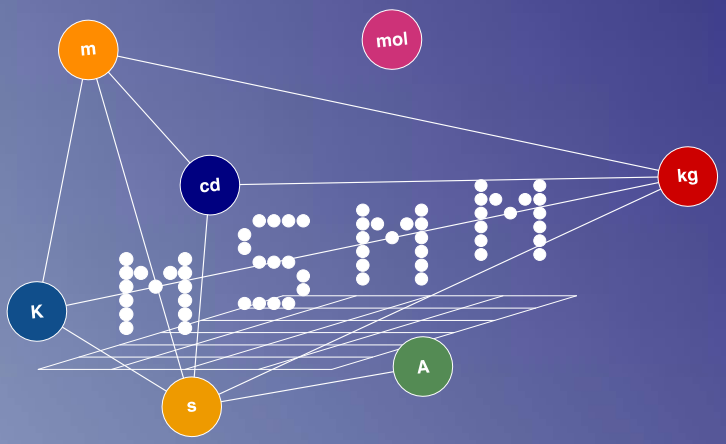


Joint Virtual Workshop of ENBIS and MATHMET

Mathematical and Statistical Methods for Metrology MSMM 2021

31 May – 1 June 2021

http://www.msmm2021.polito.it/

Programme Booklet

enbis

MATHMET

INRiM ISTITUTO NAZIONALE DI RICERCA METROLOGICA

Politecnico di Torino

DISMA Dipartimento di Scienze Matematiche G. L. Lagrange ECCELLENZA 2018 - 2022

SHOW CROW

AIM OF THE WORKSHOP

Recognizing the increasing need of *ad hoc* and innovative mathematical and statistical tools for current and emerging metrological applications in several areas of the Science of Measurement, in 2019 the Italian local network of ENBIS, i.e. **itENBIS**, the Italian National Metrology Institute **INRIM** and **Politecnico di Torino** organized a joint Workshop on Mathematical and Statistical Methods for Metrology (**MSMM 2019**).

Based on the success of the previous edition and in view of the foreseen collaborations between statistical and metrological European communities, the second edition of the **MSMM 2021 Workshop** is jointly organized by the **European Network for Business and Industrial Statistics** (**ENBIS**) and the **European Metrology Network (EMN) for Mathematics and Statistics** (**MATHMET**), under the local organization of INRIM and Politecnico di Torino.

Highlights of the Workshop include invited talks, as well as plenary, specialized and contributed sessions concerning the main topics of Mathematics and Statistics for metrological applications. Session topics include, but are not limited to:

- Uncertainty and measurement quality evaluation
- Regression and inverse models
- nterlaboratory data evaluation
- High dimensional, dynamic and complex models
- Bayesian models
- Simulations and virtual experiments, computational methods and digital twins
- Artificial Intelligence, Machine Learning, Big Data analytics
- Statistical engineering
- Data and software reliability
- Design of experiments, sampling and sequential design
- Digitalization in Metrology
- Time series analysis
- Conformity assessment, reliability and quality control
- Chemometrics
- Biostatistics

A joint **introductory session** is specifically dedicated to the presentation of the mission and main recent activities conducted by **ENBIS** and **MATHMET**, respectively, in order to foster collaborative interactions between the two bodies.

The official language of the Workshop is English.

A VIRTUAL WORKSHOP

Due to the world-wide infection of COVID-19, MSMM 2021 is held as online Workshop hosted in a virtual conference centre.

Virtual presentations are offered in either of the following modes:

- **Oral presentations**: scheduled within the specialized and contributed sessions, oral presentations are held live (synchronous presentations). They will last 15 minutes each plus 5 minutes for Q&A. No pre-recorded video of the oral presentations is required.
- **Poster presentations:** uploaded within the poster room, poster presentations are provided in the form of a 5 minute pre-recorded video (asynchronous presentations). A dedicated chat service is integrated into each poster in order to allow Q&A. Note that the poster session is open and accessible for the whole duration of the Workshop.

The virtual conference centre site will be open throughout the event.

We encourage you to visit the poster room and to meet and chat with other participants in the virtual coffee corner.

Please leave your comments and suggestions in the Information Desk.

We are grateful to ShowCrow (https://www.showcrow.de) who enabled us to exploit this exciting virtual environment to host the Workshop.

CO-CHAIRS

Francesca Pennecchi (INRIM)
Grazia Vicario (Politecnico di Torino)

SCIENTIFIC COMMITTEE

João Alves e Sousa (IPQ)
Markus Baer (PTB)
Rossella Berni (Università di Firenze)
Walter Bich (INRIM)
Alen Bosnjakovic (IMBIH)
Oriano Bottauscio (INRIM)
Maurice Cox (NPL)
Severine Demeyer (LNE)
Stephen L. R. Ellison (LGC)
Clemens Elster (PTB)
Nicolas Fischer (LNE)
Bernard Francq (GSK)
Alistair Forbes (NPL)
Fiorenzo Franceschini (Politecnico di Torino)
Maurizio Galetto (Politecnico di Torino)
Rainer Göb (University of Wuerzburg)
Sebastian Heidenreich (PTB)
Katy Klauenberg (PTB)
Gertjan Kok (VSL)
Alessandra Manzin (INRIM)
Francesca Pennecchi (INRIM)
Antonio Pievatolo (IMATI – CNR)
Jacek Puchalski (GUM)
Amalia Vanacore (Università di Napoli)
Adriaan van der Veen (VSL)
Grazia Vicario (Politecnico di Torino)
Luca Zilberti (INRIM)

LOCAL ORGANIZING COMMITTEE

Antonella Bianchi (Politecnico di Torino)
Graziano Coppa (INRIM)
Eleonora Crevacore (Politecnico di Torino)
Elisabetta Melli (INRIM)

Contact

For any requirement, please contact msmm2021@polito.it

THE PROGRAMME

(with hyperlink to the abstracts)

Oral Presentations – all times listed are CEST (UTC+2)

Day 1 (31 May)

9:00-9:20 — **Welcome**

Francesca Pennecchi (MSMM 2021 co-chair, Istituto Nazionale di Ricerca Metrologica, IT)
Grazia Vicario (MSMM 2021 co-chair, Politecnico di Torino, IT)
Pietro Asinari (Scientific Director, Istituto Nazionale di Ricerca Metrologica, IT)

9:20-10:20 — **Joint ENBIS/MATHMET session**

(Chairs: Francesca Pennecchi and Grazia Vicario)

Murat Caner Testik (ENBIS President) and **Antonio Pievatolo** (ENBIS Former President)
Markus Bär (MATHMET Chair), **Sebastian Heidenreich** (MATHMET Secretary), **Galina Kulikova** (EMN-Manager) and **Katy Klauenberg** (Measuring Uncertainty Training)

10:25-11:25

	Progress on the GUM framework - Update from the JCGM WG1 **(Chairs: Walter Bich and Maurice Cox)**	**Machine Learning for Metrology I** **(Chair: Sebastian Heidenreich)**
10:25-10:45	ID 131 *GUM Part 6 – Developing and using measurement models. An outline* **Walter Bich** (Istituto Nazionale di Ricerca Metrologica, IT)	ID 106 *Uncertainty evaluation for machine learning: metrology requirements and open challenges* **Andrew Thompson** (National Physical Laboratory, UK)
10:45-11:05	ID 109 *Simple informative prior distributions for type A uncertainty evaluation with small samples* **Maurice Cox** (National Physical Laboratory, UK)	ID 71 *Deep Learning for inverse problems – applying ensemble learning for uncertainty quantification* **Lara Hoffmann** (Physikalisch-Technische Bundesanstalt, DE)
11:05-11:25	ID 130 *Transferability of GUM-S1 type A uncertainties - a Bayesian perspective* **Gerd Wübbeler** (Physikalisch-Technische Bundesanstalt, DE)	ID 187 *Deep Learning based instance segmentation: application to agglomerated titanium dioxide particles measured by scanning electron* **Paul Monchot** (Laboratoire national de métrologie et d'essais, FR)

11:25-11:40 — **15 min break**

11:40-12:40

	Sensor calibration **(Chair: João Alves e Sousa)**	**Inverse problems in metrology** **(Chair: Sebastian Heidenreich)**
11:40-12:00	ID 48 *Metrological redundancy in distributed measurements* **Gertjan Kok** (Van Swinden Laboratorium, NL)	ID 160 *Inversion of point clouds for holistic Screw Thread Metrology* **Anita Przyklenk** (Physikalisch-Technische Bundesanstalt, DE)
12:00-12:20	ID 138 *Co-calibration of sensor networks* **Alistair Forbes** (National Physical Laboratory, UK)	ID 132 *Model error in Bayesian inversion* **Maren Casfor Zapata** (Physikalisch-Technische Bundesanstalt, DE)
12:20-12:40	ID 129 *A novel method for Callendar-Van Dusen interpolation of temperature calibration points* **Graziano Coppa** (Istituto Nazionale di Ricerca Metrologica, IT)	ID 136 *Invertible neural networks for grazing incidence X-ray fluorescence parameter reconstruction* **Nando Farchmin** (Physikalisch-Technische Bundesanstalt, DE)

12:40-13:30 — **50 min lunch**

13:30-14:30	***INVITED LECTURE*** ***Classical and Bayesian optimization for efficient experimental designs in metrology*** **Blaza Toman (National Institute of Standards and Technology, US)** **(Chair: Francesca Pennecchi)**		
14:35-15:35	**Uncertainty I** **(Chair: Walter Bich)**	**Designs of measurement experiments** **(Chair: Grazia Vicario)**	**Metrology in chemistry and chemometrics** **(Chair: Stephen Ellison)**
14:35-14:55	ID 180 *A knowledge-based evaluation of measurement non-repeatability* **Carlo Carobbi** (Università degli studi di Firenze, IT)	ID 64 *Optimal designs for hypothesis testing with heteroscedastic experimental groups* **Marco Novelli** (Università di Bologna, IT)	ID 162 *In the avantgarde of a reliable methodology for automatic identification of microplastics by micro-ATR-FTIR spectroscopy* **Vanessa Morgado** (Universidade de Lisboa, PT)
14:55-15:15	ID 176 *Uncertainty of thermodynamic properties available via online data banks: Vapor pressure as case study* **Maricarmen Lecuna** (Politecnico di Torino, IT)	ID 70 *Evaluating erosion performance of cold-sprayed coatings by Design of Experiments* **Elisa Verna** (Politecnico di Torino, IT)	ID 177 *Monte Carlo bottom-up evaluation of the uncertainty of complex sample preparation: Elemental determination in sediments* **Ricardo Bettencourt da Silva** (Universidade de Lisboa, PT)
15:15-15:35	ID 80 *Uncertainty expression by finite information quantities* **Luca Callegaro** (Istituto Nazionale di Ricerca Metrologica, IT)	ID 73 *A GUI for Bayesian sample size determination* **Jörg Martin** (Physikalisch-Technische Bundesanstalt, DE)	ID 145 *Discriminant analysis of vegetable oils by TGA-GC/MS combined with chemometrics and data fusion without sample pretreatment* **Xia Zhou** (National Institute of Metrology, Beijing, CHN)
15:40-16:40	**Uncertainty II** **(Chair: Alistair Forbes)**	**Human exposure to electromagnetic fields and ionizing radiations** **(Chair: Oriano Bottauscio)**	**Measurements on nominal and ordinal scales** **(Chair: Amalia Vanacore)**
15:40-16:00	ID 133 *Approximating Gaussian Process regression models using banded matrices* **Kavya Jagan** (National Physical Laboratory, UK)	ID 124 *Factors relating to gradient coil and radiofrequency induced heating within implanted orthopaedic devices during MRI* **Jenny Wooldridge** (National Physical Laboratory, UK)	ID 55 *Interlaboratory comparison of nominal data on macroscopic examination of welds* **Tamar Gadrich** (ORT Braude College, Karmiel, ISR)
16:00-16:20	ID 68 *How to improve linear interpolation uncertainty of humidity profiles* **Pietro Colombo** (Università degli studi di Bergamo, IT)	ID 161 *Identification of main factors impacting human exposure in inductive power transfer systems* **Lionel Pichon** (CentraleSupélec - Université Paris-Saclay, Sorbonne Université, FR)	ID 113 *Entropy-based explanations of multidimensionality in ordinal responses* **Leslie Pendrill** (RI.SE Research Institute of Sweden, SWE)
16:20-16:40	ID 81 *Uncertainty estimation by bootstrap sampling of area shape function in nano-indentation testing* **Giacomo Maculotti** (Politecnico di Torino, IT)	ID 98 *Radiation dose estimation via the contaminated Poisson and negative binomial methods in partial-body exposures* **Adam Errington** (Durham University, UK)	ID 168 *Simultaneous inference for comparing classifier performance via kappa-type coefficients* **Amalia Vanacore** (Università di Napoli "Federico II", IT)

16:45-16:50 **Conclusion of day 1**

16:50-17:15 **Scientific coffee time with MATHMET: open discussion, questions and answers on the research topics of the MATHMET Strategic Research Agenda**
(Chair: Sebastian Heidenreich)

This is an informal session organized by MATHMET Members to offer
(a) an introduction to the MATHMET Strategic Research Agenda (SRA),
(b) descriptions of the main research topics,
(c) an open discussion with the aim to collect input and feedback from stakeholders and end-users

Day 2 (1 June)

9:00-9:10 **Welcome to day 2**

9:10-10:10 ***INVITED LECTURE***
Hybrid Twins for empowering performance-based engineering based on advanced real-time physics, informed AI and smart-metrology
Francisco Chinesta (École Nationale Supérieure d'Arts et Métiers ParisTech, FR)
(Chair: Grazia Vicario)

10:15-11:15

Digital twins and virtual experiments (Chair: Alessandra Manzin)	**Flow simulation in metrology** (Chair: Sonja Schmelter)
10:15-10:35 ID 59 *"Biodigital Twins": optimizing orthopaedic implants* **Michael Gasik** (Aalto University Foundation, FIN)	ID 123 *Simulation of temperature measurement of inhomogeneous flows by ultrasonic flow meters* **Gertjan Kok** (Van Swinden Laboratorium, NL)
10:35-10:55 ID 170 *In silico experiments to guide magnetic hyperthermia pre-clinical tests* **Marta Vicentini** (Politecnico di Torino, IT)	ID 163 *Enhancement of multiphase flow simulations by turbulence damping at the gas-liquid interface* **Jiri Polansky** (Czech Technical University, Prague, CZE)
10:55-11:15 ID 191 *Virtual sensors development for real-time quality assessment in continuous production* **Manolo Venturin** (EnginSoft SpA, IT)	ID 178 *Prediction of the flow downstream of a 90°-elbow with arbitrary curvature radius and its effect on the accuracy of flow meters* **Andreas Weissenbrunner** (Physikalisch-Technische Bundesanstalt, DE)

11:15-11:30 **15 min break**

11:30-12:30

A Quality Management System for data and software - Update from the EMPIR MATHMET Project (Chair: Peter Harris)	**11:30-12:30** **Artificial Intelligence in pharma industry** (Chairs: Bernard Francq and Dan Lin)
ID 172 *A MATHMET Quality Management System for data and software* **Keith Lines** (National Physical Laboratory, UK)	11:30-11:50 ID 195 *Deep Drug Discovery* **Djork-Arné Clevert** (Bayer AG, Machine Learning Research, Berlin, DE)
This is a special session organized by MATHMET Members to offer: *(a) an introduction to the MATHMET Quality Management System (QMS) for metrology software and data,* *(b) descriptions of case studies being used by different MATHMET partners to apply the QMS,* *(c) a round table with the aim to collect input and feedback from stakeholders and end-users*	11:50-12:10 ID 143 *Aggregation in Cell Culture – App development for cell clumping scoring* **Edouard Duquesne** (Sanofi SA, Vitry-sur-Seine, FR) 12:10-12:30 ID 197 *Protein language modeling and transfer learning applied to predict TCR-epitope affinity* **Gurpreet Singh** (GlaxoSmithKline plc, Upper Providence, PA, USA)

12:30-13:20	**50 min lunch**		
13:20-14:20	**Hyperthermia techniques - Update from the EMPIR RaCHy Project** **(Chair: Alessandra Manzin)**	**Machine Learning for Metrology II** **(Chair: Nicolas Fischer)**	**Examples of measurement uncertainty evaluation - Update from the EMPIR EMUE Project** **(Chair: Maurice Cox)**
13:20-13:40	ID 158 *Uncertainty budget for acoustic characterization of tissue mimicking materials* **Piero Miloro** (National Physical Laboratory, UK)	ID 72 *Applying deep learning in metrology - an overview over some potentials and challenges* **Jörg Martin** (Physikalisch-Technische Bundesanstalt, DE)	ID 182 *Limitations of uncertainty propagation -- Measurement uncertainty for the routine determination of aqua regia extractable metals in soil* **Stephen L R Ellison** (LGC Limited, UK)
13:40-14:00	ID 120 *Simulation guided design of a TEM applicator for in vitro RF hyperthermia* **Ioannis Androulakis** (Department of Radiotherapy, Erasmus MC Cancer Institute, Rotterdam, NL)	ID 89 *A Gaussian Process approach to uncertainty evaluation for machine learning* **James Donlevy** (National Physical Laboratory, UK)	ID 196 *The role and use of measurement uncertainty in addressing specification requirements: medical temperature examples* **John Greenwood** (United Kingdom Accreditation Service, UK)
14:00-14:20	ID 183 *Modelling of iron oxide nanocubes for magnetic hyperthermia application* **Riccardo Ferrero** (Istituto Nazionale di Ricerca Metrologica, IT)	ID 156 *Is There Consistency in ML Interpretability?* **Ashish Sundar** (National Physical Laboratory, UK)	ID 198 *Evaluation of measurement uncertainty in totalization of volume measurements in drinking water supply networks* **Alvaro Ribeiro** (Laboratório Nacional de Engenharia Civil, Lisboa, PRT)
14:25-15:25	**Quantitative imaging - Update from the EMPIR QUIERO Project** **(Chair: Luca Zilberti)**	**Machine Learning for Metrology III** **(Chair: Markus Bär)**	**Regression and prediction** **(Chair: Maurizio Galetto)**
14:25-14:45	ID 95 *Optimisation of data acquisition for cardiac Magnetic Resonance Fingerprinting* **Constance Gatefait** (Physikalisch-Technische Bundesanstalt, DE)	ID 175 *The role of uncertainty in data-driven turbulence modelling* **Andrea Ferrero** (Politecnico di Torino, IT)	ID 78 *Generalization of least square method for straight line regression – A new approach* **Jacek Puchalski** (Central Office of Measures, Warsaw, POL)
14:45-15:05	ID 101 *Three dimensional MRF obtains highly repeatable and reproducible multi-parametric estimations in the healthy human brain* **Matteo Cencini** (IRCCS Stella Maris and IMAGO7 Foundation, Pisa, IT)	ID 69 *Forest embeddings for gene expression data modeling of tumor stage and survival in bladder cancer* **Mauro Nascimben** (Enginsoft Spa & University of Eastern Piedmont, IT)	ID 146 *Tensor based modelling of human motion in a collaborative human-robot approach* **Philipp Wedenig** (Joanneum Research, Graz, AUT)
15:05-15:25	ID 100 *Towards a metrological characterisation of electric properties tomography* **Alessandro Arduino** (Istituto Nazionale di Ricerca Metrologica, IT)	ID 134 *Convolutional neural network performance in the presence of physiological ECG noise* **Jenny Venton** (National Physical Laboratory, UK)	ID 153 *Modeling pyroelectric sensor signals for predicting proximity* **Franz Moser** (Joanneum Research, Graz, AUT)

15:30-16:10	**Extension of JCGM 106:2012 framework to industrial processes** **(Chair: Katy Klauenberg)**	**Multilevel measurement for business and industrial workforce development** **(Chair: Antonio Pievatolo)**	
15:30-15:50	ID 144 *Conformity assessment of lots in the framework of JCGM 106:2012* **Rainer Göb** (Universität Würzburg, DE)	ID 65 *Multilevel measurement for business and industrial workforce development: Part I* **Jan Morrison** (TIES: Teaching Institute for Excellence in STEM, Cleveland, OH, USA)	
15:50-16:10	ID 116 *Mathematical tools for a better analysis of the covariance in industrial data* **Peggy Courtois** (Deltamu, Cournon d'Auvergne, FR)	ID 66 *Multilevel measurement for business and industrial workforce development: Part II* **William Fisher** (Living Capital Metrics LLC, Sausalito, CA, USA)	
16:15-16:30	**Conclusion**		

Poster Presentations

The virtual poster room is always open during the Workshop

ID 114 *A measure of the statistical homogeneity of turbulent simulations*
Massimo Germano (Duke University, Durham, NC, USA)

ID 112 *Forecasting the COVID-19 epidemic integrating symptom search behavior: an infodemiology study*
Eugenio Alladio (Università degli Studi di Torino, IT)

ID 194 *Measurement system analysis of the CSLT measurement system - An experiment to defect diagnoses in deep drilled shafts*
Eric Ho (City University of Hong Kong, CHN)

ID 76 *Automated ML Toolbox for cyclic sensor data*
Tanja Dorst (ZeMA gGmbH, Saarbrücken, DE)

ID 115 *Intrinsic and metrological correlations on the risks of false conformity decisions*
Luciana Separovic (Faculdade de Ciências Farmacêuticas, Universidade de São Paulo, BRA)

ID 122 *A local-integral approach to electric properties tomography*
Luca Zilberti (Istituto Nazionale di Ricerca Metrologica, IT)

ID 139 *Understanding neural network classifications: Local Interpretable Model-agnostic Explanations (LIME)*
Hamza Zaheer (National Physical Laboratory, UK)

ID 148 *A probability-box based approach for measurement problems*
Tathagata Basu (Durham University, UK)

GUM Part 6 – Developing and using measurement models. An outline

Walter Bich

Key words: JCGM, Measurement uncertainty, Measurement models

1 Overview of JCGM GUM-6:2020

The Joint Committee for Guides in Metrology, JCGM, published in 2021 the document JCGM GUM-6:2020, *GUM Part 6 – Developing and using measurement models* [1]. As indicated by the title, the document is devoted to giving guidance on the development of measurement models, a key step (often the most difficult) to obtain, by the subsequent use of the model, a measurement result, *i.e.*, an estimate of the measurand and an associated uncertainty (or a complete probability density function, PDF).

The guide has been prepared by Working Group 1 of the JCGM, and has benefited from detailed reviews undertaken by member organizations of the JCGM and National Metrology Institutes.

This presentation gives an overview of the document, outlining its main points of interest and discussing some specific topics.

The document is made of 13 clauses and 6 annexes, as follows:

1. Scope
2. Normative references
3. Terms and definitions
4. Conventions and notation
5. Basic principles
6. Specifying the measurand

Walter Bich
Istituto nazionale di ricerca metrologica, Strada delle Cacce 96, 10135 Torino I, and JCGM-WG1
e-mail: w.bich@inrim.it

7. Modelling the measurement principle
8. Choosing the form of the measurement model
9. Identifying effects arising from the measurement
10. Extending the basic model
11. Statistical models used in metrology
12. Assessing the adequacy of the measurement model
13. Using the measurement model

The annexes are

A Glossary of principal symbols
B Modelling of dynamic measurements by linear time-invariant systems
C Modelling random variation
D Representing polynomials
E Cause-and-effect analysis
F Linearizing a measurement model and checking its adequacy

An extensive bibliography and an alphabetical index complete the document.

