (03/2022).

“Some measurement challenges in data analytics

- Development of big data analytics and visualisation platforms
- Turning large data sets collected in distribution grids into actionable information
- Lossless data compression
- Provide measurement data to support the development of machine learning algorithms for short-term load forecasting
- …”

9.4 APPENDIX D: EMPIR projects related to CM&VM

There are a few EMPIR (European Metrology Programme for Innovation and Research) projects in which computer-aided modelling plays an important role. Some extracts from publishable summaries are shown below.

9.4.1 MIMAS: Procedures allowing medical implant manufacturers to demonstrate compliance with MRI safety regulations

Overview

Medical implants represent a multi-billion market across Europe. A majority of the 50 million EU citizens carrying a medical implant will likely need a magnetic resonance imaging (MRI) scan during the lifetime of their device. However, the powerful electromagnetic fields of MRI systems in these cases represented a unique hazard for patient safety. Therefore, it was vital for both patient wellbeing and the success of a medical implant on the market that implant manufacturers could demonstrate safety compliance for their device in an MRI environment. This project improved the competitiveness of European implant manufacturers by providing innovative, metrologically sound and legally safe methods to demonstrate the compatibility of their products with MRI safety regulations. New, high resolution anatomical models of implant carriers were developed, using virtual surgery techniques to position the device in the patient. A comparison to the less accurate, but simpler and cheaper state-of-the art techniques to create a computer of an implant patient is given. Researchers and implant manufacturers are, for the first time, to choose the proper approach to meet their specific requirements. A new Medical Device Development Tools was developed and regulatorily approved during the project. It is commercially available and provides implant manufacturers with a clear and legally safe pathway to obtain regulatory approval for their innovative devices. Beyond solutions for today, also completely new, and potentially disruptive approaches towards personalised implant safety assessments were investigated. On a proof-of-concept level it was demonstrated, how sensor-equipped implants interfaced to parallel-transmit capable MR scanners could not only combine patient safety with optimised image quality, but simultaneously make manufacturers more and clinical personnel less responsible for safety of an implant carrying patient. A complete assessment of the patient hazards due to gradient-induced heating of large implants in MRI was achieved and possible test procedures and simplified analyses were described. The ground is thus prepared for standardisation bodies to include this hitherto uncovered subject into their normative documents and for test laboratories to offer the relevant test procedures and equipment.

Objectives related to CM:

1. To develop anatomical models of human subjects with realistic medical implants and millimetre resolution. The models to be sufficiently detailed for use within silico medicine concepts, with resolution to be determined according to image analysis needs.
2. To develop validated computational tools for the numerical simulation of electromagnetic fields (EMF) and temperature distributions in a virtual human subject during MRI exposure. The computational tools should be able to process high-resolution anatomical models.
3. To investigate numerically and experimentally the hazards associated with the interaction between bulk metallic implants and switched magnetic fields in the kilohertz regime. In addition, to develop a reference set-up for testing metallic implant heating, using switched magnetic-field gradients of a few mT/m with a target gradient uncertainty below 5%.
4. To develop and apply a suitable statistical method to demonstrate MRI compliance for small (<10 cm) orthopaedic implants without extensive testing or numerical modelling, by determining an upper limit for the hazard associated with the new implant by comparison with a similar surrogate implant, which has already been fully

assessed, thus enabling small manufacturers of a large variety of similar small metallic implants to dramatically reduce their costs for compliance demonstration.

9.4.2 QUIERO: Quantitative MR-based imaging of physical biomarkers

Overview

With more than 30 million scans per year in European countries, Magnetic Resonance Imaging (MRI) is one of the most important tomographic tools adopted in clinical practice. Nevertheless, standard MRI results mostly have a qualitative nature (i.e. they display a contrast between different tissues, which must be interpreted by a specialist on visual inspection) that limits their objectivity and comparability. The project will evaluate the suitability of two MR-based emerging techniques, Electrical Properties Tomography (EPT) and Magnetic Resonance Fingerprinting (MRF), to bring a "quantitative revolution" in MRI, so that each image pixel is associated with the measurement (including uncertainty) of one or more tissue parameters.

Objectives related to CM

1. To develop, improve and implement numerical algorithms for use in EPT and MRF and to characterise their performance. For EPT, both local relationships and global inversion methods will be considered and compared; for MRF, statistical template-free methods will be evaluated as an alternative to traditional dictionary-based techniques.
2. To make EPT and MRF suitable for practical use in the analysis of "high impact" clinical conditions. Basic EPT techniques will be improved to handle the partial knowledge of the phase of the magnetic field and mainly applied to the analysis of diseases that cause significant changes in dielectric properties (e.g. cerebral ischemia). The application of MRF will be extended to the heart region through methods able to suppress artefacts caused by physiological motion and moving fluids.
3. To evaluate the accuracy of EPT and MRF procedures in magnetic resonance experiments under controlled conditions. Heterogeneous phantoms, composed of soft semisolid materials mimicking the properties of human tissues (e.g. conductivity, relative permittivity, longitudinal and transverse relaxation times in the order of 1 S/m, 50, 1000 ms and 50 ms respectively), will be specifically developed and used for this purpose. The target uncertainties required are 20 % for EPT and 10 % for MRF.
4. To fully characterise EPT and MRF as diagnostic tools under real-world conditions, including determining, for the target organs selected, the inter-and intrasubject physiological variability and minimum threshold for the detection of anomalies due to diseases. The variability of tissue properties will be taken into account and advanced statistical techniques and in vivo assessments will be applied. The synergistic use of EPT and MRF will be explored to optimise diagnosis and specific computer-aided diagnostics approaches will be developed.