Guidance is given in the development of a measurement model, from the identification of the measurand and of the measurement principle to the inclusion of all the effects arising from the specific experimental method and setup. Several examples accompany the development of the reasoning.

The selection of the most suitable form of a given model is also given attention, by showing advantages and risks of re-parametrisation. Fitness-for-purpose, ease of manipulation, simplicity, numerical stability and efficiency in terms of computational cost are among the aspects considered. The effects of re-parametrisation on the measurement result are discussed, as well as the appropriate statistical transformations needed to ensure the integrity of the result.

Various kinds of statistical models are treated, including models for the analysis of time series, study of homogeneity, and Bayesian models. Guidance is also given about model selection and the evaluation of model uncertainty.

The adequacy of the chosen model is another important aspect. Help is given in fulfilling this delicate task.

Guidance on using a measurement model includes warnings on inappropriate linearisation, or on using the model beyond the range for which it has been validated.

The annexes treat disparate topics, from the modelling of dynamic measurements and random variation, to the representation of polynomials and cause-and-effect anaysis.

References

1. BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP and OIML Guide to the expression of uncertainty in measurement – Part 6: Developing and using measurement models. JCGM GUM-6:2020. URL https://www.bipm.org/en/publications/guides/

Uncertainty evaluation for machine learning: metrology requirements and open challenges

Andrew Thompson

Key words: Uncertainty evaluation, machine learning, metrology

1 Introduction

It is often necessary or desirable in metrology to infer information about a quantity from a number of contributions which are easier to measure. There are many scenarios in which it is challenging to build or perform computations on an analytical measurement model based upon physical understanding, making the use of machine learning (ML) in this context appealing. However, for ML to be adopted in metrology, measurements based upon it must be accompanied by reliable quantitative assessment of uncertainty.

Traditional uncertainty evaluation of the kind systematized by the GUM framework [1] assumes that the measurement model is known. Uncertainty evaluation methods in which model parameters are learned from data have also been explored in the metrology community [2]. Uncertainty evaluation in an ML context is particularly challenging since little can typically be assumed about the model. Moreover, models are often heavily overparametrised and based on large volumes of data.

2 Requirements capture

We describe in detail a framework of seven requirements which methods for uncertainty evaluation for ML in metrology should satisfy, illustrated in Figure 1.

Andrew Thompson
National Physical Laboratory, Hampton Road, Teddington, TW11 0LW, UK
e-mail: andrew.thompson@npl.co.uk

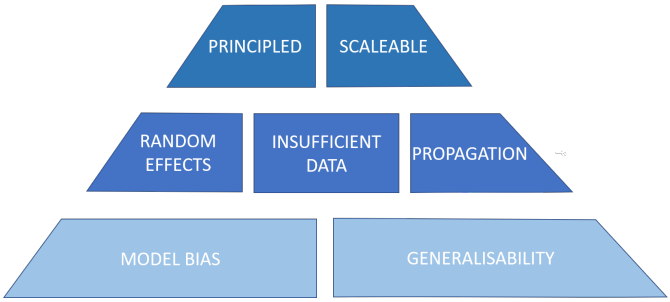


Fig. 1 Requirements for machine learning uncertainty evaluation in metrology

The base of the pyramid concerns confidence in the model itself, while the middle layer of the pyramid focuses upon the need to capture the various sources of uncertainty. The top of the pyramid captures the overarching aim, which is uncertainty evaluation that is both principled and scaleable.

The requirements were identified through an exploration of the ways in which the uncertainty evaluation challenge in ML departs from the GUM paradigm, informed by current literature and our own numerical testing.

3 Open challenges

We also describe some of the open challenges for the metrology community that arise naturally from the requirements capture. These challenges include metrics for assessing overfitting and underfitting, more principled Bayesian inference for neural networks, more scaleable Gaussian processes, and methods which combine propagated uncertainties with model uncertainty.

Acknowledgements This work is supported by the UK Government's Department of Business, Energy and Industrial Strategy (BEIS).

References

1. BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP, and OIML. Evaluation of measurement data – Guide to the expression of uncertainty in measurement (GUM 1995 with minor corrections). *Joint Committee for Guides in Metrology (JCGM)*, 100, 2008.
2. AB Forbes and JA Sousa. The GUM, Bayesian inference and the observation and measurement equations. *Measurement*, 44(8):1422–1435, 2011.

Simple Informative Prior Distributions for Type A Uncertainty Evaluation with Small Samples

Anthony O'Hagan[1] and Maurice G Cox[2]

Key words: Type A, small sample, prior distribution, Bayesian analysis

The result of a measurement, including the expression of uncertainty in the measurement, should represent a carefully considered opinion based on the metrologist's experience and expertise, as well as on the data and other information sources. This is the position of the GUM, and the requirement for such judgement is clear in the case of Type B uncertainty evaluation. However, when making Type A evaluations, involving statistical analysis of data, the GUM and its various supplements implicitly consider the data to be the only relevant information. This is unfortunate, and arguably unscientific, when, as is frequently the case, the sample size is small and the metrologist could bring other relevant information to bear.

Bayesian statistical methods allow the use of prior information in addition to the data in Type A evaluation, and have been advocated by Cox and Shirono (2017) and van der Veen (2018), amongst others. However, prior information is in principle subjective; another metrologist may bring different prior information to the analysis, leading to a different measurement result. As in other fields, there is some resistance in the metrology community to embrace Bayesian methods using substantive, subjective prior probability distributions.

We identify four desiderata that a Bayesian method employing substantive prior information should satisfy.

* *Justification*. The prior distribution should be based on judgements that are open to scrutiny and justified by reference to the metrologist's prior information and experience.

* *Simplicity*. The Bayesian procedure that derives the distribution of the measurand and summaries such as median and characteristic uncertainty should be documented in a peer-reviewed source. It should be simple to use and to be replicated.

* *Sufficient benefit*. The benefits of incorporating the prior information, for instance in terms of reduced measurement uncertainty, should be sufficient to warrant the use of the Bayesian procedure.

* *Verification*. The consistency of the prior distribution and the data should be capable of verification.

[1] Anthony O'Hagan
University of Sheffield, Sheffield, UK, e-mail: a.ohagan@sheffield.ac.uk
[2] Maurice G Cox
National Physical Laboratory, Teddington, UK, e-mail: maurice.cox@npl.co.uk

We present two informative prior distributions, denoted by ICS3 and ICS8, for use in the most basic of all Type A evaluations, where the data comprise a sample of indications assumed to be normally distributed. They represent prior information about the variance of the normal errors, in a simple form that is readily justified in practice. Formally, both require only an estimate, v, of the error variance and a judgement that it is highly likely to lie within simple bounds around v. For ICS3, the judgement is that the variance lies within a factor of 9 above or below v, while for ICS8 it is that it lies within a factor of 3 above or below v. We argue that one or other of these judgements will be eminently reasonable in practical application in measurement and calibration laboratories (Justification). We present simple (indeed, trivial) formulae for incorporating these prior distributions in the Type A evaluation (Simplicity).

We show the effect of these prior distributions in two examples (Sufficient benefit).

The first example is of a single Type A evaluation and we compute the typical reduction in characteristic uncertainty when using one of these prior distributions, relative to the standard GUM analysis that is equivalent to using a noninformative prior. For instance, for a sample of size 4 using the ICS3 prior distribution the median reduction in uncertainty is 23 %, while with the ICS8 prior distribution it is 27 %.

The second example involves a measurement model with six inputs, all of which are evaluated by Type A assessment with samples of size 4. Using ICS3 prior distributions for all six evaluations reduces uncertainty in the measurand by more than 30 %, while the use of ICS8 distributions yields a reduction of more than 40 %.

Finally, we present simple guidance for monitoring the validity of these prior distributions over time when used in a sequence of measurements (Verification).

Type A evaluation with small samples has been seen as a challenging problem for practical metrology, with a recent special meeting organised by the Joint Committee on Guides in Metrology on this topic. We believe that our proposed simple prior distributions can represent a substantial contribution to addressing this challenge.

Acknowledgement

This work was supported by an ISCF (Industrial Strategy Challenge Fund) Metrology Fellowship grant provided by the UK Government's Department for Business, Energy and Industrial Strategy (BEIS).

References

1. Cox, M., Shirono, K.: Informative Bayesian Type A uncertainty evaluation, especially applicable to a small number of observations. Metrologia **54**, 642-652 (2017)
2. van der Veen, A.M.H.: Bayesian methods for Type A evaluation of standard uncertainty. Metrologia **55**;5, 670-684 (2018)

Deep Learning for Inverse Problems – Applying Ensemble Learning for Uncertainty Quantification

Lara Hoffmann[1]* and Clemens Elster[1]

Key words: deep learning, uncertainty quantification, inverse problems, ensemble learning, computational optical form measurements

1. Motivation

Deep learning methods are based on neural networks and have become a popular and powerful tool for data analysis. They have been successfully applied in many different domains including computational imaging, healthcare or natural language processing. The main drawback of deep neural networks is their black-box character which can cause a loss of trust in their predictions. A possible remedy is to estimate the uncertainty associated with a prediction, which will be discussed in this talk based on an ensemble learning method.

2. Background

Several techniques have been developed to establish confidence in the predictions of deep neural networks. Apart from thorough testing, these techniques can be structured into two main groups.

The first group focuses on explaining why a certain prediction occurs and is known under the keyword “explainability methods”. Examples comprise sensitivity analysis, layer-wise relevance propagation [Lapuschkin(2019)] or the use of the Fisher information to detect unusual input data [Martin(2020)]. The second group aims at an “uncertainty quantification” associated with the prediction of a trained network. In the machine learning community, uncertainty is commonly divided into its aleatoric and its epistemic part. The former addresses the non-reducible part of the uncertainty which arises, for example, when predictions are made for noisy input data; the latter quantifies the model uncertainty of a trained network which could in principle be reduced when using more training data. Among the most popular

* e-mail: lara.hoffmann@ptb.de
[1] Physikalisch Technische Bundesanstalt, Abbestr. 2-12, 10587 Berlin, Germany

techniques for uncertainty quantification in deep learning are dropout methods [Gal(2016)] and ensemble methods [Lakshminarayanan(2017)], respectively.

Ensemble learning is based on training multiple deep neural networks in parallel. All trained networks are used to make predictions and to assign uncertainties associated with those predictions. Ensemble learning is seen as a promising technique; it often yields better predictions than when using just a single trained network, and it provides an uncertainty associated with a prediction at the same time. Furthermore, this technique is easy to implement and scales well to high-dimensional problems. In this talk we introduce and discuss the technique of ensemble learning, and illustrate its application in terms of a large-scale inverse problem from computational optical form measurements [Hoffmann(2021)].

References

1. Gal, Y., Ghahramani, Z. Dropout as a Bayesian approximation: Representing model uncertainty in deep learning. PMLR, pp. 1050-1059 (2016).
2. Hoffmann, L., Fortmeier, I., Elster, C. Deep Neural Networks for Computational Optical Form Measurements. arXiv preprint arXiv:2103.01259 (2021).
   http://www.rsc.org/dose/title of subordinate document. Cited 15 Jan 1999
3. Lakshminarayanan, B., Pritzel, A., Blundell, C. Simple and Scalable Predictive Uncertainty Estimation using Deep Ensembles. NIPS, pp. 6405-6416 (2017).
4. Lapuschkin, S., Wäldchen, S., Binder, A. et al. Unmasking Clever Hans predictors and assessing what machines really learn. Nat Commun 10, 1096 (2019).
5. Martin, J., Elster, C. Detecting unusual input to neural networks. Appl Intell (2020). [DOI: 10.1007/s10489-020-01925-8]

Transferability of GUM-S1 Type A Uncertainties - a Bayesian perspective -

Gerd Wübbeler[1]*, Manuel Marschall[1] and Clemens Elster[1]

Key words: GUM-S1, Type A uncertainty evaluation, transferability, Bayesian inference

1. Summary

The Type A uncertainty evaluation of Supplement 1 to the GUM [JCGM-101(2008)] is equivalent to a Bayesian inference that applies a particular non-informative prior [Elster&Toman(2009)]. We demonstrate in terms of examples that, caused by this implicit, automatic choice of the prior, the transferability of results obtained by the GUM-S1 Type A approach is challenged [Wübbeler(2020a)]. A possible remedy of this issue is to adopt a fully subjective Bayesian perspective instead. In this context, we propose a simple Monte Carlo procedure which can be applied for a subjective Bayesian uncertainty evaluation in linear models [Wübbeler(2020b)].

2. Background

Transferability represents a key requirement for the evaluation of measurement uncertainty in metrology. This is particularly relevant for ensuring the traceability of measurements performed on the shop floor level to the international system of units (SI). We have demonstrated in terms of generic examples that the GUM-S1 Type A uncertainty evaluation does not always satisfy the requirement of transferability, i.e. reusing the probability distribution produced by GUM-S1 in a subsequent uncertainty exercise can lead to inadequate results for non-linear models [Wübbeler(2020a)]. Furthermore, already for linear models the Type A uncertainty evaluation of GUM-S1 can produce unsatisfactory solutions. Similar conclusions have already been reached in [Giaquinto(2016)].

* e-mail: gerd.wuebbeler@ptb.de
1 Physikalisch-Technische Bundesanstalt, Braunschweig & Berlin, Germany

The reason for the lack of transferability is due to the automatic choice of a non-informative prior implicitly made by the Type A uncertainty evaluation of GUM-S1. This lack of transferability can be generalized to Bayesian Type A uncertainty evaluations utilizing other non-informative or vaguely informative priors. On the other hand, the Monte Carlo method used to calculate the "propagation of distributions" is an adequate method for approximating the distribution of a function of random variables given the joint distribution for those random variables. We therefore argue that GUM-S1 ought to carefully distinguish between the numerical method for "propagating distributions" and the uncertainty evaluation in the presence of data. Regarding the latter, the current approach taken in GUM-S1 is not recommended and a fully subjective Bayesian inference proposed instead.

Conducting a subjective Bayesian uncertainty evaluation requires statistical skills and, moreover, often needs to be evaluated by methods such as Markov chain Monte Carlo (MCMC) (see, e.g. [Demeyer(2020)]). In [Wübbeler(2020b)], we have proposed a simple Monte Carlo procedure which can be applied for a subjective Bayesian uncertainty evaluation in linear models. The procedure is capable of accounting for prior knowledge about the measurand and the employed measurement device in an accessible way, and it provides a coherent treatment of a small number of observations, including the case of a single observation only. To ease the application of the proposed Bayesian uncertainty evaluation open source Python software has been made available.

References

1. Demeyer, S., Fischer, N., & Elster, C.: Guidance on Bayesian uncertainty evaluation for a class of GUM measurement models. Metrologia, *58*(1), 014001 (2020).
2. Elster, C., Toman, B.: Bayesian uncertainty analysis under prior ignorance of the measurand versus analysis using the Supplement 1 to the Guide: a comparison. Metrologia, 46(3), 261-266, 2009.
3. Giaquinto, N., Fabbiano, L.: Examples of S1 coverage intervals with very good and very bad long-run success rate. Metrologia, 53, S65 (2016).
4. Joint Committee for Guides in Metrology: Evaluation of Measurement Data-Supplement 1 to the Guide to the Expression of Uncertainty in Measurement'-Propagation of Distributions Using a Monte Carlo Method (BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP and OIML, JCGM 101:2008)
5. Wübbeler, G., Elster, C.: On the transferability of the GUM-S1 type A uncertainty. Metrologia, 57(1), 015005 (2020a).
6. Wübbeler, G., Marschall, M., & Elster, C.: A simple method for Bayesian uncertainty evaluation in linear models. Metrologia, 57(6), 065010 (2020b).

Deep Learning based instance segmentation: application to agglomerated titanium dioxide particles measured by scanning electron microscopy

Paul Monchot[1], Loïc Coquelin[1], Khaled Guerroudj[1], Nicolas Feltin[2], Alexandra Delvallée[2], Loïc Crouzier[2] and Nicolas Fischer[1]

Key words: Scanning Electron Microscopy, Mask R-CNN, Deep Learning, Particle Size Distribution, Instance Segmentation, TiO_2, Agglomerate

1 Motivation

In many cases, the estimation of particle size distribution of a nanoparticle population remains a major challenge for the industrial development of nanomaterials. Titanium dioxide (TiO_2) in nanoparticulate form is produced in very large quantities for intensive use in many applications (food, paint, construction products, etc.), and their characterization becomes primordial.

Thus, this work proposes a methodology to automate the size characterization of titanium dioxide particles imaged by scanning electron microscopy (SEM) as shown in figure 1. Today, SEM imaging is widely used in laboratories and manufacturing industries and is considered in metrology as a reference technique capable of reliably determining the size, size distribution and shape of nanoparticles. However, the study of this type of particles remains challenging because of their non-spherical shape and their ability for agglomeration. The commonly used image processing methods such as watershed or active contour lack of generalizability. Indeed, the adjustment of the parameters of such algorithms is almost systematically required and slows down the automation of the processing chain. Moreover, these traditional methods are not efficient when performing segmentation over non smooth convex objects that strongly overlap. Due to this complexity, the characterization of this

e-mail: firstname.lastname@lne.fr
[1] Data Science and Uncertainty Department, National Laboratory of Metrology and Testing, 29 avenue Roger Hennequin, Trappes, 78197, France
[2] Department of Materials science, National Laboratory of Metrology and Testing, 29 avenue Roger Hennequin, Trappes, 78197, France

type of content is not automated and is frequently performed manually by experts in nanometrology.

2 Method and results

To overcome this problem, we propose a methodology relying on state of the art deep learning algorithm, namely Mask-RCNN (He et al. 2017), coupled with transfer learning. Moreover, specific hyper-parameters tuning, a small architecture modification and a task specific database augmentation technique (allowing to generate "fake" SEM images and thus increasing the training set diversity) have been performed in order to enhance segmentation accuracy.

To evaluate the performance of the proposed method, we use a test set composed of 19 images (for a total of 3741 particles) and calculate the detection score (84%), the mAP value (60.6) and the Sørensen–Dice coefficient (Sorensen 1948) (0.95). The results confirm that the approach conducts to robust and accurate segmentation of agglomerated titanium dioxide particles. Based on these accurate predictions, scanning electron microscopy sample characterization commonly relies on derived quantities such as the projected equivalent diameter. For this measurand, more than 96 % of measurements show an error of less than 5 % and 51 % of measurements show an error of less than 1 % when tested against ground truths segmentation.

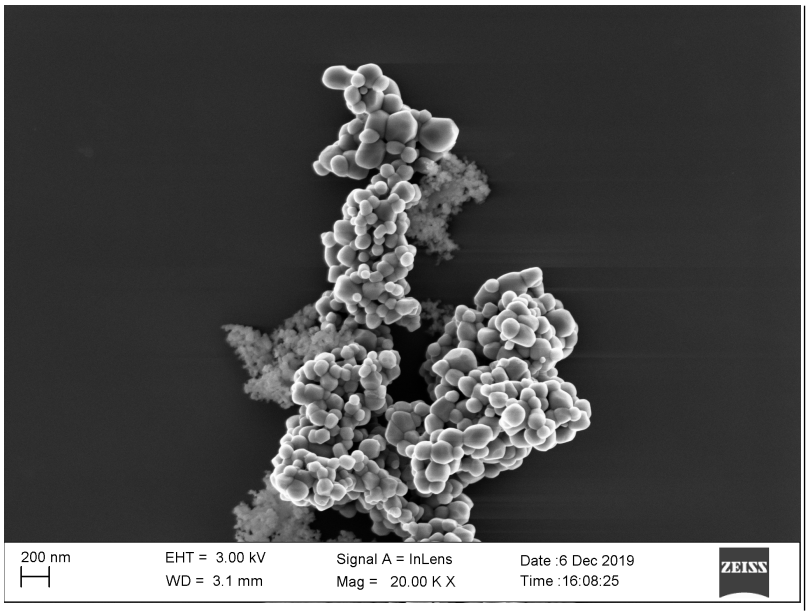


Fig. 1 Example of SEM image showing a mixture of TiO_2 and SiO_2 particles

References

He, K., Gkioxari, G., Dollár, P. & Girshick, R. (2017), Mask r-cnn, *in* 'Proceedings of the IEEE international conference on computer vision', pp. 2961–2969.

Sorensen, T. A. (1948), 'A method of establishing groups of equal amplitude in plant sociology based on similarity of species content and its application to analyses of the vegetation on danish commons', *Biol. Skar.* **5**, 1–34.

Metrological redundancy in distributed measurements

Gertjan Kok[1] and Peter Harris[2]

Key words: redundancy, distributed measurements, metrics, uncertainty, temperature, factory of the future, testbed

1. Introduction

In the EMPIR project Metrology for the Factory of the Future [Met4FoF, 2018] it is investigated how redundancy in sensor networks can be quantified and exploited in different contexts. In [Kok, 2020a] and [Kok, 2020b] the impact of redundancy on aggregated measurements was studied. In this contribution, redundancy in distributed measurements will be addressed. Metrics quantifying the amount of redundancy will be proposed. The testbed of the company SPEA will be presented as an example of a distributed temperature measurement in a factory of the future setting. Measurement data from this testbed will be modelled and evaluated in various ways. The question regarding how much redundancy is present in the data will also be addressed.

2. Quantifying metrological redundancy

The term redundancy can indicate various related concepts. In information and communications technology (ICT) systems it can mean that multiple systems are operating in parallel, having multiple, independent paths in a network, or storing data at multiple locations. In a metrological context we define it as the property that there are multiple, independent ways of deriving the value of the measurand Y from a set of sensor measurements $\boldsymbol{X}$, quantified by a notion of gradual uncertainty increase when sensors fail or are removed.

Two metrics that can be used to quantify redundancy in a metrological context are:

RedExcess(U_{req}) maximum number of arbitrarily selected sensors that can be removed without compromising the requirement U_{req} on the measurement uncertainty of the measurand

and

$$\text{RedUnc}(m) = \max_{|\boldsymbol{W}|=m,\ s\in S} \left(u\left(Y(s,\boldsymbol{X}/\boldsymbol{W})\right) - u\left(Y(s,\boldsymbol{X})\right)\right).$$

[1] Gertjan Kok
VSL, Thijsseweg 11, 2629 JA Delft, The Netherlands, e-mail: gkok@vsl.nl
[2] Peter Harris
NPL, Hampton Road, Teddington, UK, e-mail: peter.harris@npl.co.uk

Here, $u\big(Y(s,\boldsymbol{X}/\boldsymbol{W})\big)$ denotes the uncertainty of the distributed measurand Y at location s when using the set of sensor measurements $\boldsymbol{X}/\boldsymbol{W}$, where $\boldsymbol{W}$ denotes the set of sensors which is removed from the full set $\boldsymbol{X}$. Informally, RedUnc(m) is the maximum uncertainty increase at an arbitrary location when removing m arbitrary sensors from the sensor network.