9.4.3 MedalCare: Metrology of automated data analysis for cardiac arrhythmia management

Overview

The aim of the project is to develop a novel validation strategy of cardiac arrhythmia classification algorithms based on multiparametric data analysis of electrocardiography (ECG) data through metrological research. A novel synthetic reference database will be developed that will enable to investigate, for the first time, the uncertainty of modern data analysis approaches, such as machine learning in medicine and to contribute to standardising machine learning methods in health applications, specifically by establishing a novel metrological validation platform of such algorithms manifesting a digital traceability chain.

Objectives related to CM

1. To develop synthetic ECG (electrocardiography) reference data of a virtual population. This would involve existing biophysical modelling frameworks to develop a synthetic

ECG reference dataset allowing the assessment of uncertainty of automated data analysis methods such as machine learning (ML). An ECG-database of a representative virtual population including healthy variations and selected pathologies will be generated.

2. To carry out the uncertainty analysis of reference data by assessing the sensitivity of different parameters on results of the biophysical modelling resulting in an uncertainty evaluation of the synthetic ECG data. For this, the influence of the model input parameters, such as anatomies, conduction blocks, tissue conductivity, infarct and fibrotic tissues, will be assessed.
3. To assess and compare the effect of different classification approaches focusing on uncertainty analysis along two directions: the influence of uncertainty of features of ECG data on the output of the classification algorithm and the influence of wrongly labelled training data on the output of the classification. The project would investigate whether hidden features can be detected by modern ML-approaches for "quantitative classification" of ECG.
4. To carry out thorough investigation of clinical application of multi-parametric data analysis that includes detection and classification of cardiac ischemia and arrhythmia. A comparison of performance of experienced physicians.

9.4.4 RaCHy: Radiotherapy Coupled with Hyperthermia

Overview

The integration of radiotherapy with hyperthermia requires experimental studies to accurately assess the biological mechanisms involved at a cellular level (e.g., the inhibition of DNA repair mechanisms caused by heat exposure). The increased understanding of the involved biological mechanisms will allow clinicians to prescribe the required thermal and radiation doses (magnitude and homogeneity), according to the individual patient's needs. The project is aimed at providing metrological support to achieve the maximum synergistic advantages in the integration of radiotherapy (RT) oncology with different hyperthermia (HT) techniques, based on high intensity Therapeutic Ultrasound (TUS), Electromagnetic Radiation (EMR) and the use of magnetic nanoparticles (MNPs) excited by AC magnetic fields. One of the main aims of this project is to experimentally demonstrate how the use of different sources to generate hyperthermia, combined with radiotherapy, could result in a successful treatment of the whole tumour. This approach requires excellent knowledge and control of the temporal and spatial distributions of temperature increases, and of the radiation dose during and after the treatments.

Objectives related to CM

1. To develop heat delivery systems for hyperthermia treatments (TUS, EMR and MNPs) for use with radiotherapy. 2D and 3D measurement set-ups and validated modelling tools will be developed to estimate the spatial-temporal distribution of energy deposition.
2. To develop innovative analytical tools for biological assessment by using chemical metrology multimodal techniques as suitable non-invasive and non-ionising tissue diagnosis tools, and mass spectrometry combined with imaging modalities at nanometre resolution.
3. To facilitate the review of Biological Equivalent Dose (BED) concept related to the radiotherapy combined with hyperthermia. The role of control parameters such as the energy deposition in tissues, the radiation dose and the duration of the hyperthermia and/or radiation treatment will be taken into account.

9.4.5 ATMOC: Traceable metrology of soft X-ray to IR optical constants and nanofilms for advanced manufacturing

Overview

The optics and semiconductor industries have been using innovative materials and complex nanostructures whose optical properties are difficult to measure and often not accurately known. This project will develop advanced mathematical methods to traceably characterise these materials for wavelength ranges, from soft Xray to IR. This will be achieved by creating a database of optical constants with associated uncertainties for bulk materials and ultra-thin film systems and industrially relevant datasets. This database will provide the opportunity to relevant industries to run simulations and eventually develop new materials with tailored properties.

Objectives related to CM

1. Advanced inverse modelling and virtual measurements. To develop and apply advanced mathematical models for virtual and real measurements in order to determine the optical response of the test samples developed in objective 1 and their dependence on complex nanostructures. The uncertainties associated with ab initio methods, interlayer roughness, crystal structures, model reduction techniques, surrogate modelling, machine learning and inverse modelling should also be determined.
2. Determination of optical response functions and assembling a database. To determine the optical constants and the corresponding measurement uncertainties of thin stratified layer systems and to estimate the geometrical parameters of these complex nanostructures in the soft Xray to IR spectral range. In addition, to assemble a database of optical constants, dielectric tensors and estimated geometrical parameters, including both measurement values and virtual/simulated measurement data.

EURAMET e.V.
Bundesallee 100
38116 Braunschweig
Germany

Phone: +49 531 592 1960
Fax: +49 531 592 1969
E-mail: secretariat@euramet.org

The EMPIR initiative is co-funded by the European Union's Horizon 2020 research and innovation programme and the EMPIR Participating States