3. Case study: SPEA testbed

In the testbed of the Italian company SPEA [SPEA, 2018] 196 small temperature sensors are simultaneously calibrated by comparison with 64 reference sensors. The temperature field is not perfectly homogeneous and is thus measured at 64 locations and predicted at 196 other locations using interpolation methods. Seven different mathematical models have been used to model the data under different assumptions. The models include simple models assuming a constant temperature measured with known or unknown measurement uncertainty, nearest-neighbour modeling of the data and Gaussian processes with different kernels. The models will be discussed together with the resulting fits and interpolation uncertainty. The models will also be used to assess the amount of redundancy present in the testbed.

Acknowledgement

This work has been performed in the EMPIR 17IND12 project ‘Metrology for the Factory of the Future’. This project has received funding from the EMPIR programme co-financed by the Participating States and from the European Union's Horizon 2020 research and innovation programme.

References

[Kok, 2020a] G. Kok and P. Harris: "Quantifying Metrological Redundancy in an Industry 4.0 Environment," 2020 IEEE International Workshop on Metrology for Industry 4.0 & IoT, Roma, Italy, 2020, pp. 464-468, doi: 10.1109/MetroInd4.0IoT48571.2020.9138235.

[Kok, 2020b] G. Kok and P. Harris, "Uncertainty Evaluation for Metrologically Redundant Industrial Sensor Networks," 2020 IEEE International Workshop on Metrology for Industry 4.0 & IoT, Roma, Italy, 2020, pp. 84-88, doi: 10.1109/MetroInd4.0IoT48571.2020.9138297.

[Met4FoF, 2018] EMPIR (European Metrology Programme for Innovation and Research) project ‘Metrology for the Factory of the Future’ (17IND12, Met4FoF), 2018 – 2021, www.met4fof.eu

[SPEA, 2018] Blog post on Met4FoF website presenting the SPEA testbed, 11 June 2018, https://www.ptb.de/empir2018/met4fof/information-communication/blog/detail-view/?tx_news_pi1%5Bnews%5D=13&tx_news_pi1%5Bcontroller%5D=News&tx_news_pi1%5Baction%5D=detail&cHash=f47a890ea864fa24603143f50ae09ac1

Inversion of point clouds for holistic Screw Thread Metrology

Anita Przyklenk and Sebastian Schädel and Martin Stein

Key words: Inversion, Holistic Data Evaluation, Parametrization, Screw Thread Metrology, Coordinate Measuring Machines

Abstract

We present a constrained non-linear least-squares optimization algorithm for the application in coordinate metrology, which we particularly use for holistic 3D screw thread calibration [Schädel(2019)]. Therefore, we have implemented the so-called FUNKE code by [Sourlier(1992)] to minimize the objective function $Q(\mathbf{W})$ depending on the parameter vector $\mathbf{W} \in \mathbb{R}^{(2n+g+l)}$ with the numbers of measurement points n, geometry parameters g and position and orientation parameters l in the three-dimensional Euclidean space.

The FUNKE code has some advantages over conventional Orthogonal Distance Regression (ODR) methods [Turner(1999)]. For example, a separation of dimension, form and position/orientation parameters is performed, which allows a fundamental improvement of determining the actual dimension and form deviations of objects under inspection in coordinate metrology. The parameter vector is specified as $\mathbf{W} = (\mathbf{U}, \mathbf{p}, \mathbf{A}, \mathbf{T})$, where $\mathbf{U} = (u_1, v_2, ..., u_n, v_n)$ denotes surface-, $\mathbf{p}$ geometry-

Anita Przyklenk
Physikalisch-Technische Bundesanstalt (PTB), Bundesallee 100, 38116 Braunschweig, Germany,
e-mail: anita.przyklenk@ptb.de

Sebastian Schädel
ZEISS Industrial Metrology, Carl-Zeiss-Straße 22, 73447 Oberkochen, Germany,
e-mail: sebastian.schaedel@zeiss.com

Martin Stein
Physikalisch-Technische Bundesanstalt (PTB), Bundesallee 100, 38116 Braunschweig, Germany,
e-mail: martin.stein@ptb.de

, **A** orientation- and **T** position parameters. Let $\mathbf{P} \in \mathbb{R}^{3n}$ be the gathered point cloud and $\mathbf{F} \in \mathbb{R}^{3n}$ be footprints on the form element $\mathbf{S}_j(\mathbf{U}_j, \mathbf{p}_j, \mathbf{A}_j, \mathbf{T}_j)$ of a screw thread with $j = 1, 2$ for the upper and the lower screw surface. Then the residual is defined as the distance

$$\Delta\mathbf{X} = \mathbf{P} - \mathbf{F}(\mathbf{W}) \in \mathbb{R}^{3n} \tag{1}$$

to be minimized. However, the objective functions in $Q(\mathbf{W}) = \min(||\Delta\mathbf{X}(\mathbf{W})||)$ are easy to change by the implementation of different geometries in parametric representation. In particular, starting from Sourlier's idea of using parameterized form elements, we developed a measurement strategy on coordinate measuring machines that refers to the same form element as during the fitting and the evaluation of dimension and form deviations [Schädel(2019)]. A drawback of the parametric representation is that two additional parameters per measured point in the Euclidean 3D space are needed, so that a high-dimensional overdetermined problem has to be solved

$$\mathbf{J}^{\mathrm{T}}(\mathbf{W})\mathbf{J}(\mathbf{W})\Delta\mathbf{W} = \mathbf{J}^{\mathrm{T}}(\mathbf{W})\Delta\mathbf{X}(\mathbf{W}) \tag{2}$$

with $\mathbf{J}(\mathbf{W}) \in \mathbb{R}^{3n\times(2n+g+l)}$. This problem is overcome by converting full matrices into sparse form. Any zero elements are squeezed out in dependence on constraints and symmetries of the l positioning parameters, which leads to a reduction of the matrix dimensions. Since the point clouds of interest are composed of a maximum of $\sim 100\,000$ coordinates and initial parameters are easy to guess due to auxiliary measurements and known nominal values for the geometry, a sufficiently accurate result is usually provided after a few iterations. In the application it turned out that calibration results obtained by the novel and conventional rather two-dimensional method agree remarkably [Przyklenk(2021)]. As soon as screw thread gauges have shape deviations, the new method becomes advantageous due to the significantly increased resolving capability. In addition, the uncertainty has been reduced in relation to the conventional method by at least a quarter depending on the tested gauge.

The algorithm presented here, in combination with area-based measuring strategies and holistic interpretation of inversion results, will expand PTB's range of calibrations in the field of coordinate metrology.

References

[Przyklenk(2021)] Przyklenk A, Schädel S and Stein M (2019) Verification of a calibration method for 3D screw thread metrology. Measurement Science and Technology, 10.1088/1361-6501/abead2.

[Schädel(2019)] Schädel S, Wedmann A and Stein M (2019) Advanced screw thread metrology using an areal measuring strategy and a holistic evaluation method. Measurement Science and Technology, 10.1088/1361-6501/ab1501.

[Sourlier(1992)] Sourlier D and Bucher A (1992) Exact bestfit algorithm applicable to sculptured surfaces or to any non-regular surfaces in parametric form. tm - Technisches Messen, 59(7-8), 10.1524/teme.1992.59.78.293

[Turner(1999)] Turner D A (1999) The approximation of Cartesian coordinate data by parametric orthogonal distance regression. University of Huddersfield, Dissertation

Co-calibration of sensor networks

Alistair Forbes

Key words: calibration, Gaussian processes, sensor networks, uncertainty quantification

Collocation studies are used to assess the performance of air quality sensors, for example, NO_2 diffusion samplers. In a collocation study, a number of test sensors to be assessed are placed along side one or more calibrated reference sensors. Measurements are gathered at regular time intervals from all the sensors and these are then used to calibrate the test sensors relative to the calibrated sensor. Standard approaches to calibrate the test sensors usually involve a series of independent analyses involving the data from the reference sensor and each of the test sensors. However, there are advantages in analysing the data for the complete ensemble of sensors since all the sensors are nominally measuring the same signal. In this paper, we discuss an ensemble model of a collocation study of the form

$$y_i = a_k + b_k z_{j,k} + \epsilon_i, \quad \epsilon_i \in \mathrm{N}(0, \sigma_i^2),$$

where y_i is the observation recorded by sensor $k = k(i)$ at time $j = j(i)$, $z_{j,k}$ is the underlying signal at time j at the location of the kth sensor, a_k and b_k are calibration parameters associated with the kth sensor and σ_i is (an estimate of) the standard deviation of the random effects associated with the ith measurement. We assume there is informative prior information $p(\boldsymbol{a}, \boldsymbol{b})$ available about the calibration parameters and that for at least one of the reference sensors, the prior information is informative based on a prior calibration experiment for example.

In a *strict* collocation experiment, it is assumed that at any given time all sensors see the same signal so that $z_{j,k} = z_j$. In an *approximate* collocation experiment, we do not insist that all sensors see the same signal, but that the separate signals are correlated with the correlation strength determined by some metric related to the spatial separation of the sensors. This correlation can be modelled as $\boldsymbol{z}_j \in \mathrm{N}(\mathbf{1} z_{j,0}, V_j)$, where $\mathbf{1}$ is an $n_k \times 1$ vector of 1s and V_j is an $n_K \times n_K$ variance matrix encoding the correlation in the signals $z_{j,k}$ at time t_j. The term $z_{j,0}$ represents the spatial mean at the jth time while the variance matrix V_j controls the degree of smoothness of the departure from

Alistair Forbes
National Physical Laboratory, Teddington UK, TW11 0LW, e-mail: alistair.forbes@npl.co.uk

the spatial mean as a function of spatial separation [1, 4]. If V_j has Cholesky factorisation $V_j = L_j L_j^\top$, we can write

$$z_{j,k} = z_{j,0} + \boldsymbol{\ell}_{j,k}\boldsymbol{e}_j, \quad \boldsymbol{e}_j \in \mathrm{N}(0, I), \tag{1}$$

where $\boldsymbol{\ell}_{j,k}$ is the kth row of L_j.

Assuming that the σ_i are known, point estimates $\hat{\boldsymbol{\alpha}}$ of the model parameters $\boldsymbol{\alpha} = \langle \boldsymbol{z}_0, \{\boldsymbol{e}_j\}, \boldsymbol{a}, \boldsymbol{b}\rangle$ can be found by maximising the posterior distribution $p(\boldsymbol{\alpha}|\boldsymbol{y})$. The variance matrix associated with these point estimates can be calculated from the Gaussian approximation $\mathrm{N}(\hat{\boldsymbol{\alpha}}, V_{\boldsymbol{\alpha}})$ to $p(\boldsymbol{\alpha}|\boldsymbol{y})$ constructed from a quadratic approximation to $-\log p(\boldsymbol{\alpha}|\boldsymbol{y})$ evaluated at $\hat{\boldsymbol{\alpha}}$. However, the interaction between the scale correction parameters $\boldsymbol{b}$ and the prior information on $\boldsymbol{z}_j = \langle z_{j,0}, \boldsymbol{e}_j\rangle$ leads to very biased point estimates. However, the model can be presented as one in which $\boldsymbol{y}$ arises according to the model

$$y_i|\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{z} \in \mathrm{N}(C\boldsymbol{a} + D(\boldsymbol{b})\boldsymbol{z}, \sigma_i^2),$$

with priors $p(\boldsymbol{a}, \boldsymbol{b})$ and $p(\boldsymbol{z})$ where $\boldsymbol{z} \sim \mathrm{N}(\boldsymbol{z}_0, V_{\boldsymbol{z}})$.

The form of the model allows the marginalised posterior $p(\boldsymbol{a}, \boldsymbol{b}|\boldsymbol{y})$ to be determined analytically (up to a normalising constant) [2, 4]. Point estimates of $\boldsymbol{a}$ and $\boldsymbol{b}$ determined by maximising $p(\boldsymbol{a}, \boldsymbol{b}|\boldsymbol{y})$ are much more useful since the ratio of the number of observations to the number of parameters is much larger and the good asymptotic properties [3] of maximum likelihood estimation come into play.

In this paper we look at different computational approaches for the co-calibration of sensor networks from collocation experiments that exploit the structure in the model to improve computational efficiency.

Acknowledgements This work is supported by the UK Governments Department of Business, Energy and Industrial Strategy (BEIS) through the National Measurement System programme for Data Science.

References

1. N. Cressie and C. K. Wikle. *Statistics for Spatio-Temporal Data*. Wiley, Hoboken, New Jersey, 2011.
2. A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin. *Bayesian Data Analysis*. Chapman & Hall/CRC, Boca Raton, Fl., second edition, 2004.
3. H. S. Migon and D. Gamerman. *Statistical Inference: an Integrated Approach*. Arnold, London, 1999.
4. C. E. Rasmussen and C. K. I. Williams. *Gaussian Processes for Machine Learning*. MIT Press, Cambridge, Mass., 2006.

Model error in Bayesian inversion

M. Casfor Zapata and S. Heidenreich and N. Farchmin and P. Hagemann

Key words: model error, measure transport, Bayesian inversion

1 Introduction

Bayesian inversion is a powerful tool for solving inverse problems, but limited by the complexity of the forward model. For computationally expensive forward problems the forward model usually has to be simplified. The inaccuracy of the forward model, the so called model error or model uncertainty is often ignored in Bayesian setting what propagates to an error in the posterior distribution as discussed in [Brynjarsdóttir and O'Hagan, 2014].

Including the model error in the Bayesian setting, as it is done for example in [Kennedy and O'Hagan, 2001], allows us to use a simplified forward model $f \approx F$ for faster computation, but even so get a posterior distribution close to the exact one. In contrast to the approach of [Kennedy and O'Hagan, 2001] we use the concept of transport maps to describe the model error distribution and the posterior distribution. This approach provides a more flexible description of the model error and allows us to sample fast from the posterior distribution. Furthermore machine learning methods can simply be applied. In the presented contribution we include the model error into the Bayesian framework using a transport map given by the model error function $M = F - f$. We further solve the resulting inverse problem using a transport map given by an Invertible Neural Network (INN) [Hagemann, 2021] .

M. Casfor Zapata
Physikalisch-Technische Bundesanstalt, Abbestraße 2-12, 10587 Berlin,
e-mail: maren.casfor@ptb.de

S.Heidenreich
Physikalisch-Technische Bundesanstalt, Abbestraße 2-12, 10587 Berlin,
e-mail: Sebastian.Heidenreich@ptb.de

2 Methods

The used concept of transport maps [Marzouk et al., 2016] consists in a transformation between two distributions as in Figure 1. We replace the common sampling algorithm (MCMC) by a transport map T which approximates the posterior distribution by the push forward

$$T_{\#}p_0 = p_0 \circ T^{-1} |\det \nabla_x T^{-1}|$$

of a standard normal distribution $p_0 = N(0,1)$. The transport map T is given by an Invertible Neural Network (INN) [Hagemann, 2021] and trained by minimizing the KL-Divergence of the push forward $T_{\#}p_0$ and the posterior π^y.

To include the model inaccuracy as presented in [Calvetti et al., 2018] the model error is included as a stochastic variable distributed by the push forward $M_{\#}\pi$. Therefore the model error function M is used as a transport map to generate samples from the push forward distribution $M_{\#}\pi$, for $\pi = T_{\#}p_0$.

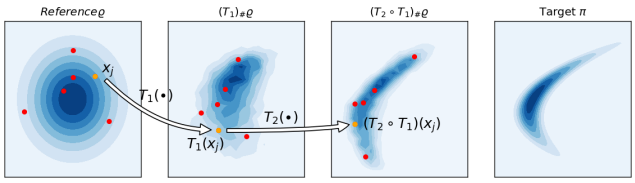


Fig. 1 Transformation of a reference densitiy ρ to the target density π unsing transport maps

3 Application

A possible application can be an inverse problem in scatterometry where the simulation by a Finite Element Method (FEM) is very expensive. The model error function would then be given by the difference of the FEM simulation on a fine mesh and a coarse mesh. The reconstruction is then based on the FEM simulation on the coarse mesh and samples from the model error distribution.

References

[Brynjarsdóttir and O'Hagan, 2014] Brynjarsdóttir, J. and O'Hagan, A. (2014). Learning about physical parameters: the importance of model discrepancy. *Inverse Problems*, 30(11).

[Calvetti et al., 2018] Calvetti, D., Dunlop, M., Somersalo, E., and Stuart, A. (2018). Iterative updating of model error for bayesian inversion. *Inverse Problems*, 34(2).

[Hagemann, 2021] Hagemann, P. (2021). Invertible neural networks versus mcmc for posterior reconstruction in grazing incidence x-ray fluorescence. to appear.

[Kennedy and O'Hagan, 2001] Kennedy, M. C. and O'Hagan, A. (2001). Bayesian calibration of computer models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(3).

[Marzouk et al., 2016] Marzouk, Y., Moselhy, T., Parno, M., and Spantini, A. (2016). *Sampling via Measure Transport: An Introduction*. Springer International Publishing, Cham.

A novel method for Callendar-Van Dusen interpolation of temperature calibration points

Graziano Coppa[1]

Key words: temperature metrology, Callendar-Van Dusen, spline

1. Introduction and scope

The Callendar-Van Dusen (CVD) equation (Van Dusen, 1925) has been used for decades in the interpolation of platinum resistance thermometers (PRTs) calibration points. This empirical equation, which represents the extension to $t < 0$ ℃ of the original Callendar equation (Callendar, 1899), can be written in several forms, one of the most used being

$$R = R0\ (1 + A\ t + B\ t^2 - 100\ C\ t^3 + C\ t^4), \quad (1)$$

where *R* is the resistance of the PRT at temperature *t*, *R0* is the resistance of the PRT at 0 °C, and *A*, *B* and *C* are coefficients of the interpolation. For temperatures above 0 °C, the equation simplifies as the original Callendar equation:

$$R = R0\ (1 + A\ t + B\ t^2). \quad (2)$$

There are currently no prescriptions on the mathematical method of interpolation for PRT calibration points, therefore there are a number of sub-optimal methods in use, which often are geared towards the optimization of the fit only for calibration subranges. This works inquires into a novel method of interpolation that minimizes the squared sum of fit residuals along all the calibration range, therefore representing a means of uncertainty reduction in an overall calibration budget.

2. Method

The interpolation system presented in this work makes use of piecewise Constrained Least Squares (CLS) and consists in the familiar least squares method to which a system of constraints is applied.

Basically, the whole calibration range is interpolated by a spline, made up by the two forms of the CVD equations (Eqn. 1 for $t < 0$ ℃ and Eqn. 2 for $t > 0$ ℃). The constraints set is composed by the continuity conditions of Eqns 1 and 2, as well as

[1] Graziano Coppa
INRiM, Strada delle Cacce, 91, Torino, e-mail: g.coppa@inrim.it

their first derivative, at $t = 0$ ℃. Other constraints are given by the equality conditions of coefficients of Eqns 1 and 2.

The problem can be written as the minimization of the objective function:

$$\|\boldsymbol{Ax} - \boldsymbol{b}\|^2$$

constrained by

$$\boldsymbol{Cx} = \boldsymbol{d},$$

where $\boldsymbol{x}$ is the unknown vector of CVD coefficients, dimension n, of which only 3 are linearly independent; $\boldsymbol{A}$ is the Vandermonde matrix of the observations (Shores, 2018), dimensions $m \times n$, where m is the number of calibration points; $\boldsymbol{b}$ is the observations vector (resistance measurements), dimension m; $\boldsymbol{C}$ is the constraints matrix, dimension $p \times n$, with p representing the number of constraints; $\boldsymbol{d}$ is the response vector of constraints, dimension p.

3. Conclusions

The theoretical construction of the system, based on the null-space method (Benzi *et al.*, 2005), will be shown, as well as its practical implementation (Excel and R functions). Also, a comparison with other methods of interpolation already in use will be provided (e.g. Figure 1) using a case study.

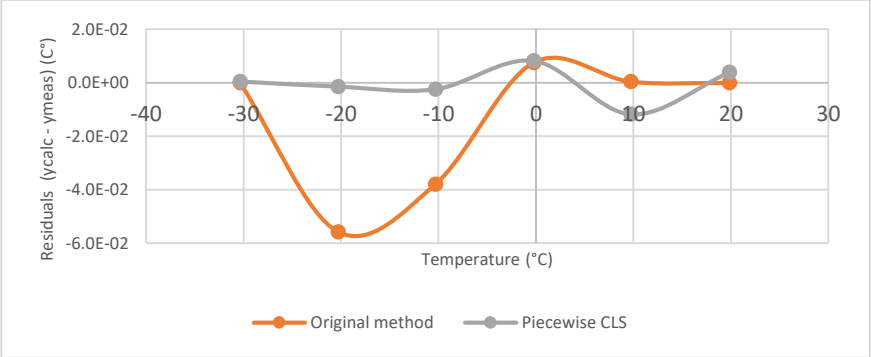


Figure 1: Example of fit residual comparison between one commonly used method of CVD interpolation and the one proposed in this work

References

1. Benzi M, Golubt GH, Liesen J. 2005. Numerical solution of saddle point problems. *Acta Numerica*, 14: 1–137. https://doi.org/10.1017/S0962492904000212.
2. Callendar HL. 1899. XIII. Notes on platinum thermometry . *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 47(285): 191–222. https://doi.org/10.1080/14786449908621251.
3. Shores TS. 2018. *Applied Linear Algebra and Matrix Analysis*. *Springer*. Springer International Publishing: Cham.

4. Van Dusen M V. 1925. Platinum-resistance thermometry at low temperatures. *Journal of the American Chemical Society*, 47(2): 326–332. https://doi.org/10.1021/ja01679a007.

Invertible Neural Networks for Grazing Incidence X-Ray Fluorescence Parameter Reconstruction

Nando Farchmin and Paul Hagemann and Sebastian Heidenreich

Key words: uncertainty quantification, invertible neural networks, measure transport, X-ray fluorescence

Grazing Incidence X-Ray Fluorescence

As a non-destructive technique, grazing incidence X-ray fluorescence (GIXRF) [Hoenicke et al., 2020] is of particular interest for many industrial applications, such as investigating the shape of micro chip layouts. Mathematically, the reconstruction of nanostructures can be rephrased as an inverse problem. Given grazing incidence X-ray fluorescence measurements, we recover the distribution of the shape parameters of a grating, where it appears crucial to take a Bayesian perspective to account for measurement errors. We consider a silicon nitride (Si_3N_4) lamellar grating on a silicon substrate. The grating oxidized in a natural fashion resulting in a thin SiO_2 layer. A cross-section of the lamellar grating is shown in Fig. 1, left. It can be characterized by seven parameters namely the height (h) and middle-width (cd) of the line, the sidewall angle (swa), the thickness of the covering oxide layer (t_t), the thickness of the etch offset of the covering oxide layer beside the lamella (t_g) and additional layers on the substrate (t_s, t_b). A significant cause of uncertainty are noisy measurements, which are characterized by a user defined error model.

Nando Farchmin
Physikalisch-Technische Bundesanstalt, Abbestr. 2-12, 10587 Berlin, e-mail: nando.farchmin@ptb.de

Paul Hagemann
Technische Universität Berlin, Strasse des 17. Juni 135, 10623 Berlin,

Sebastian Heidenreich
Physikalisch-Technische Bundesanstalt, Abbestr. 2-12, 10587 Berlin

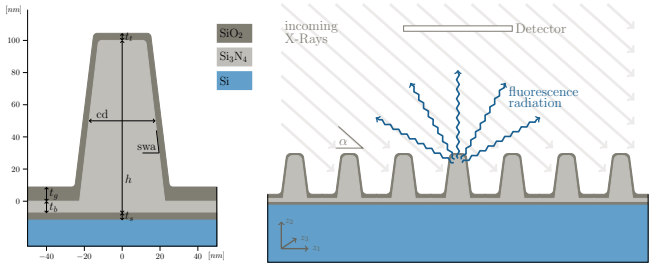


Fig. 1 Left: Cross-section of one grating line with characterizing parameters. Right: Cross-section of the grating with incoming X-rays and emitted fluorescence radiation.

Measure Transport via Invertible Neural Networs

A common approach to recover the distribution of the parameters are MCMC based algorithms. Instead, in this paper, we use invertible neural networks (INNs) [Dinh et al., 2017] within the general concept of transport maps [Marzouk et al., 2016]. This means that we sample from a reference distribution and seek a diffeomorphic transport map, or more precisely its approximation by an INN, which maps this reference distribution to the problem posterior. This approach has some advantages over standard MCMC–based methods. First, given a transport map, which is computed in an offline step, the generation of independent posterior samples essentially reduces to sampling the freely chosen reference distribution. Additionally, observations indicate that learning the transport map requires less time than generating a sufficient amount of independent samples via MCMC. Second, although the transport map is conditioned on a specific measurement, it can serve as a good initial guess for the transport related to similar measurements or as a prior in related inversion problems. Hence the effort to find a transport map for different runs within the same experiment reduces drastically.

Citations and References

References

[Dinh et al., 2017] Dinh, L., Sohl-Dickstein, J., and Bengio, S. (2017). Density estimation using real NVP. In *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings*.

[Hoenicke et al., 2020] Hoenicke, P., Andrle, A., Kayser, Y., Nikolaev, K., Probst, J., Scholze, F., Soltwisch, V., Weimann, T., and Beckhoff, B. (2020). Grazing incidence-x-ray fluorescence for a dimensional and compositional characterization of well-ordered 2d and 3d nanostructures. *Nanotechnology*, page 505709.

[Marzouk et al., 2016] Marzouk, Y., Moselhy, T., Parno, M., and Spantini, A. (2016). Sampling via measure transport: An introduction. *Handbook of Uncertainty Quantification*, page 1–41.

Classical and Bayesian optimization for efficient experimental designs in metrology.

Blaza TOMAN

National Institute of Standards and Technology

USA

Experiments in metrology are often constrained in time or in other ways that make standard experimental designs such as complete factorial or randomized block designs unsuitable. In such situations optimal experimental design techniques which are rooted in classical or Bayesian decision theory, together with some of the usual tools of classical design such as blocking and randomization, can be used to obtain efficient, fit for purpose designs. Bayesian methods are particularly useful when there is related information based on prior experimental studies or on expert opinion which can be utilized to produce a more efficient and useful design.

These useful experimental design techniques, both classical and Bayesian, will be discussed in full detail and applied to specific examples of metrology experiments such as toxicity assays studied using a 96 well plate, sequential binary testing of nuclear detectors, two-phased experiments using mass spectrometry, and linked Key Comparisons.

A knowledge-based evaluation of measurement non-repeatability

Carlo Carobbi

Key words: measurement non-repeatability, type A evaluation of standard uncertainty, knowledge-based probability density function, Bayesian inference, conjugate prior.

1. Measurement non-repeatability and its evaluation

Measurement non-repeatability quantifies the variability of measurement results obtained under specified repeatability conditions. Measurement non-repeatability is an essential contribution to measurement uncertainty in every field of experimental activity.

In the context of testing and calibration if a stable item is re-tested or re-calibrated, the new measurement results are expected to be compatible with the old ones. Two distinct operators should provide compatible measurement results when testing or calibrating the same item. Measurement non-repeatability is then a reference for qualification of personnel. Monitoring measurement non-repeatability contributes to assuring the validity of test and calibration results. In an accreditation regime [ISO/IEC 17025], measurement non-repeatability must be kept under statistical control. Periodic assessments are carried out by the accreditation body aimed at verifying, through an appropriate experimental check, the robustness of the estimate of measurement repeatability [SINAL DT-0002/6].

The Guide to the expression of Uncertainty in Measurement (GUM) [GUM] provides type A evaluation of standard uncertainty as the tool to quantify measurement non-repeatability. Type A evaluation is based on a frequentist approach, thus implying that information on the quality of the estimate of measurement uncertainty must be conveyed to the user. This is done in terms of effective degrees of freedom. The GUM Supplement 1 (GUMS1), devoted to numerical methods for quantification of measurement uncertainty, adopts a knowledge based (in contrast to frequentist) approach to model measurement non-repeatability. The quality of the estimate of measurement uncertainty is accounted for by the available prior knowledge, which eventually determines the width of the coverage interval.

The GUM and the GUMS1 approaches are inconsistent. They produce substantially different results when measurement non-repeatability is a significant contribution to measurement uncertainty and the number of measurements used for its estimate is low [Bich (2012)].

The use of numerical methods for professional (accredited) evaluation of measurement uncertainty is expected to increase in the future. Indeed, the GUMS1 numerical method, which is based on the propagation of probability distributions, accounts for possible non-linearity of the measurement model, is simple, less prone to mistakes (partial derivatives are not required), provides all the available information about the measurand in terms of its probability distribution. Further, the use of numerical methods is practically unavoidable when the measurement model is complex and/or the measurand is an ensemble of scalar quantities (vector). On the extreme the analytical method (based on the law of propagation of uncertainty) is consolidated and the one predominantly adopted nowadays. A further point of strength of the analytical method is its great pedagogical value.

The Working Group 1 of the Joint Committee for Guides in Metrology, responsible for the maintenance of the GUM and its supplements, did not yet identify a way out of the inconsistence between the GUM and the GUMS1. A drawback deriving from direct transposition of the GUMS1 approach to the GUM approach, is that calculation of the spread of measured values about the mean would require at least four measurements (instead of two, as for calculation of sample standard deviation). This is generally not acceptable (e.g., in destructive testing), particularly if implemented as a standard (mandatory) method.

In fact, in the GUMS1 theoretical framework the knowledge-based approach has not been employed for quantification of measurement non-repeatability. The standard probability density function associated with the sample of measured values is indeed a Student's *t* probability density function derived through Bayesian inference assuming a normal probability model and *no prior knowledge about measurement non-repeatability*. Non-informative priors are indeed assigned to both the mean and the variance.

It should be borne in mind that an evaluation of sample standard uncertainty based on fewer than four measurements represents a poor estimate of measurement non-repeatability, regardless of being based on the GUM approach or on the GUMS1 approach. More directly, it is not a matter of dispute between frequentists and Bayesians.

What is proposed here is a knowledge-based approach to the evaluation of measurement non-repeatability. Prior information about measurement non-repeatability may be available from a model of the sources of variability in the measurement chain, from experience gained in the use of similar measurement chains (especially in the context of standard test methods under a statistical control regime), from interlaboratory comparisons (if systematic effects are negligible). As in the GUMS1, use is made in this work of Bayesian inference since it provides a straightforward method to incorporate prior knowledge. To obtain analytical results, useful in the framework of the law of propagation of uncertainty, a normal probability model is assumed with a non-informative prior for the mean and a conjugate prior for the variance.

References

1. ISO/IEC 17025, Conformity Assessment–General Requirements for the Competence of Testing and Calibration Laboratories, Int. Org. Standardization, Geneva, Switzerland (2017).
2. SINAL DT-0002/6, GUIDA AL CALCOLO DELLA RIPETIBILITÀ DI UN METODO DI PROVA ED ALLA SUA VERIFICA NEL TEMPO, Rev. 0, Dicembre 2007.
3. GUM: BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP, and OIML, Guide to the Expression of Uncertainty in Measurement, JCGM 100:2008.
4. GUMS1: BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP, and OIML, Supplement 1 to the 'Guide to the Expression of Uncertainty in Measurement' – Propagation of Distributions Using a Monte Carlo Method, JCGM 101:2008.
5. Walter Bich et al 2012 Metrologia **49** 702.

Optimal designs for hypothesis testing with heteroscedastic experimental groups

Alessandro Baldi Antognini, Rosamarie Frieri, Marco Novelli and Maroussa Zagoraiou

Key words: ANOVA, Ethics, Power, Response-adaptive randomization, Wald test

Comparing the means of several experimental groups is an old and well known problem in the statistical literature which finds applications in many fields. The past decades has seen a flourishing literature about the design of experiments for treatment comparisons that, however, is almost exclusively devoted on estimation precision. Specifically, in the context of the linear homoscedastic model, balancing the allocations among groups is usually considered 'universally optimal', since it optimizes the well-known alphabetical criteria for treatment effects' estimation (see e.g. [3, 5]). However, in the case of several heterogeneous experimental groups, or when inference focuses on treatment contrasts, the balanced allocation could be highly inefficient [1, 4, 7]. Furthermore, in the context of clinical trials, balance could be strongly inappropriate also from an ethical viewpoint since the demand of individual care should lead to assign more patients to the most effective treatment(s).

Reliably detecting the actual treatment difference is a crucial problem in statistical inference however, until recently, little has been done on this topic in the design of experiment literature. Generalizing the results in [1, 4, 6], Baldi Antognini et al. in [2] presented a unified framework for deriving optimal designs for testing the efficacy of several competing treatments which also encompasses the general ANOVA set-up with heteroscedastic errors. Specifically, the ensuing optimal designs are generalized Neyman allocations involving only two experimental groups. Moreover, by considering the ordering of the treatment effects - which is of particular interest in the clinical context for ethical reasons - the authors derive optimal designs maximizing the power of the classical Wald test of homogeneity subject to the ethical constraint that the allocation proportions should reflect the ordering of the treatment efficacies. Albeit in general the treatment effects ordering is a-priori unknown, the proposed allocations are locally optimal designs that can be implemented with response-adaptive randomization procedures provided that a smoothing techniques is properly applied. The advantages of the suggested allocations are illustrated both theoretically and through several numerical examples including nor-

Department of Statistical sciences, University of Bologna, Via Belle Arti 41, 40126, Bologna e-mail: marco.novelli4@unibo.it

mal, binary, Poisson and exponential data (with and without censoring). The comparisons with other designs suggested in the literature confirm that the proposed optimal allocations provide good performance in terms of both statistical power and ethical demands.

References

1. Baldi Antognini, A., Novelli, M., Zagoraiou, M.: Optimal designs for testing hypothesis in multiarm clinical trials. Stat Meth Med Res, **28**, 3242–3259 (2019)
2. Baldi Antognini, A., Frieri, R., Novelli, M., Zagoraiou, M.: Optimal designs for testing the efficacy of heterogeneous experimental groups. Submitted.
3. Begg, C.B., Kalish, L.A.: Treatment Allocation for Nonlinear Models in Clinical Trials: The Logistic Model. Biometrics, **40**, 409–420 (1984)
4. Frieri, R., Zagoraiou, M. : Optimal and ethical designs for hypothesis testing in multi-arm exponential trials. Stat. Med. In press. (2021)
5. Silvey, S.D.: Optimal Designs. Chapman & Hall, London (1980)
6. Singh, S.P., Davidov, O. (2019). On the design of experiments with ordered treatments. J R Statist Soc Series B, **81** 881900.
7. Tymofyeyev, Y., Rosenberger, W. F., Hu, F.: Implementing optimal allocation in sequential binary response experiments. J Am Stat Assoc, **102**, 224-234 (2007)

In the avantgarde of a reliable methodology for automatic identification of microplastics by micro-ATR-FTIR spectroscopy

Vanessa Morgado[1], Carla Palma[2] and Ricardo Bettencourt da Silva[3]

Keywords: microplastics, micro-ATR-FTIR, examinology, uncertainty, validation

1. Statement

Environment contaminated by micro(plastics) issue is very trending nowadays.

Plastics are ubiquitous and have a wide range of applications due to their versatile properties. Its production reached over 368 million tonnes worldwide and 57.9 million tonnes in Europe in 2019 [Birley(1988), PlasticsEurope(2019)]. However, plastic materials are somehow discarded into aquatic environments, becoming responsible for over 62% of the marine litter's composition worldwide [Tekman(online reference)].

The scientific community has become interested in monitoring the presence of these contaminants in different matrices, namely, surface water, column water, seafloor sediment, and beaches. The impact of microplastics in open ocean, rivers, estuarine areas, and coastal regions compartments is only possible to understand if this contamination is characterized adequately and objectively regarding the physical and chemical properties (i.e. polymer type) of particles.

[1] Vanessa Morgado
Instituto Hidrográfico, R. Trinas 49, 1249-093 Lisboa, Portugal;
Centro de Química Estrutural, Faculdade de Ciências, Universidade de Lisboa, Campo Grande, 1749-016 Lisboa, Portugal, e-mail: vmmorgado@fc.ul.pt
[2] Carla Palma
Instituto Hidrográfico, R. Trinas 49, 1249-093 Lisboa, Portugal, e-mail: carla.palma@hidrografico.pt
[3] Ricardo Bettencourt da Silva
Centro de Química Estrutural, Faculdade de Ciências, Universidade de Lisboa, Campo Grande, 1749-016 Lisboa, Portugal: rjsilva@fc.ul.pt

1.1 Microplastics ID

The type of polymer is largely identified by Fourier-Transformed-Infrared spectroscopy, FTIR, where the acquired infrared, IR, spectrum works as a molecular fingerprint of the plastic. Identifying microplastics from the IR spectrum can be a challenge in some cases, especially when a biofilm covers the particles or whenever spectral inconsistencies appear due to differences in plastic additives and copolymers, ageing, or other coating types.

This work aimed at developing and validating a methodology towards the automatic identification of microplastics by micro-Attenuated Total Reflectance-FTIR, micro-ATR-FTIR, overcoming the complexity and time-consuming of a manual interpretation of characteristic spectral bands. The automatic identification of the IR spectra can be supported on a fast mathematical comparison between the unknown microparticle and reference spectra using an agreement index such as correlation, which is more commonly used in routine analysis than other more complex classifications methods. This work describes the development and validation of the use of different correlation algorithms for spectra correlation based on the modelling of their distribution by the Bootstrap method. Particle spectrum is initially checked for spectral contamination by biofilm and the level of attenuation of characteristic spectral bands of the polymer, leading to a removal of about 26% of the polymer spectra. The methodology was implemented in a user-friendly *MS-Excel* spreadsheet to define and validate statistically sound criteria for accurate and automatic identification of microplastics with a true positive rate, *TP*, not lower than 95%, and a false positive rate, *FP*, not greater than 5%. Different data sets of polymers were used to define and validate the identification criteria. The quality of the identification was expressed as an identification uncertainty quantified by the Likelihood Ratio, *LR*(+), which combines the *TP* and *FP* in a single metric (*TP*/*FP*). Considering the defined criteria for the *TP* and *FP*, the analysis fitness for purpose can be controlled by assessing if the *LR*(+) is greater than 19 [Morgado(2021)]. For instance, assuming positive and negative results are equally likely, a *LR*(+) of 19 indicates that the identification is 19 times more likely true than false.

The methodology for identifying microplastics with adequate uncertainty was successfully applied to the identification of Polyethylene, PE, Polyethylene terephthalate, PET, Polypropylene, PP, and Polystyrene, PS, microparticles from sediments collected in Portuguese rivers, improving the preliminary results on the polymer type identification already reported [Gomes(2020)].

References

1. Birley, A.W., Heath, R.J., Scott, M.J.: Plastics Materials, Properties and Applications. Springer Science+Business Media, New York (1988)
2. Morgado, V., Palma, C., Bettencourt da Silva, R.J.N.: Microplastics identification by infrared spectroscopy – Evaluation of identification criteria and uncertainty by the Bootstrap method. Talanta **224**, 121814 (2021). DOI: 10.1016/j.talanta.2020.121814
3. PlasticsEurope Association of Plastics Manufactures: Plastics – the Facts 2020 An analysis of European plastics production, demand and waste data (2019) Available via http://www.plasticseurope.org

4. Tekman, M.B., Gutow, L., Macario, A., Haas, A., Walter, A., Bergmann, M.: Alfred-Wegener-Institut Helmholtz-Zentrum für Polar- und Meeresforschung. Litterbase. Online portal for marine litter available via: https://litterbase.awi.de/
5. Gomes, G.B.; Morgado, V.; Palma, C.: Preliminary Data on Polymer Type Identification from Estuarine Environmental Samples. In: Cocca, M.; Pace, E.D.; Errico, M.E.; Gentile, G.; Montarsolo, A.; Mossotti, R.; Avella, M. (eds.) Proceedings of the 2nd International Conference on Microplastic Pollution in the Mediterranean Sea., pp. 170-174. Springer, Switzerland, (2020).

Uncertainty of thermodynamic properties available via online data banks: Vapor pressure as case study

Lecuna, M.[1] and Sassi, G.[2]

Key words: uncertainty evaluation, montecarlo simulation, vapor pressure data

Online databanks are a convenient source of information whenever thermodynamic properties are required for calculations, in both scientific and engineering environment. The data listed in these repositories is usually reported along with a range of validity and their sources. The values of the thermodynamic properties can usually be retrieved from databases and handbooks, in the form of tables, curves or correlations with regression parameters. Most of the times, no information is given on the uncertainty of the reported estimates, let alone regression parameters in the case of correlations [Dong (2005)]. Evaluating the sources of uncertainty in any model is fundamental to verify the significance of the results for a specific application. Knowing the uncertainty, makes the difference between trusted values and random values. Depending on the model where the property is required and its application (the use of results), the contribution to the total uncertainty of thermodynamic properties and constants can surpass the contributions of other experimental input quantities. In this work, the methods described in the "Guide for the expression of uncertainty in measurement" [JCGM (2008)] are used to evaluate the uncertainty of a thermodynamic property of a pure substance, calculated based on the constants and references reported in two important online data banks (NIST and Dortmund Databank) [NIST (2017)] [Dortmund (2016)]. The vapor pressure is considered as case study and several definitions with their relevant sources of uncertainty are presented.

The aim of the work is to highlight the limitations of properties data commonly used in engineering estimations, as well as the importance of accounting for their uncertainty. Following established metrological guidelines, a full procedure have been developed, with particular considerations aimed to account for the 'quality' of the realization of the definition of the measurand as a source of uncertainty of the property.

[1] Maricarmen Lecuna
Politecnico di Torino, Corso Duca degli Abruzzi 24, e-mail: maricarmen.lecuna@polito.it
[2] Guido Sassi
Politecnico di Torino, Corso Duca degli Abruzzi 24, e-mail: guido.sassi@polito.it

1. Data and Methodology

Antoine Equation model was considered as measurand equation based on the availability of data in many databases and databanks. Monte Carlo algorithm was implemented to estimate the uncertainty of the regression parameters and the coefficient of determination was introduced into the uncertainty analysis to account for the suitability of the model. Finally, a comparison was performed, between the results of vapor pressure estimated with databanks correlations and the results obtained with regressions over different sets of raw experimental data (retrieved from the databanks references). The Antoine equation parameters of five substances, namely acetone, acetonitrile, ethanol, butanol and methanol, were listed from NIST and Dortmund databanks and the references associated to them were identified. A total of ten data sets were identified, with articles dating back to 1926.

Both the linearized and non-linearized equations were used to estimate the parameters of regression: A, B and C constants. Montecarlo simulations have been performed to estimate the probability distributions of each regression parameter. After obtaining the uncertainty of each parameter, an uncertainty budget of the vapor pressure can be built. Besides the regression parameters, the temperature is the only influence quantity. An additional influence quantity was introduced in the budget to account for the suitability of the model to the experimental data, i.e., the realization of the definition of the measurand as souce of uncertainty. The suitability of the model was estimated as function of the coefficient of determination of the models obtained in Montecarlo simulations.

The sensitivity to the dataset was analysed performing the same procedure with different sets of data. The obtained results were similar to those reported by the databanks in terms of precision, however differences in the uncertainty were observable based on the number and distribution of the available experimental points. The form of the equation (linearized or non linear) was observed to play an important role in the increase of the uncertainty of the property.

With this work we evidence the importance of relying on experimental data and raise awareness of the impact that introducing correlation constants could have on results, if a proper quantification of uncertainty is omitted. The use of a case study illustrates how datasets could be improved just by applying formal metrological procedures to already available raw data.

References

1. Dong,Q, Chirico, R.D., Yan, X., Hong, X., Frenkel, M. Uncertainty reporting for experimental thermodynamic properties. In: Journal of Chemical and Engineering Data, vol. 50, no. 2. pp. 546–550 (2005).
2. Dortmund Data Bank Available at:, www.ddbst.com (2016).
3. Joint Committee for Guides in Metrology (JCGM), Evaluation of measurement data: Guide to the expression of uncertainty in measurement. no. September, p. 120 (2008).
4. National Institute of Standards and Technology, NIST Chemistry WebBook, NIST Standard Reference Database Number 69. Gaithersburg MD (2018)

Evaluating erosion performance of cold-sprayed coatings by Design of Experiments

Elisa Verna[1], Roberto Biagi[2], Marios Kazasidis[3], Maurizio Galetto[1], Edoardo Bemporad[2], and Rocco Lupoi[3]

Key words: Cold spray, Erosion, Design of Experiments

Extended abstract

Solid Particle Erosion (SPE) occurs when solid particles dragged by fluid means impinge on the surface of pneumatic or hydraulic systems resulting in loss of their mass. Multiple potential solutions to the problem have been investigated, including the modification of the machinery design and the use of filtration systems. From the perspective of materials engineering, the alleviation of SPE in metals has been widely encountered with the fabrication of coatings [1], using methods such as laser cladding, plasma spray, high-velocity oxyfuel and cold spray (CS) [2]. CS is a low-temperature particle deposition process in which microscale powder particles are accelerated to a high velocity (300 to 1200 m/s) through a De-Laval nozzle dragged by a high-pressure propelling gas (commonly nitrogen or helium) toward a target substrate [3]. The advantages of this powerful technique over other thermal spray processes include minimizing potential phase changes, preserving the original feedstock properties, and preventing coating defects, such as surface oxides and other inclusions.

[1] Elisa Verna, Maurizio Galetto
DIGEP, Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Torino, Italy, e-mail: elisa.verna@polito.it, maurizio.galetto@polito.it

[2] Roberto Biagi, Edoardo Bemporad
Mechanical and Industrial Engineering Department, University of Rome "ROMA TRE", Via Vasca Navale 79, 00146 Rome, Italy, e-mail: r.biagi90@gmail.com, edoardo.bemporad@uniroma3.it

[3] Marios Kazasidis, Rocco Lupoi
Department of Mechanical and Manufacturing Engineering, Trinity College Dublin, The University of Dublin, Parsons Building, Dublin 2, Ireland, e-mail: kazaside@tcd.ie, lupoir@tcd.ie

Amongst nickel-based superalloys, Inconel 718 (IN718) is commonly used in applications where stability at elevated temperatures, high long-time creep strength, and corrosion resistance in aggressive environments are required [3].

The complex nature of the SPE mechanism involves several variables related to the erodent properties (e.g., feed rate, velocity, size, hardness, impingement angle) and variables related to the intrinsic properties of the coating materials. As a result, identifying a standard method for predicting the erosion performance of coatings is extremely difficult. To this end, one of the most widely adopted techniques to assess the effect of control variables on coating erosion resistance is the Design of Experiments (DoE) [4], [5]. To date, few empirical studies have investigated the erosion performance of cold spray coatings through experimental design [6], which is a major innovative aspect of this study.

The authors adopted a general full factorial design to investigate the SPE resistance of cold sprayed IN718-Ni composite coatings. The coatings were deposited onto substrates via high-pressure CS using nitrogen as the propulsive gas, with inlet pressure at 3.0 MPa and gas temperature at 1000 °C. The effect of the impingement angle (°), the erodent size (μm) and the erodent feed rate (g/min) on the erosion rate (mg/min) of the coating was investigated.

Empirical results showed that the erodent feed rate and the impingement angle produced a statistically significant effect on the erosion rate, as well as the interaction between erodent size and impingement angle. Furthermore, a mathematical model relating such variables with the erosion rate was identified by stepwise regression analysis. This model effectively allows predicting the erosion performance of real components deposited with IN718-Ni coating under working conditions and properly designing the in-service conditions of new cold-sprayed components. The set of parameters optimizing the erosion rate was accordingly derived by performing an optimization, and then validated. Finally, the worn surfaces of the coatings were observed by SEM in order to understand the erosion mechanism.

References

1. Patnaik, A., Satapathy, A., Chand, N., Barkoula N. M., Biswas S., Solid particle erosion wear characteristics of fiber and particulate filled polymer composites: A review, Wear, 268(1), 249–263 (2010)
2. Alidokht, S., Vo, P., Yue, S., Chromik, R. R., Erosive wear behavior of Cold-Sprayed Ni-WC composite coating, Wear, 376–377, 566–577 (2017)
3. R. Biagi, E. Verna, E. Bemporad, M. Galetto, and R. Lupoi, Cold spraying of IN 718-Ni composite coatings: microstructure characterization and tribological performance, In: Materials Science Forum, Vol. 1016, pp. 840-845, Trans Tech Publications Ltd (2021)
4. Montgomery, D. C., Design and analysis of experiments, 9th Ed. New York: John Wiley & Sons, (2017)
5. Verna, E., Biagi, R., Kazasidis, M., Galetto, M., Bemporad, E., Lupoi, R., Modeling of erosion response of cold-sprayed In718-Ni composite coating using full factorial design, Coatings, 10(4), 335 (2020)
6. Hu, H.X., Jiang, S.L., Tao, Y.S., Xiong, T.Y., Zheng, Y.G., Cavitation erosion and jet impingement erosion mechanism of cold sprayed Ni-Al2O3 coating, Nucl. Eng. Des., 241(12), 4929–4937 (2011)

Monte Carlo bottom-up evaluation of the uncertainty of complex sample preparation: Elemental determination in sediments

Vanessa Morgado[1], Carla Palma[2] and Ricardo Bettencourt da Silva[3]

Key words: Measurements in chemistry; Bottom-up evaluations; Sample preparation uncertainty; Monte Carlo Method; Sediments

1. Introduction

Many chemical analyses involve a complex sample preparation and some, based on an instrumental method of analysis such as spectrometric and chromatographic methods, are affected by matrix effects. The objective interpretation of the results of these analyses, performed in the framework of a research or of a conformity assessment, requires the quantification of the measurement uncertainty. Top-down assessments of the measurement uncertainty are known to involve the oversimplification of the measurement process and a pessimist quantification of the uncertainty [Silva (2006)]. This work presents a novel methodology for the bottom-up modelling of the performance of complex analytical operations, such as sample digestion or extraction, by the Monte Carlo simulation of their performance independently of the performance of the other analytical steps [Morgado (2020)].

[1] Vanessa Morgado
Instituto Hidrográfico, R. Trinas 49, 1249-093 Lisboa, Portugal;
Centro de Química Estrutural, Faculdade de Ciências, Universidade de Lisboa, Campo Grande, 1749-016 Lisboa, Portugal, e-mail: vanessa.morgado@hidrografico.pt
[2] Carla Palma
Instituto Hidrográfico, R. Trinas 49, 1249-093 Lisboa, Portugal, e-mail: carla.palma@hidrografico.pt
[3] Ricardo Bettencourt da Silva
Centro de Química Estrutural, Faculdade de Ciências, Universidade de Lisboa, Campo Grande, 1749-016 Lisboa, Portugal: rjsilva@fc.ul.pt

2. Methodology

The developed modelling requires the replicate analysis of items, where at least some should have a known reference value, in n different days (e.g. $n = 10$). The simulation of m measured values (e.g. $m = 16000$), considering all analytical steps except the sample preparation, of the analysis of each sample in the n days allows the determination of ($n \times m$) differences between n sample means with m simulations of between-days means. These differences allow simulating the between-days precision component of the measurement error, E_b, and the mean analyte recovery, $\bar{R}$, from the analysis of reference materials. The E_b and $\bar{R}$ from the analysis of various items are pooled simulating the complex probability density of these uncertainty components. For the post-validation analysis of unknown samples, the simulated uncertainty of all analytical steps except sample preparation is merged with the simulated E_b and $\bar{R}$, and results corrected for the $\bar{R}$ reported as relevant percentile intervals or the distribution of simulated measured values.

3. Results

The developed methodology was successfully applied to the determination of total or acid-extractable As (following OSPAR or EPA 3051A methods, respectively) [OSPAR (2015), EPA (2007)] in sediments where E_b was simulated from the analysis of one Certified Reference Material, CRM, and three sediment samples and $\bar{R}$ simulated from the analysis of the CRM and two spiked samples. The evaluated uncertainty is fit for environmental monitoring considering performance criteria defined for Quasimeme proficiency tests [Eurachem (2015)]. The developed measurement models were successfully cross-validated by randomly extracting data from the validation set subsequently used to check the compatibility between estimated and reference values for 95% or 99% confidence level. The observed success rate of these assessments is compatible with the confidence level of the tests.

References

1. EPA: Method 3051A – Microwave Assisted Acid Digestion of Sediments, Sludges, Soils and Oils, EPA, USA (2007)
2. Eurachem/CITAC Guide: Setting and Using Target Uncertainty in Chemical Measurement, Eurachem (2015) (ISBN 978-989-98723-7-0).
3. Morgado, V., Palma, C., Silva, R. B.: Monte Carlo Bottom-up Evaluation of Global Instrumental Quantification Uncertainty: Flexible and user-friendly computational tool. Chemosphere 258, 127285 (2020)
4. OSPAR Commission: JAMP Guidelines for Monitoring Contaminants in Sediments, Agreement Ref. no. 2002-16, OSPAR (2015)
5. Silva, R. B., Santos, J. R., Camões, M. F.: A New Terminology for the Approaches to the Quantification of the Measurement Uncertainty. Accred. Qual. Assur. **10**, 664-671 (2006)

Uncertainty expression by finite information quantities

Luca Callegaro, Francesca Pennecchi and Walter Bich

Key words: uncertain digits, finite information quantities

1 Introduction

Concepts like "certain" and "uncertain" digits of a quantity value are immediate in the mind of the experimenter, for example when writing a measurement outcome from a digital instrument. In the GUM framework the state of knowledge about a measurand *i.e*, a real-valued measurement result, is expressed by a probability distribution (PDF).

In most cases, only location and scale parameters of the PDF (typically, the mean and the standard deviation) are given, truncated to a finite precision, by rounding "to the significant digits".

Here we discuss a recently proposed alternative to the expression of uncertainty of a quantity value directly in terms of the degree of certainty of its digits.

2 Finite information quantities

In [Del Santo and Gisin, 2019], the authors introduce the concept of *finite information quantities* (FIQ) in an epistemological framework. In short, a FIQ ranging in

Luca Callegaro
Istituto Nazionale di Ricerca Metrologica (INRIM), Torino, Italy, e-mail: l.callegaro@inrim.it

Francesca Pennecchi
Istituto Nazionale di Ricerca Metrologica (INRIM), Torino, Italy, e-mail: f.pennecchi@inrim.it

Walter Bich
Istituto Nazionale di Ricerca Metrologica (INRIM), Torino, Italy, e-mail: w.bich@inrim.it

the interval $[0,1]$ is expressed by the binary number $Q = 0.Q_1Q_2Q_3\ldots$, where the individual bits Q_k are independent Bernoulli random variables; q_k is the probability of the case $Q_k = 1$. The probabilities are limited to rational numbers, which have a finite information content. A FIQ Q is thus specified by

$$\boldsymbol{q} = [q_1, q_2, \ldots, q_k, \ldots, q_M, \tfrac{1}{2}, \tfrac{1}{2}, \ldots], \tag{1}$$

the vector of probabilities of its bits Q_k; it is assumed that $q_k = \frac{1}{2}$ for $k > M$, *i.e.*, all bits beyond position M have a 50 % probability of being either 0 or 1 and therefore carry no information. Only a finite number M of probabilities q_k are needed to completely specify Q. Calculations on FIQs can be achieved with proper combinatorial logic units; in [Callegaro et al, 2020] we proposed the algorithm for adding FIQs and multiplying by a numerical constant, which are sufficient operations for linear measurement models.

3 Limitations

The FIQ concept, as per definition (1), is incomplete, because any arithmetical operation on a FIQ, such as the change of measurement unit, destroys independence among the bits Q_k [Callegaro et al, 2020]. As a consequence, the description of a quantity value in therms of a FIQ is incomplete unless the dependencies among the bits Q_k are also taken into account.

Whether this limitation can be overcome is still unclear. A possible way to extend FIQs including dependencies among the bits is to express the FIQs in the form of a binary decision tree. It has been shown [Lewis, 1975] that such a structure can properly map to a PDF with an arbitrary degree of numerical precision. What is still missing is a way to perform arithmetical calculations with binary trees, which is presently under investigation.

References

Callegaro L, Pennecchi F, Bich W (2020) Comment on "physics without determinism: Alternative interpretations of classical physics". Phys Rev A 102:036,201

Del Santo F, Gisin N (2019) Physics without determinism: Alternative interpretations of classical physics. Phys Rev A 100:062,107

Lewis TG (1975) Distribution sampling for computer simulation. Lexington Books, Lexington, MA, USA, chapter 8

A GUI for Bayesian sample size determination

Jörg Martin[1*], Clemens Elster[*]

Key words: Sample size determination, experimental design, Bayesian statistics

1. Overview

We introduce a graphical user interface (GUI) that is open source, written in Python, easy to use and allows the user to perform sample size determination based on solid Bayesian statistics without having any foreknowledge in the latter. The computation of the sample size is based on a Bayesian sample size criterion that was developed by the authors in [Martin(2020)].

2. Motivation and outline of the talk

Taking measurements can be a costly and time-consuming activity where the financial and temporal cost is usually directly proportional to the length of the measurement series. Moreover, repeating a measurement series that failed to produce a result with a sufficient accuracy can even be more costly and is sometimes not even possible. For this reason, the determination of an optimal length of the measurement series, known as sample size determination, is one of the most important subjects of experimental design [Desu(2012)].

Despite its basic importance, however, sample size determination has hardly received much attention within metrology. As experimental design is usually done before any measurements are taken, it usually must be based on some sort of foreknowledge [Santis(2007)]. Bayesian statistics is a consistent mathematical framework for such purposes but learning it might be deterrent for someone who only wants to find a sample size.

The presented GUI tries to bridge this gap between theory and practice and enables users to use Bayesian sample size planning without any need of knowledge of Bayesian statistics [Martin(2021)]. Moreover, it is the first implementation of a criterion that was previously introduced by the authors as the variation of the posterior variance criterion (VPVC) in [Martin(2020)] and leads to a more reliable sample size planning than previous sample size criteria [Wang(2002)].

The talk splits in two parts. In the first part we introduce the underlying VPVC and demonstrate along an example that it performs more reliable than previous

[1] e-mail: joerg.martin@ptb.de

[*] Physikalisch-Technische-Bundesanstalt, Abbestr. 2, 10587 Berlin

criteria. In the second part, the usage of the GUI together with the elicitation of prior knowledge is demonstrated along the same example.

References

1. M. Desu, Sample size methodology. Elsevier, 2012, Amsterdam.
2. J. Martin and C. Elster The variation of the posterior variance and Bayesian sample size determination. Statistical Methods & Applications, 1613-981X, 2020.
3. J. Martin and C. Elster GUI for Bayesian sample size planning in type A uncertainty evaluation. Measurement Science and Technology, 2021, in press.
4. De Santis, Fulvio. Using historical data for Bayesian sample size determination. Journal of the Royal Statistical Society: Series A (Statistics in Society) 170.1 (2007): 95-113.
5. F. Wang and A. Gelfand. A simulation-based approach to Bayesian sample size determination for performance under a given model and for separating models. Statistical Science 17.2 (2002): 193-208.

Discriminant Analysis of Vegetable Oils by TGA-GC/MS Combined with Chemometrics and Data Fusion without Sample Pretreatment

Xia Zhou[1], Xiaoting Chen, Xiuqin Li, Chaonan Han, Qinghe Zhang *

In this study, volatile oxidation product profiles of six vegetable oils (soybean, rapeseed, peanut, sunflower, olive and camellia) were established by thermogravimetric-gas chromatography/mass spectrometry (TGA-GC/MS) and further subjected to statistical analyses by using both unsupervised (principal component analysis) and supervised (principal component analysis-linear discriminant analysis) chemometrics techniques. The results indicated that volatile oxidation product profiles could discriminate all of the studied vegetable oils clearly except the slight overlap of olive oil and camellia oil. In order to improve the discrimination results, data fusion was applied by combining GC/MS dataset and TGA dataset with ratio modulation. At a weight ratio of 0.9: 0.1 (GC/MS to TGA), the prediction ability of the classification model was 96.2%, which can effectively distinguish these vegetable oils. The method established in this work is simple, rapid, and need no sample pretreatment, which provides a new idea for vegetable oil authentication and quality assessment.

Key words: vegetable oil, TGA-GC/MS, chemometrics, data fusion, classification

1. Introduction

Due to the wide variety of vegetable oils available in supermarkets, authentication of vegetable oils is of growing concern in food quality and safety. Recent advances in mass spectrometry together with improvements in chemometric tools have allowed non-target analysis to be developed in food authentication (Sales, 2017). Currently, one of the critical challenges in non-target approach based on mass spectrometry is that the sample pretreatment like extraction or purification is usually required to be compatible with mass spectrometer (Cavanna, 2018). Ideally, in order to obtain more informative results, no sample pretreatment would be preferred for the analysis of samples.

The main objective of the present work was to develop a non-target classification method for vegetable oils from six different botanical origins (soybean, rapeseed, peanut, sunflower, olive and camellia) based on thermogravimetric-gas chromatography/mass spectrometry (TGA-GC/MS).

[1] Xia Zhou
Xia Zhou, Division of Chemical Metrology and Analytical Science, National Institute of Metrology, Beijing, 100029, China, e-mail: zhouxia@nim.ac.cn

2. Results and discussion

From the GC/MS spectra, 15 volatile oxidation products were identified by matching with mass spectra of commercial mass spectral library (NIST), including alkanals (hexanal, heptanal, octanal, nonanal), alkenal (2-heptenal), ketones (2-pentanone, 2-heptanone, and 2-octanone), alkanes (heptane, octane, nonane, decane, undecane), alkene (1-octene) and 2-butyltetrahydrofuran. PCA analysis of GC/MS dataset showed that different vegetable oils were clearly classified into five groups mainly according to their botanical origin, with overlapping being observed for olive oils and camellia oils, which may be due to their similar composition of fatty acid (Shen, 2020). The prediction rate of PCA-LDA model based on GC/MS dataset was 92.5%. Finally, 2 olive oils were falsely recognized as camellia oils; 2 camellia oils were falsely recognized as olive oils.

In order to improve the discrimination results, data fusion was applied by combining GC/MS dataset and TGA dataset with ratio modulation. At a weight ratio of 0.9: 0.1 (GC/MS to TGA), the prediction rate of the classification model was 96.2%, with 1 olive oil being falsely recognized as camellia oil and 1 camellia oil being falsely recognized as olive oil. Figure 1 shows the LDA analysis results of GC/MS dataset and data fusion of GC/MS and TGA.

The results demonstrated that TGA-GC/MS combined with chemometrics and data fusion could effectively discriminate vegetable oils from six different botanical origins. The method established in this work is simple, rapid, and need no sample pretreatment, which provides a new idea for vegetable oil authentication and quality assessment.

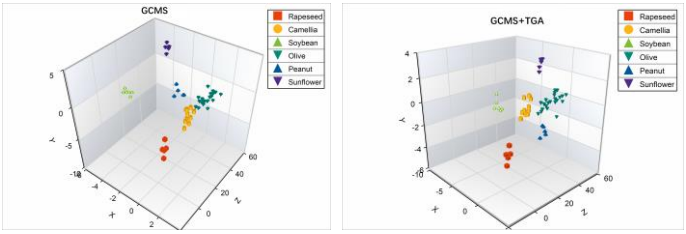


Figure 1: LDA analysis results of GC/MS dataset and data fusion of GC/MS and TGA.

References

1. Cavanna, D., Righetti L., Elliott C., Suman, M.: The scientific challenges in moving from targeted to non-targeted mass spectrometric methods for food fraud analysis: A proposed validation workflow to bring about a harmonized approach. Trends Food Sci. Technol. **80**, 223--241 (2018)
2. Sales, C., Cervera, M. I., Gil, R., Portolés, T., Pitarch, E., Beltran, J.: Quality classification of spanish olive oils by untargeted gas chromatography coupled to hybrid quadrupole–time of flight mass spectrometry with atmospheric pressure chemical ionization and metabolomics-based statistical approach. Food Chem., **216**, 365--373 (2017)
3. Shen, M., Zhao, S., Zhang, F., Huang, M., Xie, J.: Characterization and authentication of olive, camellia and other vegetable oils by combination of chromatographic and chemometric techniques: role of fatty acids, tocopherols, sterols and squalene. **247**, 411--426 (2021)

Approximating Gaussian Process regression models using banded matrices

Kavya Jagan and Alistair Forbes

Key words: Gaussian Processes, uncertainty evaluation, banded matrices

Gaussian Process (GP) regression [4] is a flexible way of modelling data that is not limited by having to specify a parametrized functional form. GPs are typically used to model a variable that has spatial (and/or temporal) dependence. GP models are used to describe the values of the variable at different spatial locations as being correlated and describing the strength of the correlation in a flexible way as a function of spatial separation through a covariance kernel. The idea of GP regression is to use data collected at known, fixed locations and the correlation behaviour to predict the response at other locations within an area of interest. A particular advantage of GP regression is that the uncertainties associated with predicted values can also be evaluated.

The correlation structure in data is captured by GPs through a covariance kernel function that depends on hyper-parameters. A realisation of the covariance kernel for a particular data set defines the covariance matrix associated with the data. One of the main challenges of applying GPs in practice is that the covariance matrix becomes too large to store and manipulate for large data sets.

To improve the scalability of GPs and hence their applicability to metrology problems, we explore a number of approximations to the GP covariance kernel. We focus in particular on data that is equally spaced (either spatially or temporally) allowing us to construct approximations of the covariance matrix using banded matrices. We describe using the Yule-Walker equations [4] to obtain banded approximations of the inverse of the covariance matrix, optimising relative entropy [2] to obtain banded approximations of the covariance matrix or its inverse and deriving asymptotic banded approximations of the covariance matrix for time series data [3].

The structure of the approximate covariance matrices can be exploited in the algorithms used to make predictions, for instance by using Fast Fourier Transforms to perform matrix multiplication [1], leading to algorithms that have a complexity of

Kavya Jagan
National Physical Laboratory, Hampton Road, Teddington, TW11 0LW, UK e-mail: kavya.jagan@npl.co.uk

Alistair Forbes
National Physical Laboratory, Hampton Road, Teddington, TW11 0LW, UK e-mail: alistair.forbes@npl.co.uk

$O(m)$ where m is the number of data points, rather than $O(m^3)$ for full matrix approaches, where the "Big-Oh" notation is often used to represent the computational complexity in terms of floating point operations or flops [1]. The methods described will be illustrated using gear surface roughness data sets involving over 20,000 data points.

Acknowledgements This work is supported by the UK Government's Department of Business, Energy and Industrial Strategy (BEIS) through the National Measurement System programme for Data Science.

References

1. Gene H Golub and Charles F Van Loan. *Matrix computations*, volume 3. JHU press, 2013.
2. Robert M Gray. *Entropy and information theory*. Springer Science & Business Media, 2011.
3. J Pasupathy and RA Damodar. The Gaussian Toeplitz matrix. *Linear algebra and its applications*, 171:133–147, 1992.
4. C. E. Rasmussen and C. K. I. Williams. *Gaussian Processes for Machine Learning*. MIT Press, Cambridge, Mass., 2006.

Factors Relating to Gradient Coil and Radio Frequency Induced Heating within Implanted Orthopaedic Devices during MRI

J Wooldridge[1], I Partarrieu[1], M Cox[1], A Arduino[2], L Zilberti[2], U Zanovello[2], M Chiampi[2], V Clementi[3] and O Bottauscio[2]

Key words: MRI, Implant heating, Parametric modelling

1. Introduction

Within the EU, approximately 50 million citizens have a medical implant of some kind [Lidgren (2020)], with a majority requiring a magnetic resonance imaging (MRI) scan during the lifetime of the device [Bhuva (2020)]. Under the EU Medical Device Regulations [MDR (2017)], MRI safety labelling is required for all implants to ensure regulatory compliance, assessing each device through standardised testing procedures. In this work, we aim to identify the implant-related factors that affect tissue heating, for both radio frequency (RF) and gradient coil (GC) switching, where possible providing parametric analyses to enable more efficient and less time-consuming evaluations of every size and shape of device. In this way, we establish a subset of worst-case scenario configurations and investigate the uncertainty in extrapolating the results from simplified geometries to real devices.

2. Methods

GC heating models were constructed using a hybrid finite element-boundary element approach as described in Bottauscio (2015), and the RF heating finite element simulations were performed separately in Comsol Multiphysics v5.5 [Comsol (2019)]. A range of simplified implant geometries and common orthopaedic implants were simulated, alongside investigations into different MRI sequences and a variety of implant and phantom materials properties. For the RF simulations both 1.5 T and 3 T models were constructed.

3. Results

3.1 GC Heating

A series of ten simplified and seven clinical sequences were simulated. An index of thermal stress, I_s, was constructed for each sequence, related to the square of the

[1] J Wooldridge, email: jenny.wooldridge@npl.co.uk; I Partarrieu, email: ignacio.partarrieu@npl.co.uk; M Cox, email: maurice.cox@npl.co.uk
National Physical Laboratory, Hampton Road, Teddington, Middlesex, TW11 0LW, UK
2 A Arduino, email: a.arduino@inrim.it; L Zilberti, email: l.zilberti@inrim.it; U Zanovello, email: u.zanovello@inrim.it; M Chiampi, email: m.chiampi@inrim.it; O Bottauscio, email: O.bottauscio@inrim.it
Istituto Nazionale di Ricerca Metrologica, Str. delle Cacce, 91, 10135 Torino TO, Italy
3 V Clementi, email: valeria.clementi40@gmail.com
IRCCS Istituto Ortopedico Rizzoli, Laboratorio di Tecnologia Medica, Via diBarbiano 1/10, 40136 Bologna, Italy

time derivative of the applied magnetic field. The power P deposited in the implant and therefore the implant heating, was found to scale with I_s^2 for a wide range of implants. P is determined by the amplitude of the eddy currents within the implant, which is related to both the implant volume and its average cross section with respect to the direction of the applied magnetic field. Implant conductivity also directly affects the heating, scaling linearly with electrical conductivity until the skin effect starts playing a role within the metal.

3.2 RF Heating

As with the GC heating, the RF fields induce currents within the implant, resulting in heating caused by associated ohmic losses. However, the heating behaviour in this case is predominantly defined by the "antenna effect" in which heating is maximised in linear conductors where the induced currents form resonant standing waves caused by reflections at the ends of the conductor. In considering the phantom material properties, within the GC simulations the only relevant parameter of interest is the thermal exchange coefficient, which modifies the heat transfer between the implant and phantom. In contrast, within the RF simulations, the phantom electrical conductivity and permittivity directly play a role in defining the resonant length of the linear conductor and therefore the implant heating.

3.3 Parametric Models

Considering the GC heating, separate parametric models were constructed to estimate the temperature rise in 3D implants (e.g., hip, knee and shoulder prostheses, along with spheroids as approximations of complex implants) and 2D implants (orthopaedic plates and cuboids as their simplified counterparts). The models consider the parameters of implant volume, external surface area, aspect ratio, electrical conductivity, amplitude of the GC magnetic field, the sequence thermal stress and the thermal conductivity of the phantom material. Within the RF simulations, multivariate polynomials, represented in terms of Chebyshev series, were fitted to a series of cylinder and cuboid models using least-squares methods, the results of which were used to assess the feasibility of approximating realistic implant geometries with simplified substitutions.

Acknowledgements: This project has received funding from the EMPIR programme co-financed by the Participating States and from the European Union's Horizon 2020 research and innovation programme (grant 17IND01 – MIMAS).

References

1. AN Bhuva, R Moralee, JC Moon, CH Manistry, "Making MRI available for patients with cardiac implantable electronic devices: growing need and barriers to change", European Radiology, 30(3):1378-1384, 2020
2. O Bottauscio, M Chiampi, J Hand and L Zilberti, "A GPU Computational Code for Eddy-Current Problems in Voxel-Based Anatomy", IEEE Transactions on Magnetics, 51(3):1-4, 2015
3. COMSOL Multiphysics© v5.5, www.comsol.com, COMSOL AB, Stockholm Sweden
4. L Lidgren, D Bushan Raina, M Tagil and K Elizabeth Tanner, "Recycling implants: a sustainable solution for musculoskeletal research", Acta Orthopaedica, 91(2):125, 2020
5. MDR, "Regulation (EU) 2017/745 of the European Parliament and of the Council of 5 April 2017 on Medical Devices, amending Directive 2001/83/EC, Regulation (EC) No 178/2002 and Regulation (EC) No 1223/2009 and Repealing Council Directives 90/385/EEC and 93/42/EEC", Official Journal of the European Union L 117/1, 2017 https://eur-lex.europa.eu/eli/reg/2017/745/oj

Interlaboratory comparison of nominal data on macroscopic examination of welds

Tamar Gadrich[1], Ilya Kuselman[2], Ivana Andrić[3]

Key words: nominal data**,** CATANOVA, interlaboratory comparison, welding imperfections, macroscopic examination

1. Developed statistical technique

The nominal property value of a phenomenon, body or substance is a word or alphanumerical code given for identification reasons, where the property has no magnitude. Nominal variables are coded by exhaustive and disjoint classes or categories with no natural ordering. Therefore, nominal data are related to categorical data.

A statistical technique for interlaboratory comparisons of nominal data that are influenced by two independent factors (variables) and their possible interaction was developed based on two-way categorical analysis of variation CATANOVA [Anderson and Landis (1980)]. The technique includes decomposition of the total variation of the nominal data for a cross-balanced design according to two factors and their interaction, as well as according to the categories of the response.

2. An application for the macroscopic examination of welds

The developed technique was applied for the interlaboratory comparison of the macroscopic examination of weld imperfections (the nominal response variable) caused by failures in the welding process, with five categories according to ISO 6520-1:2007. The comparison was organized using 12 photographs of the macrostructures of the welds as test items distributed to three participating laboratories (factor 1) and examined by experienced technicians, as well as by novices (factor 2). Important is that there is not any hierarchy of the categories and/or hierarchy of the factors.

Since the test items in the present study are not samples of a substance, material or a thing but are identical images, there is no question about their chemical or physical homogeneity. The assign value of the testing property and its uncertainty are not objectives in this study, as well as a score of a laboratory proficiency based on a deviation of the laboratory results from the assigned value. The consensus only of the comparison participants is discussed here.

Analysis of the obtained data showed a consensus between laboratories and between technicians and no interaction between them. Therefore, the proficiency of the participants of the comparison can be considered as satisfactory.

It was found that the weld imperfections from two categories (cavities and inclusions) were examined with low variation, while the examination results of imperfections belonging to the three other categories (cracks, lack of fusion, and geometric shape errors) had significantly larger variations, i.e., were much more difficult to identify [Gadrich et al (2020)].

[1] Tamar Gadrich

Tamar Gadrich, ORT Braude College, Department of Industrial Engineering and Management, P.O. Box 78, 51 Snunit St., 2161002 Karmiel, Israel, e-mail: tamarg@braude.ac.il

[2] Ilya Kuselman

Ilya Kuselman, Independent Consultant on Metrology, 4/6 Yarehim St., Modiin, Israel, e-mail: ilya.kuselman@bezeqint.net

[3] Ivana Andrić

Ivana Andrić, Mechanical and Metallographic Laboratory, Department of Welding Testing and Technology (ZIT Ltd), Zagreb, 6 Rakitnica St., 10040 Zagreb, Croatia, e-mail: ivana.andric@zit-zg.hr

The proposed calculations are based on the formulas requiring the simplest mathematical operations and can be performed using a routine Excel sheet. The developed technique is applicable for other nominal properties and can be adjusted for proficiency testing of laboratories participated in the comparison.

References

1. Anderson, R.J., Landis, J.R.: CATANOVA for multidimensional contingency tables: Nominal-scale response. Comm. Statist. Theory Methods **9**, 1191--1206 (1980).
2. Gadrich, T., Kuselman, I., Andrić, I.: Macroscopic examination of welds: Interlaboratory comparison of nominal data. SN Applied Sciences (2020) doi.org/10.1007/s42452-020-03907-4
3. ISO 6520-1:2007. Welding and allied processes - Classification of geometric imperfections in metallic materials, Part 1.

Application of Gaussian Processes to the uncertainty evaluation of humidity measurements looks effective and competitive with respect to standard (linear) methods.

Pietro Colombo[1] Alessandro Fassò[2]

Key words: Gaussian process 1, Linear interpolation 2, Uncertainty estimation 3, GRUAN 4, Humidity profiles 5.

1. Introduction

GCOS Reference Upper Air Network (GRUAN) [Dirksen et al. (2014)] aims to provide long-term, highly accurate measurements of the atmospheric profiles of fundamental climate variables, such as wind, humidity, and temperature. Some of the data are collected using Vaisala RS41 radiosondes launched from different sites around the globe. Since missing values often appear in these vertical strips of climate variables, estimation techniques to fill the gaps are required. One of the main goals of GRUAN is to evaluate measurement uncertainty. This uncertainty evaluation allows for the comparability of different techniques and measures data reliability. There are countless reasons to have highly accurate, reliable, and accountable data of essential climate variables. They include understanding climate changes at different atmosphere levels; quantifying climate trends to characterize the properties of atmospheric columns; calibrating measurement instruments; and providing accountable data to understand the extent to which business opportunity is possible.

1.1 Main objective

The objective of this paper is to understand the interpolation uncertainty of the climate variable "relative humidity". The radiosonde community generally assesses variability by using linear interpolation (LINT). The success of this technique is due to its performance and easy implementation. Here we propose a new approach based on Gaussian Process estimation (GP) to estimate the "missing values" and understand the "uncertainty", in contrast with the long-established procedure of linear interpolation. We demonstrate GP's superior performance and explain why GP should be preferred to LINT for the "relative humidity" assessment.

1.2 Data description

Data provided by Vaisala radiosondes at various GRUAN sites are given as vertical profiles with 1-second resolution. Different processing stages are then applied to this "raw data sensing" to disentangle noise from signal, while a regular grid of measurements is maintained. Measurements gaps in the vertical profiles of measurements occur for various reasons, such as the mere presence of physical obstacles, meteorological conditions, and sensor-related reasons. These gaps may have different lengths and/or altitudes; moreover, they increase in frequency with the horizontal distance from the launch site [Fassò et al. (2020)]. The dataset used in this study is composed of 178 tables of approximately 5,000 observations each, for a total of more than one million measurements. The dataset's variables considered here are: second time from radiosonde launch [s], temperature in Kelvin [K], relative humidity [%], and physical position of the radiosondes along profiles, such as latitude [deg], longitude [deg], and altitude [m]. A gap in one variable might not correspond to a gap in another one. The number of missing values for variable humidity is "only" 739 for the entire data set used, or less than one over thousand. Hence, they are deleted and ignored for the rest of the paper.

2. Methods

This study follows the general estimation procedure proposed by [Fassò et al. (2019)] for temperature profiles; however, results and conclusions differ considerably due to the well-known difference between the atmospheric dynamics of temperature and humidity. The procedure involves two stages. First, a block-bootstrap like procedure is applied to the dataset simulating new missing values along the profiles of the variable of interest. This procedure randomly generates gaps of different lengths and altitudes, allowing the creation of "multiple missing scenarios". Second, the simulated missing values are estimated using some Gaussian Processes (GPs) and Linear Interpolators (LINTs). The performance of the two approaches is evaluated as the difference between estimated simulated missing values and actual values in the missing place in terms of Root Mean Square Error (RMSE), Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE). The settings of different predictor variables are also tested for both approaches.

2.1 Justification

The reasons for using a GP for the interpolation of vertical profiles are many. A GP allows to make inferences on relations between input and targets with a relatively simple framework in which the scientist is only asked to optimize some initial conditions [Rasmussen (2003)]. Moreover, it provides an estimation of the data prediction uncertainty. Finally, it shows higher performance in managing different variability scenarios. In general, a GP S(x) is a stochastic process presenting the property that, for any collection of the

[1] **Pietro Colombo**
Università degli studi di Bergamo, viale G.Marconi 5, 24044, Bergamo (Italy), e-mail: pietro.colombo@unibg.it
[2] **Alessandro Fassò**
Università degli studi di Bergamo, viale G.Marconi 5, 24044, Bergamo (Italy), e-mail: alessandro.fasso@unibg.it

locations x1…xn, the joint distribution of S={S(x1),..,S(xn)} is a Multivariate Gaussian Distribution". Such processes are defined by both a *mean function* $\mu(x) = E[S(x)]$ and the *covariance function* $\gamma(x_i, x_j) = \mathrm{cov}\left(S(x_i), S(x_j)\right)\ \ i, j = 1, n \quad and \ \ i \neq j$.

In our context, we denote the observed humidity at time t by y(t). Here, t is the time in seconds from the radiosondes' take off. The observed humidity is assumed to be equal to S(t), which denotes the unobserved "true" humidity, plus a zero mean Gaussian error $\varepsilon(t)$. Moreover, for some basis functions h(t) applied to the matrix dimension dxt of d predictors and t observations , we assume that $E(y(t)) = \beta' h(t)$, where β is the vector of the coefficients of the predictors. It is widely known [Diggle et al.(1998)] that the covariance function affects surface smoothness generated by GP. Hence, using different covariance structures and initial values for the covariance parameters (usually length scale parameter, signal variance and noise variance) defines models for an uncountable number of variability patterns. In synthesis, a single GP describes the joint realization of relative humidity at different locations through the specification of a covariance structure and a function basis. The choice and the combinations of these ingredients can accurately predict many variability patterns. However, linear interpolation is the simple approach currently used at GRUAN to interpolate the missing values. Thus, the comparison we put in place aims to compare the performance of the new approach against an established procedure.

3. Conclusions

This paper provides evidence that GP has superior performance than LINT in predicting and evaluating the uncertainty of the missing values of the essential climate variable relative 'humidity'. This result differs from the one obtained by [Fassò et al. (2020)] for temperature, in which the LINT has been shown to be equivalent to the GP approach. The paper gives details about GP model selection, validation, and performance comparisons.

References

1. Diggle, Peter J., Jonathan A. Tawn, and Rana A. Moyeed. "Model-based geostatistics". Journal of the Royal Statistical Society: Series C (Applied Statistics) 47.3 (1998): 299-350.
2. Dirksen, R. J., et al. "Reference quality upper-air measurements: GRUAN data processing for the Vaisala RS92 radiosonde." Atmospheric Measurement Techniques 7.12 (2014): 4463-4490.
3. Fassò, Alessandro, Michael Sommer, and Christoph von Rohden. "Interpolation uncertainty of atmospheric temperature profiles." Atmospheric Measurement Techniques 13.12 (2020): 6445-6458.
4. Madonna, Fabio, et al. "Radiosounding HARMonization (RHARM): a new homogenized dataset of radiosounding temperature, humidity and wind profiles with uncertainty." Earth System Science Data Discussions (2020): 1-38.
5. Rasmussen, Carl Edward. "Gaussian processes in machine learning." Summer school on machine learning. Springer, Berlin, Heidelberg, 2003.

Identification of main factors impacting human exposure in inductive power transfer systems

Paul Lagouanelle[1], Fabio Freschi[2] and Lionel Pichon[3]

Key words: Inductive power transfer, EMC, Metamodeling

1. Introduction and context

In recent years, inductive power transfer (IPT) systems have been widely developed in several fields such as biomedical engineering, consumer electronics and automotive industry.

With such increasing use, the human exposure to the radiated electromagnetic fields from these systems has to be deeply investigated and measured [Cirimele(2018)]. The properties of shields and materials used in such systems are very dependent on many parameters (thickness, number of layers, electromagnetic properties, etc.) and these parameters are corrupted by many uncertainties (measurements errors, approximations) that may have a significant impact on the shielding effect and/or the global efficiency.

For such uncertainty propagation studies, statistical methods based on Monte Carlo simulations may provide reliable results. With this approach, a large set of inputs are considered and many evaluations of a model response are needed. This leads to a heavy computational cost in case of complex system configurations. To avoid the computational burden and deal with a large variability of data, it can be very useful to build adequate meta-models (or surrogate models) [Bilicz(2016)] [Knaisch(2016)][Marelli(2014)].

[1] Paul Lagouanelle
GeePs – Group of electrical engineering - Paris, CNRS, CentraleSupélec, Université Paris-Saclay, Sorbonne Université, 91192 Gif-sur-Yvette
e-mail: Paul.Lagouanelle@geeps.centralesupelec.fr
[2] Fabio Freschi
Dipartimento Energia, Politecnico di Torino, 10129 Torino, Italy,
e-mail: fabio.freschi@polito.it
[3] Lionel Pichon
GeePs – Group of electrical engineering - Paris, CNRS, CentraleSupélec, Université Paris-Saclay, Sorbonne Université, 91192 Gif-sur-Yvette
e-mail: Lionel.Pichon@geeps.centralesupelec.fr

2. Methodology and results

In order to manage the variability of parameters involved in the assessment of level exposure due to inductive power transfer systems for electric vehicles, the use of surrogate models based on Kriging and Polynomial chaos expansions are presented and applied to simplified but realistic inductive power transfer configurations.

The first case is the power system developed in Politecnico di Torino for dynamic applications [Lagouanelle(2019)]. The second case, studied in INRiM, addresses the exposure related to a system located under a light vehicle [Lagouanelle(2021)]. Meta-models are built from 3D computational results. A sensitivity analysis using Sobol's indices allows to identify the more impacting geometrical of physical parameters involved in the configuration. The study has been performed in the framework of the European project MICEV (Metrology for Inductive Charging of Electric Vehicles).

Acknowledgment

The results here presented are developed in the framework of the 16ENG08 MICEV Project. The latter received funding from the EMPIR programme cofinanced by the Participating States and from the European Union's Horizon 2020 research and innovation programme.

References

1. S. Bilicz, S. Gyimóthy, J. Pávó, P. Horváth, K. Marák, Uncertainty quantification of wireless power transfer systems, IEEE Wireless Power Transfer Conference (WPTC), Aveiro, Portugal 5-6 May 2016.
2. V. Cirimele, M. Diana, F. Freschi, M. Mitolo, "Inductive power transfer for automotive applications: state-of-the-art and future trends, IEEE Trans. on Industry Applications, Vol. 54, n°5, pp 4069-, 2018.
3. K. Knaisch, P.Gratzfeld, Gaussian process surrogate model for the design of circular, planar coils used in inductive power transfer for electric vehicles, IET Power Electron., Vol. 9, Iss. 15, pp. 2786-2794, 2016.
4. P. Lagouanelle, V-L Krauth, L. Pichon, Uncertainty Quantification in the Assessment of Human Exposure near Wireless Power Transfer Systems in Automotive Applications, Automotive 2019, Torino, July 2-4th, 2019.
5. Paul Lagouanelle, Oriano Bottauscio, Lionel Pichon, Mauro Zucca. Impact of parameters variability on the level of human exposure due to inductive power transfer. IEEE Transactions on Magnetics, 2021.
6. S. Marelli, and B. Sudret, UQLab: A framework for uncertainty quantification in Matlab, Proc. 2nd Int. Conf. on Vulnerability, Risk Analysis and Management (ICVRAM2014), Liverpool, United Kingdom, 2014, 2554-2563.

Entropy-based explanations of multidimensionality in ordinal responses

Leslie Pendrill[1]* Jeanette Melin[1], and Stefan Cano[2]

Key words: Entropy, Information, Metrology, Measurement system analysis, Rasch, Cognition, Task difficulty, Person ability, Ordinality, Multidimensionality

1 Ordinal performance metrics

Extension of traditional metrological concepts to include ordinal properties is needed in a widening and increasingly important set of scientific studies and applications (such as in education and healthcare [Pendrill 2019], EMPIR NeuroMET2).

The Rasch [1961] approach is well known as a suitable Generalised Linear Model (GLM, [McCullagh 1980]) for the metrology of ordinal responses in such applications. It provides an analogous separation of object attribute (task difficulty, δ) from instrument attribute (agent ability, θ) as in traditional measurement systems, obtained by making a logistic regression of eq. 1 to the raw ordinal data. The probabilistic response $P_{success}$ is how well the classification system performs when endorsing binary categories (e.g., pass/fail) in the log-odds ratio:

$$log\left(\frac{P_{success}}{1-P_{success}}\right) = \theta - \delta + ln(\rho) \quad (1)$$

An additional factor is the discrimination, ρ, which may vary in each response. A lack of discrimination, for instance, to serial position effects (SPEs) such as primacy and recency when performing the task of recalling word lists has been claimed to be a marker of cognitive decline [Bruno *et al*., 2015].

Of particular interest is the validity of the basic Rasch GLM (eq. 1) in the presence of multidimensionality due to SPEs.

2 Explaining verbal learning list tests. Diagnosis based on PCA loading

Our work [Melin *et al.* 2021] has successively predicted task difficulty δ with Construct Specification Equations (CSEs) of explanatory variables (X) based on informational entropy (easier tasks are more ordered). From earlier studies of the

[1] Research Institutes of Sweden (RISE), 412 58, Göteborg, Sweden;
*****Correspondence: leslie.pendrill@ri.se
[2] Modus Outcomes, Spirella Building, Letchworth Garden City, SG6 4ET, UK

most basic syntax information when recalling of series of blocks or digits [Stenner *et al*. 1983], the work here addresses the increasingly rich cognitive tests of semantic verbal learning list tests [Rey's Auditory Verbal Learning Test, RAVLT].

Regarding CSEs as candidate certified reference materials, it should also be possible to predict the effects of scale distortion where discrimination to SPEs changes according to state of health as a step towards formulating a diagnostic for cognitive decline.

The additional dimensions and scale extensions in response owing to changes $\Delta\rho$ in discrimination to SPEs such as primacy and recency can be modelled in terms of an entropy-based combination of two kinds of principal component analyses (PCA) when formulating CSEs and testing uni-dimensionality, respectively, in the Rasch model: *PCA1 loading*: $L_{p,x} = \mathrm{Cov}(PC_p, X_x) \propto a_{p,x}$ where $PC_p = \sum_{x=1}^{X} a_{p,x} \cdot X_x$; *PCA2 loading*: $L_{p,x} \propto a_{p,x} \cdot \frac{\partial P_{success}}{\partial \delta} \cdot (\Delta\rho \cdot \delta)$. A comparison is shown in Figure 1 between experimental (dots) and theoretical (triangles) loading plots associated with Primacy (*Pr*) and Recency (*Rr*) for the EMPIR NeuroMET cohort.

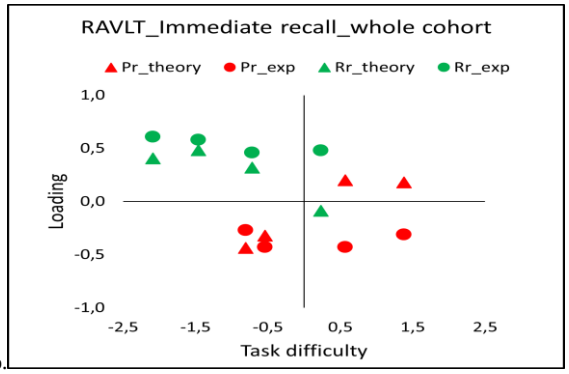


Fig. 1: PCA2 loading plots, exp. & theo.

Improved reliability and validity from these studies can in turn provide opportunities for more practical and accurate measurement in clinical practice, research, and trials and in quality-assuring ordinal properties in general.

This project 18HLT09 NeuroMET2 has received funding from the EMPIR programme co-financed by the Participating States and from the European Union's Horizon 2020 research and innovation programme.

References

1. Bruno P *et al*. 2015 *Alzheimer's & Dement.*, **1**, 81–6, http://dx.doi.org/10.1016/j.dadm.2014.11.002
2. McCullagh P 1980 *J. R. Statist. Soc.* **B, 42**, 109 – 42
3. Melin J, Cano SJ and Pendrill LR 2021, *Entropy* **23**, 212, https://doi.org/10.3390/e23020212
4. NeuroMet2 project, EMPIR 18HLT09
5. Pendrill LR 2019, *Quality Assured Measurement – Unification across Social and Physical Sciences*, ISBN: 978-3-030-28695-8 (e-book), https://doi.org/10.1007/978-3-030-28695-8
6. Rasch G 1961, *Proc. 4th Berkeley symposium on mathematical statistics and probability*, **4**, 321 - 34
7. Stenner AJ, Smith III M, & Burdick DS 1983, *J. of Educ. Measur.* **20**, 305-16

Uncertainty estimation by bootstrap sampling of area shape function in nano-indentation testing

Giacomo Maculotti[1] and Gianfranco Genta[2] and Maurizio Galetto[3]

Key words: Measurement Uncertainty, Bootstrap, Instrumented Indentation Test, Area shape function

Extended Abstract

Instrumented Indentation Test (IIT) is a quick and non-destructive method to thoroughly characterise materials' mechanical properties consisting of a loading-holding-unloading force cycle to indent a material. It relies on a depth-sensing technique to relate the displacement h of the indenter's tip in the material to the projected area A_p of the surface in contact between them, through the area shape function $A_p=f(h; \boldsymbol{a})$, where $\boldsymbol{a}$ is a parameters vector. This allows characterising the nano scale, e.g. metallic grains, and enabling quality control of advanced manufacturing, e.g. additive manufacturing, and finishing processes, e.g. thermal treatments, coatings, thin films. Thus, the evaluation of characterisation measurement uncertainty is essential to provide users with confidence in the obtained results. Barbato(2017) showed that major factors influencing the uncertainty are the stiffness of the platform and the parameters of A_p, i.e. $\boldsymbol{a}$, which need calibration. This is performed according to ISO 14577-2:2015, which presents some criticalities [Maculotti(2020)]. Amongst the others, the effect of the functional form of A_p on the metrological performances of the characterisation are unreported. Despite literature suggests some practical guidelines to choose most adequate models [Oliver(2004)], thorough metrological assessment lacks. To achieve it an uncertainty estimation of the A_p is necessary. A simulative approach based on bootstrap sampling is herein proposed to this scope. Such approach is resorted to cope with the iterative nature of the calibration routine [Galetto(2020)] and include the correlation of the raw measured

[1] Giacomo Maculotti
DIGEP, Politecnico di Torino, C.so Duca degli Abruzzi 24, 10129 Torino, Italy, e-mail: giacomo.maculotti@polito.it
[2] Gianfranco Genta
DIGEP, Politecnico di Torino, C.so Duca degli Abruzzi 24, 10129 Torino, Italy, e-mail: gianfranco.genta@polito.it
[3] Maurizio Galetto
DIGEP, Politecnico di Torino, C.so Duca degli Abruzzi 24, 10129 Torino, Italy, e-mail: giacomo.maculotti@polito.it

quantities, i.e. the applied force and the indenter displacement [Maculotti(2020)]. The uncertainty evaluation allows comparing the performances of different models of A_p available in literature (see Table 1). Models 4 to 6 are used to extend calibration validity beyond tested range. Model 1 and 4 are considered despite the rare adoption for the complexity. Experiments are carried out on a commercial state-of-the-art platform with a Berkovich indenter. Testing is performed considering the same model during both calibration and subsequent characterisation of calibrated reference materials. Results are provided in terms of calibrated parameters and mechanical characterisation. Uncertainty of the results is evaluated by bootstrap sampling the raw data generating 11,000 datasets. Results show a remarkable dependence of accuracy and precision of characterisation on the model choice (see e.g. Figure 1).

Table 1: Functional form of the models describing the projected area considered in the comparison

Model number	***Model equation***
1	$A_p(h) = a_8h^2 + a_7h + a_6h^{1/2} + a_5h^{1/4} + \cdots + a_0h^{1/128}$
2	$A_p(h) = a_2h^2 + a_1h + a_0h^{1/2}$
3	$A_p(h) = a_2h^2 + a_1h + a_0$
4	$A_p(h) = 24.5h^2 + a_7h + a_6h^{1/2} + a_5h^{1/4} + \cdots + a_0h^{1/128}$
5	$A_p(h) = 24.5h^2 + a_1h + a_0h^{1/2}$
6	$A_p(h) = 24.5h^2 + a_1h + a_0$

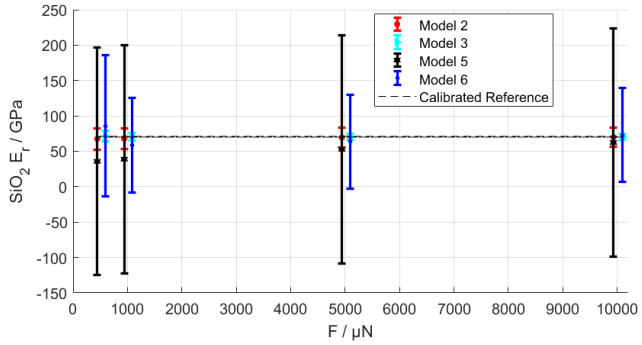


Figure 1. Comparison of characterisation of fused silica (SiO_2) reduced modulus E_r; Model 1 and 4 are unreported for introducing a significant systematic difference.

References

1. Barbato, G. Genta, G., Cagliero, R., Galetto M., Klopfstein, M.J., Lucca, D.A., Levi, R.: Uncertainty evaluation of indentation modulus in the nano-range: Contact stiffness contribution. CIRP Ann. **66**, 495—498 (2017)
2. Galetto, M., Maculotti, G., Genta, G.: Single-step calibration method for nano indentation testing machines. CIRP Ann. **69**, 429—432
3. Maculotti, G., Genta, G., Galetto, M.: Criticalities of iterative calibration procedures for indentation testing machines in the nano-range. Proc 20th EUSPEN I.C.E., Geneva 8-12 June 2020
4. Oliver, W.C., Pharr, G.M.: Measurement of hardness and elastic modulus by instrumented indentation: Advances in understanding and refinements to methodology. J. Mater. Res. **19**, 3—20 (2004)

Radiation dose estimation via the contaminated Poisson and negative binomial methods in partial-body exposures

Adam Errington, Jochen Einbeck, Jonathan Cumming

Key words: Biomarker, Dosimetry, Dispersion, Zero-Inflation, Uncertainty

1 Introduction

One of the biggest challenges in biological dosimetry is the satisfactory conversion of a measured quantity of radiation damage, such as a dicentric or foci yield, into an estimate of dose. All biodosimetric methods require a calibration curve in order to translate the observed yield of damage in cells into an estimate of radiation dose. In addition, the choice of radiation biomarker used to quantify the contracted dose through the caused cellular damage, is important in assessing the radiation sensitivity in a patient blood sample. The most commonly used biomarkers are cytogenetic, for example dicentric chromosomes or micronuclei [IAEA (2011)], or protein-based biomarkers such as the γ-H2AX assay [Horn et al (2011)]. The latter considers counts of foci, which form in cells as a result of phosphorylation of the H2AX histone following double-stranded breaks (DSBs). Using immunofluorescence microscopy, these foci become detectable and appear as visible spots under a microscope which can be counted (or scored). Typically, a laboratory will examine

Adam Errington
Department of Mathematical Sciences, Durham University, UK, e-mail: adam.errington@durham.ac.uk

Jochen Einbeck
Department of Mathematical Sciences, Durham University, UK, e-mail: jochen.einbeck@durham.ac.uk

Jonathan Cumming
Department of Mathematical Sciences, Durham University, UK, e-mail: j.a.cumming@durham.ac.uk

between 500 and 2000 cells per slide and then record the total observed foci for each slide.

2 Overdispersion and partial body exposures

The assessment of the absorbed dose is particularly reliable in cases of acute, uniform and whole-body exposures. However, the scenarios of most radiation accidents result in partial-body exposures or non-uniform dose distribution, leading to a differential exposure of lymphocytes in the body. Subsequently for the exposed individuals, their blood will contain a mixture of cells showing no radiation impact at all (which we refer to as *structural zeros*), and cells featuring a distribution of counts according to dose of exposure. As a consequence, the produced yield of dicentric or foci aberrations in a patient blood sample will become overdispersed and therefore no longer conform to the Poisson distribution. In the context of dicentrics, there do exist some methods to infer the degree of partial body exposure from the overdispersion. Specifically, these are Dolphin's contaminated Poisson method (based on the score equations of the zero-inflated Poisson model) [Dolphin (1969)] and the Qdr method [Sasaki and Miyata (1968)]. Although the contaminated Poisson is able to account for any overdispersion arising from zero-inflation, there may still remain alternate sources of overdispersion which cannot always be removed. For instance, experimental factors which generate variability in the scoring process of cells are all absorbed by the dispersion value. Such dispersion-generating effects are generally present for both manual and automated scoring of γ-H2AX foci. Therefore it remains that, in the case of partial body exposures, there are no statistical procedures to follow for the γ-H2AX assay.

This work focuses on updating the contaminated Poisson method by using the score equations of the zero-inflated negative binomial type 1 (ZINB1) regression, enabling an estimate of the radiation dose to be found in the presence of both zero-inflation and overdispersion. As an extension, we discuss and compare how to measure the uncertainty associated with a given dose estimate via the delta, Merkle and ISO method [Merkle (1983); Ainsbury et al (2017)]. The results we present are based on simulated zero-inflated Poisson and NB1 data, with the non-inflated part being generated using an external γ-H2AX whole-body calibration curve.

References

Ainsbury E, Higueras M, Puig P, Einbeck J, Samaga D, Barquinero J (2017) Uncertainty of fast biological radiation dose assessment for emergency response scenarios. Int J Radiat Biol 2017;93.:127–135

Dolphin G (1969) Biological dosimetry with particular reference to chromosome aberration analysis. a review of methods. handling of radiation accidents. J Biol

Chem Proc. Int.Symp. Vienna:215–224
Horn S, Barnard S, Rothkamm K (2011) Gamma-H2AX-based dose estimation for whole and partial body radiation exposure. PLoS One 6, e25113
IAEA (2011) Cytogenetic dosimetry: applications in preparedness for and response to radiation emergencies. International Atomic Energy Agency Vienna
Merkle W (1983) Statistical methods in regression and calibration analysis of chromosome aberration data. J Biol Chem Radiat Environ Biophys. 21:217–233
Sasaki M, Miyata H (1968) Biological dosimetry in atom bomb survivors. J Biol Chem Nature 220:1189–1193

Simultaneous inference for comparing classifier performance via kappa-type coefficients

Amalia Vanacore, Maria Sole Pellegrino and Armando Ciardiello

Key words: Classifier accuracy, kappa-type coefficients, Multi-class classifications, Multiple comparison

1 Introduction

An important issue in classification problems is the comparison of classifiers predictive performance, commonly measured as proportion of correct classifications and often referred to as accuracy or similarity measure.

In the last decades Cohen Kappa coefficient [3], correcting the proportion of observed agreement with that expected by chance, has been adopted as a measure of accuracy also within the context of expert system, machine learning and data mining communities [1]. The main criticism with Cohen Kappa concerns its dependency on marginal frequencies which converts a reasonably high observed agreement into an unrealistic coefficient value [5] when dealing with extreme unbalanced marginals, a case frequently encountered in machine learning problems [4].
This paper recommends the adoption of two kappa-type coefficients that, although developed to solve the paradoxical behavior of Cohen Kappa, have not received much attention as accuracy measures in classification problems. The recommended coefficients for binary and nominal classifications are the Uniform Kappa [2] and the *AC* coefficient [6], whereas their weighted variants accounting for misclassification severity are able to deal with ordinal classifications.

Dept. of Industrial Engineering, University of Naples "Federico II", p.le Tecchio 80, 80125, Naples
e-mail: amalia.vanacore@unina.it

2 Testing accuracy of multi-class classifiers

The applicability and usefulness of the recommended kappa-type coefficients are illustrated through two case studies concerning multi-class classification problems in the field of smart viticulture and wine quality. The accuracy of four machine learning algorithms, that is Deep Neural Network, Random Forest, Extreme Gradient Boosting and Support-Vector Machines, is assessed via Uniform Kappa and AC coefficients, which are compared each other as well as against similarity measure and Cohen Kappa.

In order to identify which machine learning algorithm best-suits the problem at hand, the accuracy of the four algorithms is tested via max-*t* test [8]. The max-*t* test is a simultaneous inference procedure testing if each algorithm exceeds a performance threshold taking into account the joint distribution of test statistics and the correlation between the performance estimates. It allows to test each null hypothesis individually while maintaining the family wise error rate (FWER) bounded by the nominal significance level and provides information about which individual null hypothesis is significantly rejected [7].

In the first case study, the differences between the similarity measure and kappa-type coefficients are negligible; this result is in line with our expectations due to both nominal nature of classifications and balanced marginal frequencies. Vice-versa, the results of the second case study, dealing with unbalanced marginals and ordinal classifications, reveal the absence of paradoxical behavior in Uniform Kappa and *AC* coefficient and the positive effect of a weighting scheme accounting for misclassification severity. Such results shed light on the advantages of Uniform Kappa and *AC* as measures of classifier accuracy and especially of their weighted variants for ordinal classifications.

References

1. Ben-David, A.: Comparison of classification accuracy using Cohen's Weighted Kappa. Expert Systems with Applications **34**(2), 825–832 (2008)
2. Brennan, R.L., Prediger, D.J.: Coefficient kappa: some uses, misuses, and alternatives. Educational and Psychological Measurement **41**(3), 687–699 (1981)
3. Cohen, J.: A coefficient of agreement for nominal scales. Educational and psychological measurement **20**(1), 37–46 (1960)
4. Delgado, R., Tibau, X.A.: Why Cohen's Kappa should be avoided as performance measure in classification. PloS ONE **14**(9), e0222,916 (2019)
5. Feinstein, A.R., Cicchetti, D.V.: High agreement but low kappa: I. The problems of two paradoxes. Journal of Clinical Epidemiology **43**(6), 543–549 (1990)
6. Gwet, K.L.: Computing inter-rater reliability and its variance in the presence of high agreement. British Journal of Mathematical and Statistical Psychology **61**(1), 29–48 (2008)
7. Hothorn, T., Bretz, F., Westfall, P.: Simultaneous inference in general parametric models. Biometrical Journal: Journal of Mathematical Methods in Biosciences **50**(3), 346–363 (2008)
8. Raschka, S.: Model evaluation, model selection, and algorithm selection in machine learning. arXiv preprint arXiv:1811.12808 (2018)

Hybrid Twins for empowering performance-based engineering based on advanced real-time physics, informed AI and smart-metrology

Francisco CHINESTA

Arts et Métiers Institute of Technology, Paris
Francisco.Chinesta@nesam.eu

World is changing very rapidly. Today we do not sell aircraft engines, but hours of flight, we do not sell an electric drill but good quality holes, … and so on. We are nowadays more concerned by performances than by the products themselves. Thus, the new needs imply focusing on the real system subjected to the real loading that it experienced until the present time in order to predict the future responses and in this manner, anticipate any fortuity event or improve the performances. Here, usual modeling and simulation techniques are limited because of the fact that a model is sometimes no more than a crude representation of the reality. Artificial Intelligence irrupted and became a major protagonist in many areas of technology and society at the beginning of the third millennium, however many times it require impressive training efforts (incredible amount of data, most of them inexistent, difficult to collect and manipulate, extremely expensive in time and resources).

A highway to circumvent these difficulties and successfully accomplishing the most efficient (fast, accurate and frugal) generation of information and knowledge facilitating a real-time decision-making in engineering consists of a hybrid paradigm embracing accuracy and agility, hybridizing real-time physics and real-time data-driven modelling, while accommodating smart-metrology, for defining a new alliance in which data will enrich models and the last making smarter data and metrology (the right data, at the right scale, collected at the right place and time for the expected prupose).

“Biodigital Twins”: optimizing orthopaedic implants

Michael Gasik[1]

Key words: optimization, multi-objective optimization, multi-variate analysis, orthopaedic implants, surface, i*n vitro, in vivo*.

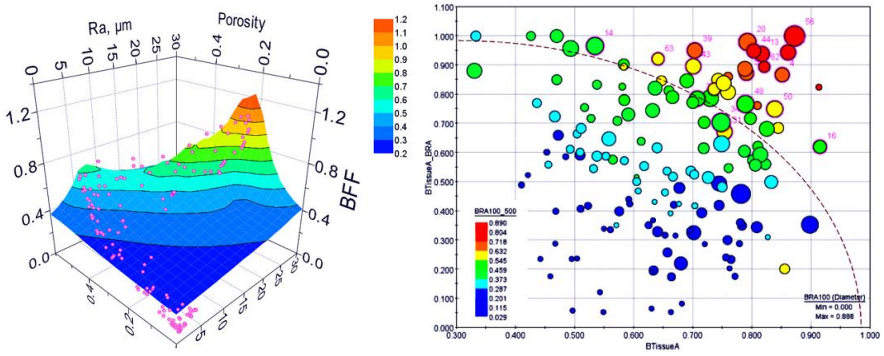
The use of implants has grown dramatically over the past years, driven by ageing of populations, and the need and desire of the patients to maintain the same level of activity and quality of life [Prasad(2017), Gasik(2020)]. Many types of biomaterials are presently available for use in different implants [Burdick & Mauck (2011), von Recum (1998)]: metallic alloys, ceramics, composites, polymers are all used, whether with or without living cells, medical substances or some other additions like antifouling or antibacterial factors.

Despite the success of orthopaedic implants, there is still a risk of foreign body reactions and infections, which are devastating complications - they drastically compromise patients' quality of life, causing pain, immobility (in the USA alone infective cases expected increase by 673% for hip and 174% for knee arthroplasties till 2030, with the annual revisions costs exceed $1,620 millions by 2020). The formation of stable and virulent biofilm by causative pathogens is central to the occurrence and the recalcitrance of these infections [Rimondini & Gasik (2018)]. Different factors (potential, surface energy, activity), local environment (pH, ionic strength), surface topography, porosity, hydrophobicity, micro-fluidics are all affecting the adhesion.

In this work we show the application of some multi-variate methods (response surface modelling, self-organizing maps, clustering, neural networks) to have a look on experimental *in vitro* and *in vivo* data aimed on finding the optimal (“ideal”) surface condition of orthopaedic titanium alloys. In ideal case, such surface can resist (minimize) risk of biofilm formation and at the same time provide the best conditions to osteointegration. Here modeFRONTIER (ESTECO SpA, Italy) – as a statistical tool [Poles & Margonari (2008)] - and LIONsolver (Reactive Search srl, Italy) as a data mining software have been used in this study.

Experiments [Gasik e.a. (2012)] have confirmed such combination of surface properties can be manufactured with available techniques, leading to least bacterial contamination (Fig. 1,a). Followed *in vivo* tests have confirmed that optimized combination is indeed beneficial for bone ingrowth and integration (Fig. 1,b).

[1] Aalto University Foundation, Espoo, FIN-00076 AALTO, e-mail: michael.gasik@aalto.fi



a b

Figure 1: Biofilm formation function (BFF) vs. porosity and roughness R_a (a) and bone/tissue contact parameters from *in vivo* tests (b)**.** BTissueA = formed bone tissue area fraction; BRA = Bone regeneration area; _100 = within 100 µm from implant surface; _100_500 = within zone of 100 to 500 µm from the implant surface.

Conclusion

The advantages of this model-free optimization are in that no model or material assumptions are required, as the analysis was made solely based on a sufficient amount of experimental data. The approach demonstrates that with such analysis it is possible to create a "biodigital twin" of the system where biological data could be aligned and correlated with engineering parameters to be reliably controlled during manufacturing.

References

Burdick J.A., Mauck R.L. (eds.). Biomaterials for tissue engineering applications: a review of the past and future trends. Springer, Wien - New York (2011)

Gasik, M., Van Mellaert, L., Pierron, D., et al.: Reduction of biofilm infection risks and promotion of osteointegration for optimized surfaces of titanium implants. Adv. Healthcare Mater. **1**, 117–127 (2012).

Gasik, M.: Biomechanical characterization of engineered tissues and implants for tissue/organ replacement applications. In: Vrana, N.E., Knopf-Marques, H., Barthes, J. (eds.) Biomaterials for Organ and Tissue Regeneration; Woodhead Publishing: Cambridge, UK, pp. 599–627 (2020).

Poles, S., Margonari, M.: modeFRONTIER as a statistical tool. Enginsoft Newsletter, Special Issue, 24-28 (2008).

Prasad, K., Bazaka, O., Chua, M., Rochford M. et al.: Metallic Biomaterials: Current Challenges and Opportunities. Materials, **10**, 884 (2017)

Rimondini L., Gasik M. Bacterial attachment and biofilm formation on biomaterials: the case of dental and orthopaedic implants. In: Biomaterials and Immune Response: Complications, Mechanisms and Immunomodulation, Ed. Vrana N.E., CRC Press, Boca Raton, USA (2018).

von Recum A.F. (Ed.), Handbook of biomaterials evaluation: scientific, technical, and clinical testing of implant materials. Taylor & Francis, USA (1998).

Simulation of temperature measurement of inhomogeneous flows by ultrasonic flow meters

Gertjan Kok[1]

Key words: simulation, flow, inhomogeneous temperature, ultrasonic flow meter, ultrasound, method bias

1. Introduction

Transit time based ultrasonic flow meters measure the flow rate of a fluid flowing through a pipe by measuring the time periods t_1 and t_2 it takes for an ultrasonic pulse to travel in both directions between two transducers installed in a pipe. From these measured time periods both fluid speed and speed of sound can be calculated. The speed of sound can be used to derive an average temperature of the fluid. Experimental results concerning this method have been published in the past, see e.g. [Klason (2014)]. In this contribution we study the theoretical bias of the method in the case of inhomogeneous temperatures assuming different use cases and sensor configurations.

2. Mathematical method description

The speed of sound depends on the fluid type, pressure and temperature, $c = c(\text{fluid},\ P,\ T)$. Assuming the fluid composition and pressure are constant and known, the non-linearly dependency of the speed of sound on temperature $c = f(T)$ can be inverted to obtain the temperature from the measured average speed of sound c_{ave}:

$$T_{\text{meas}} = f^{-1}(c_{\text{ave}}).$$

In a pipe with inhomogeneous fluid temperature $T(x, y)$ the speed of sound changes locally as function of the cross section coordinates (x, y), i.e. $c = c(x, y) = c\big(T(x, y)\big)$. The measured transit times t_1 and t_2 in both directions along the path $\wp$ with angle α and parametrized by s are in the general case given by

$$t_{1,2} = \int_{\wp} L \,/\, \big(c\big(x(s), y(s)\big) \pm v(x(s), y(s)) \cos\alpha\big) ds$$

To simplify this equation, we assumed $\alpha = 90°$ and/or $v \equiv 0$, leading to the measured average speed of sound c_{ave} given by

$$c_{\text{ave}} = \left(\int_{\wp} (c(x(s), y(s)))^{-1}\, ds \right)^{-1}.$$

[1] Gertjan Kok
VSL. B.V., Thijsseweg 11, 2629 JA, Delft, the Netherlands, e-mail: gkok@vsl.nl

On the other side, the true ‘mixed cup’ thermodynamic average temperature follows from calculating the average enthalpy h_{ave} over the pipe cross section S using the local enthalpy $h(x,y) = h(T(x,y))$ and mass density $\rho(x,y) = \rho(T(x,y))$ according to

$$h_{\mathrm{ave}} = \left(\iint_S \rho\big(T(x,y)\big) dx dy \right)^{-1} \left(\iint_S h\big(T(x,y)\big) \rho\big(T(x,y)\big) dx dy \right)$$

and subsequently inverting the relationship $h = g(T)$ of enthalpy on temperature, resulting in

$$T_{\mathrm{true}} = g^{-1}(h_{\mathrm{ave}}).$$

This thermodynamic averaging can also be compared by the approximate approaches of volumetric and mass-based averaging.

3. Case studies

In the simulation of the ultrasound based temperature measurement method we will study the theoretical bias of the method for various cases according to the following variations:

- Temperatures around 25 °C and around 300 °C;
- Various assumptions on the temperature distribution: homogeneous, small boundary layer, two circular regions, stratification in two layers, linear stratification;
- Various assumptions on the path configurations: one vertical path, three paths passing through the pipe centre, three parallel vertical paths, three parallel horizontal paths.

Acknowledgement

This work has been performed in the EMRP ENG06 and EMPIR 17IND12 projects. This project has received funding from the EMRP and EMPIR programmes co-financed by the Participating States and from the European Union's Horizon 2020 research and innovation programme.

References

1. Klason, P., Kok, G.J., Pelevic, N. et al. Measuring Temperature in Pipe Flow with Non-Homogeneous Temperature Distribution. Int J Thermophys 35, 712–724 (2014). https://doi.org/10.1007/s10765-014-1579-3

In silico experiments to guide magnetic hyperthermia pre-clinical tests

Marta Vicentini[1], Marta Vassallo[2], Riccardo Ferrero[3] and Alessandra Manzin[4]

Key words: Magnetic hyperthermia, Magnetic nanoparticles, *In silico* models, Electromagnetic simulations, Thermal simulations.

1. Introduction

Magnetic hyperthermia is a therapeutic technique, typically employed in combination with radiotherapy and chemotherapy; this uses magnetic nanoparticles (MNPs) to induce a localized release of heat in a diseased tissue, under the action of external magnetic fields with frequency variable in the range 50 kHz – 1 MHz. The heat deposited in the tissue is responsible for an increase in temperature, which should be in the order of 4-5 °C [Angelakeris(2017)], [Chang(2018)], [Ferrero(2019)]. In pre-clinical studies, typically conducted on murine models, several factors have to be taken into account to optimize heat deposition and reduce side-effects. These comprise undesired electromagnetic induction phenomena, field applicator configuration, heat transfer with the external environment and spatial distribution of MNPs within the target tissue.

In this context, we present a physics based modelling approach to support *in vivo* tests of magnetic hyperthermia [Manzin(2021)]. To this aim, we have developed *in silico* models for the evaluation of eddy current phenomena induced in the body during hyperthermia sessions and for the calculation of thermal effects, consequent to alternating field application and MNP excitation. The study is focused on computational anatomical murine models with variable size, changing the target region where the MNPs are dispersed with different dose and its location in the body.

2. Physics based models

The study is performed by means of different in-house numerical models aimed at determining the thermal effects correlated, first, to the exposure to the only

[1] Marta Vicentini
Politecnico di Torino, Torino (Italy), marta.vicentini@polito.it
Istituto Nazionale di Ricerca Metrologica, Torino (Italy)
[2] Marta Vassallo
Politecnico di Torino, Torino (Italy), marta.vassallo@polito.it
Istituto Nazionale di Ricerca Metrologica, Torino (Italy)
[3] Riccardo Ferrero
Istituto Nazionale di Ricerca Metrologica, Torino (Italy), r.ferrero@inrim.it
[4] Alessandra Manzin
Istituto Nazionale di Ricerca Metrologica, Torino (Italy), a.manzin@inrim.it

alternating magnetic field and, second, to the MNP excitation. The numerical solutions are obtained by using the finite element method (FEM) with linear shape functions, after discretizing the 3D domain (the animal body under investigation) into tetrahedral elements.

To evaluate the eddy current effects produced in biological tissues during the alternating magnetic field application, we have developed a low-frequency electromagnetic field solver, in which displacement currents are neglected. To calculate the thermal effects induced in the animal body due to the exposure to the alternating magnetic field and to the excitation of the MNPs, we have implemented a numerical code that solves the Pennes' bioheat transfer equation [Manzin(2021)].

3. Results

We have carried out *in silico* experiments to investigate 1) possible eddy current effects arising when applying the alternating magnetic field, 2) the role of field applicator geometry and position with respect to the body and 3) the heating effects due to MNP excitation, as a function of target tissue location and MNP dose. Some results are reported below.

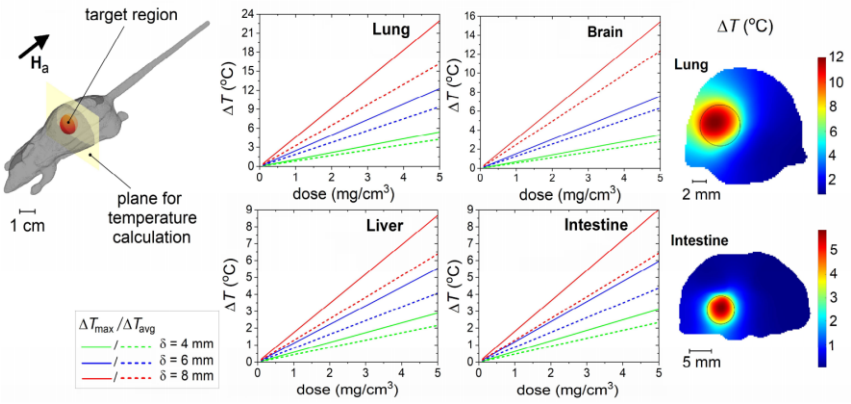


Figure 1: Left: Comparison of temperature increments reached at the heating equilibrium, by varying the mouse organ where the target region is placed, its size δ and the dose of MNPs (with specific loss power in the order of 200 W/g). Right: Maps of temperature increase for δ = 6 mm and MNP dose of 5 mg/cm^3.

References

1. Angelakeris, M: Magnetic nanoparticles: A multifunctional vehicle for modern theranostics. Biochimica et Biophysica Acta **1861**, 1642—1651 (2017).
2. Chang, D. et al.: Biologically Targeted Magnetic Hyperthermia: Potential and Limitations. Frontiers in Pharmacology **9**, 831 (2018).
3. Ferrero, R., Manzin, A., Barrera, G., Celegato, F., Coïsson, M., Tiberto, P.: Influence of shape, size and magnetostatic interactions on the hyperthermia properties of permalloy nanostructures. Scientific Reports **9**, 6591 (2019).
4. Manzin, A., Ferrero, R., Vicentini, M.: From micromagnetic to *in silico* modeling of magnetic nanodisks for hyperthermia applications. Adv. Theory. Simul. **2100013**, 1—13 (2021).

Enhancement of multiphase flow simulations by turbulence damping at the gas-liquid interface

Jiri Polansky and Sonja Schmelter

Key words: multiphase flow, stratified flow, slug flow, computational fluid dynamics (CFD), OpenFOAM, turbulence damping

1 Introduction

Multiphase flow measurement is fundamental in subsea oil and gas production. However, field measurements exhibit high measurement uncertainty (up to 20 %) [Falcone(2013)]. Within the European research project "Multiphase flow reference metrology" [Fin(2021)], the metrological characterization of multiphase flows was improved by computational fluid dynamics (CFD) techniques. The great advantage of CFD is that it gives insight into areas that are hardly accessible by measurement techniques. Therefore, simulations can help to understand the formation of multiphase flow patterns and instability as well as their influence on the measurement process. In this contribution, we present a modeling approach that enhances the treatment of turbulence at the interface between the different phases.

Jiri Polansky
Czech Technical University in Prague, Jugoslavskych partyzanu 1580/3, 160 00 Prague 6 Dejvice, Czech Republic e-mail: jiri.polansky@cvut.cz

Sonja Schmelter
Physikalisch-Technische Bundesanstalt (PTB), Abbestr. 2-12, 10587 Berlin, Germany e-mail: sonja.schmelter@ptb.de

2 Implementation of turbulence damping in OpenFOAM

In industrial applications, turbulence is often modeled by Reynolds-averaging the Navier-Stokes equations (so-called RANS approach). For this method, it is well known that turbulent viscosity is overestimated in the vicinity of the interface. To overcome this problem, [Egorov(2004)] proposed to add a source term to the ω-equation in the k-ω model, which mimics the damping of turbulence close to a solid wall. We implemented Egorov's approach in OpenFOAM on the subclass of shear stress transport (SST) models. Hence, turbulence damping is available for all SST type models that are based on the ω-equation. This includes also hybrid RANS-LES models, like delayed detached eddy simulation (DDES) and detached eddy simulation (DES).

3 Influence of turbulence damping on slug flow

When the gas velocity is increased under stratified flow conditions, waves are formed on the interface. For higher gas-liquid velocity ratio, plug and slug regimes occur. The amplitude and frequency of the waves and slugs depends on the relative velocity between the phases and the properties of fluids, such as their density and surface tension [Hanratty(2013)]. However, if the calculated velocity gradient close to the interface of the phases is underestimated, typical flow instabilities do not appear in the CFD calculations. Due to the better modeling of the velocity gradient at the gas-liquid interface, turbulence damping also enhances the development and evolution of slugs in the pipe. Comparison with gas-liquid experiments shows that turbulence damping implementation leads to a more realistic flow instability simulation.

Acknowledgements This work was supported through the Joint Research Project "Multiphase flow reference metrology". This project has received funding from the EMPIR programme co-financed by the Participating States and from the European Union's Horizon 2020 research and innovation programme.

References

[Fin(2021)] (2021) Final Publishable Report Multiphase flow reference metrology (16ENG07)
[Egorov(2004)] Egorov Y (2004) Validation of CFD codes with PTS-relevant test cases. EVOL-ECORA-D07
[Falcone(2013)] Falcone (2013) Multiphase flow metering: Current trends and future developments. Journal of Petroleum Technology 54, DOI 10.2118/71474-MS
[Hanratty(2013)] Hanratty TJ (2013) Physics of Gas-Liquid Flows. Cambridge University Press

Virtual sensors development for real-time quality assessment in continuous production

Manolo Venturin[1], Giovanni Paolo Borzi[2] and Anteneh Yemane[3]

Common trends impacting the entire European manufacturing industry, such as shorter life cycle of the products, increased complexity in design and manufacturing, higher customization requests, increasing demands on traceability or cost reduction are straining the classic approaches to quality management (generally applied as a production-inspection sequence). These limitations and the pursuit for continuous improvement and innovation motivated ENKI, an Italian company leader in the production and assembly of medical devices such as micro-catheters for angioplasty, to partner the ForZDM[4] research and innovation action in order to identify, develop and implement innovative methods, processes and technologies aimed at zero defect manufacturing (ZDM).

This work presents the methodology for the creation of a Cyber Physical System (CPS) for the real time quality control of multi-layer microtubes production. The CPS takes the form of virtual sensors presented to the production operator aimed at improving the polymer extrusion process monitoring and assist its real-time quality control.

Virtual sensors provide an estimate of physical values through either data-based or simulation-based models, that are run in real time in the guise of production process Digital Twins. Virtual sensors are useful whenever it is impractical or outright technologically impossible to introduce real sensors for process monitoring.

In the ENKI[5] case, virtual sensors are based on mathematical models obtained through multivariate data analysis techniques. However, while the models are data-based, simulation played a fundamental role in the development of the models. In fact, the model development process was carried out in three main steps (Figure 1). This work presents the applied methodology i.e. (i) the (Computational Fluid Dynamics) CFD process simulation, (ii) the experimental Design of Experiments

[1] Ing. Manolo Venturing, Ph.D.
EnginSoft Spa, via Giambellino 7 35135 Padova (ITALY), e-mail: m.venturin@enginsoft.com
[2] Ing. Giovanni Paolo Borzi
EnginSoft Spa, via Giambellino 7 35135 Padova (ITALY), e-mail: g.borzi@enginsoft.com
[3] Ing. Anteneh Yemane, Ph.D.
EnginSoft Spa, via Giambellino 7 35135 Padova (ITALY), e-mail: a.yemane@enginsoft.com
[4] https://www.forzdmproject.eu/
[5] https://www.enki-microtubes.com/

(DoE) setup and (iii) the multivariate modeling techniques applied. Finally, the in process virtual sensors implementation is described.

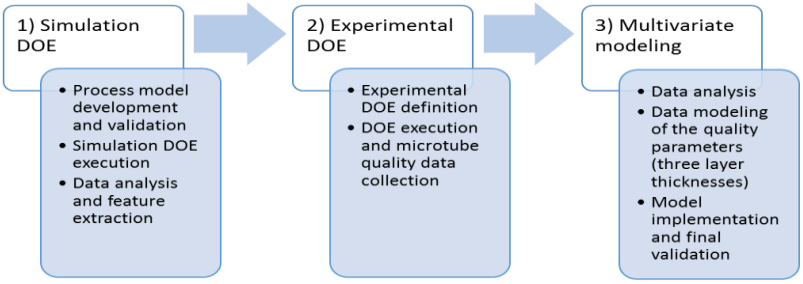


Figure 1: main steps for virtual sensors modeling

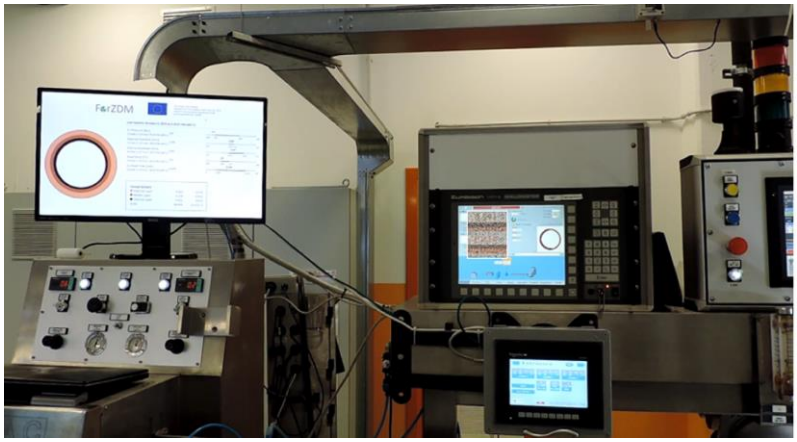


Figure 2: Virtual Sensors deployed in production

Key words: Cyber Physical System, process Digital Twin, Virtual Sensors

References

1. Broy, M.: Software engineering -- from auxiliary to key technologies. In: Broy, M., Dener, E. (eds.) Software Pioneers, pp. 10-13. Springer, Heidelberg (2002)
2. Dod, J.: Effective substances. In: The Dictionary of Substances and Their Effects. Royal Society of Chemistry (1999) Available via DIALOG. http://www.rsc.org/dose/title of subordinate document. Cited 15 Jan 1999
3. Geddes, K.O., Czapor, S.R., Labahn, G.: Algorithms for Computer Algebra. Kluwer, Boston (1992)
4. Hamburger, C.: Quasimonotonicity, regularity and duality for nonlinear systems of partial differential equations. Ann. Mat. Pura. Appl. **169**, 321--354 (1995)
5. Slifka, M.K., Whitton, J.L.: Clinical implications of dysregulated cytokine production. J. Mol. Med. (2000) doi: 10.1007/s001090000086

Prediction of the flow downstream of a 90°-elbow with arbitrary curvature radius and its effect on the accuracy of flow meters

Andreas Weissenbrunner, Sonja Schmelter, Ann-Kathrin Ekat, and Martin Straka

Key words: Computational fluid dynamics (CFD), pipe flow, elbow, bend, Reynolds-averaged Navier-Stokes (RANS), flow meter, uncertainty estimation

Abstract

For accurate flow measurements most flow meters require calibration functions. These calibration curves usually rely on the assumption that the flow profile is fully developed. In industrial applications, the velocity profiles are usually not ideally shaped (i.e. symmetric without cross flow), as they are disturbed by pipe installations such as elbows. Hence, the flow meter reading is not entirely correct. However, it could be corrected if the velocity profile was known precisely.

To obtain a complete velocity profile of a pipe cross section, laser-optical measurements or computational fluid dynamics (CFD) simulations can be used. While measurements provide highly accurate results, their set-ups are expensive and time-consuming. Therefore, they can be performed for a few specific geometries only.

Andreas Weissenbrunner
Physikalisch-Technische Bundesanstalt (PTB), Abbestr. 2-12 D-10587 Berlin, e-mail: andreas.weissenbrunner@ptb.de

Sonja Schmelter
Physikalisch-Technische Bundesanstalt (PTB), Abbestr. 2-12 D-10587 Berlin, e-mail: sonja.schmelter@ptb.de

Ann-Kathrin Ekat
Physikalisch-Technische Bundesanstalt (PTB) Abbestr. 2-12 D-10587 Berlin, e-mail: ann-kathrin.ekat@ptb.de

Martin Straka
Physikalisch-Technische Bundesanstalt (PTB), Abbestr. 2-12 D-10587 Berlin, e-mail: martin.straka@ptb.de

CFD simulations, on the other hand, can be conducted for a variety of different configurations. However, their accuracy is dependent on the validity of the assumptions of the underlying models, geometry, boundary conditions etc.

In this study, we present a procedure that enables an accurate prediction of the axial and cross flow profile downstream of a 90°-elbow with an arbitrary curvature radius. Therefore, CFD simulations with the Reynolds-averaged Navier-Stokes (RANS) model Spalart-Allmaras are performed for elbows of 22 different curvature radii R_c between 0.51 and 10 times the pipe diameter D and a Reynolds number Re $= 5 \cdot 10^4$. To validate the simulations, the numerical results are compared to Laser-Doppler Velocimetry (LDV) measurements, performed for two Re numbers (Re = $5 \cdot 10^4$, $5 \cdot 10^5$) on six cross sections downstream of an elbow with $R_c = 1.4\,D$. For a meaningful validation the uncertainties of the measurement must be sufficiently low. The uncertainties for these measurements are stated to be below 0.5 % ($k = 2$).For a more detailed description of the measurement data, see [1].

To eliminate systematic errors of the simulation results, a function that fits the simulated flow field to the measurement data is derived. It is shown that this correction function can also be applied to enhance the results with different curvature radii R_c. Thus, it allows to predict the velocity profile for any curvature radius $R_c \in [0.51\,D, 10\,D]$. By decomposing the axial flow field into an ideal profile and a disturbed part, the results can be extended to higher Reynolds numbers. Altogether, the proposed procedure allows to predict the flow field for a variety of different curvature radii and Reynolds numbers. Hence, the performance of several flow meters, such as ultrasonic and electromagnetic flow meters can be investigated under these conditions. A procedure to calculate the expected errors of flow meters as well as the uncertainty of these predictions is proposed.

References

1. M. Straka, C. Koglin, and T. Eichler. Segmental orifice plates and the emulation of the 90°-bend. tm - Technisches Messen, 87(1):18–31, 2020.

A MATHMET Quality Management System for data and software

Keith Lines, Jean-Laurent Hippolyte and Peter Harris[1], Nicolas Fischer and Sebastien Marmin[2], Stephen Ellison and Simon Cowen[3], Gertjan Kok[4], Sebastian Heidenreich and Oliver Henze[5], Francesca Pennecchi[6], Alen Bosnjakovic and Merima Čaušević [7], Joao Sousa, Carlos Pires and Olivier Pellegrino[8], and Jörg F. Unger[9]

Key words: Quality Management System, quality assurance, trust, data product, software product

1. Background

An activity of the European Metrology Network for Mathematics and Statistics [MATHMET] is to create a Quality Management System (QMS) to ensure that research outputs in the field of mathematics and statistics for metrology 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.

[1] Keith Lines, Jean-Laurent Hippolyte and Peter Harris
NPL, Hampton Road, Teddington TW11 0LW, UK, e-mail: keith.lines@npl.co.uk
[2] Nicolas Fischer and Sebastien Marmin
LNE, 1 rue Gaston Boissier, 75724 Paris Cedex 15, France,
e-mail: Nicolas.Fischer@lne.fr
[3] Stephen Ellison and Simon Cowen
LGC Limited, Queens Road, Teddington, TW11 0LY, UK,
e-mail: S.Ellison@LGCGroup.com
[4] Gertjan Kok
VSL B.V., Thijsseweg 11, 2629 JA, Delft, the Netherlands, e-mail: GKok@vsl.nl
[5] Sebastian Heidenreich and Oliver Henze
PTB, Abbestrasse 2-12, 10587 Berlin, Germany,
e-mail: sebastian.heidenreich@ptb.de
[6] Francesca Pennecchi
INRIM, Strada delle Cacce 91, 10135 Turin, Italy, e-mail: f.pennecchi@inrim.it
[7] Alen Bosnjakovic and Merima Čaušević
IMBIH**,** Augusta Brauna 2, 71000 Sarajevo, e-mail: alen.bosnjakovic@met.gov.ba
[8] Joao Sousa, Carlos Pires and Olivier Pellegrino
IPQ, Rua António Gião, 2, 2829-513 Caparica, Portugal, e-mail: jasousa@ipq.pt
[9] Jörg F. Unger
BAM, Unter den Eichen 87, 12205 Berlin, Germany, e-mail: joerg.unger@bam.de

A Special Session is proposed to include (a) an introduction to the QMS currently under development, (b) descriptions of case studies being used by different MATHMET partners to apply the QMS and thereby inform its development, and (c) a round table with the aim to collect input and feedback from stakeholders. In this context, stakeholders comprise both organisations wishing themselves to make use of the QMS as well as those wanting to have assurance in the data and software products provided by MATHMET partners.

2. QMS for data and software

An overview of the QMS and the principles behind it will be presented. The QMS is structured according to the life cycles for developing software (including the underlying models and algorithms) and builds upon existing good practices for data management, and software development, following a process-based approach [ISO(2015),ISO(2017)]. One element of the QMS is a risk assessment based on the criticality and complexity of the data or software, leading to the assignment of an *integrity level* that is used to decide quality management activities and interventions appropriate for the identified risks. Importance is given to the validation of software that constitutes a link in the traceability chain for computation-based metrology. Methodologies and tools will be included to support metrologists, scientists and engineers to meet the requirements of the QMS and thereby establish trustworthiness. Based on an ontology and existing Standards, data quality and software development flowcharts will be used.

3. Case studies

The data and software products provided by partners of MATHMET can differ greatly in terms of their functionality, the requirements set by a range of different users and customers, and whether they are developed by single organizations or through collaborations. Some of those products are developed within national programmes, others within European-funded research and development programmes such as EMPIR, and yet others through commercial contracts. A QMS is needed that is flexible enough to be used for these different purposes.

Several case studies are being undertaken to support the development of the QMS and these will be briefly described. They include software implemented in MATLAB for conformity assessment [CASoft], software developed in Python for exploiting redundancy in sensor network data [Met4FoF], software to support the analysis of data from homogeneity studies provided using Microsoft Excel, software for the calibration of pressure gauges, and data in the form of real and simulated ECG signals to represent different cardiac conditions [Medal-Care, PhysioNet].

4. Round table

The third part of the session consists of a round table with the goal to receive input and feedback from the audience. Some of the questions to be discussed include:

- Does the application of this QMS give you confidence in the data and software products provided by NMIs such as through MATHMET? Is there anything missing from the QMS that would give (greater) trust in those products?

- Would you be interested to use the QMS yourself? Why or why not?

Acknowledgement

The project 18NET05 MATHMET has received funding from the EMPIR programme co-financed by the Participating States and from the European Union's Horizon 2020 research and innovation programme.

References

1. CASoft https://www.lne.fr/en/software/CASoft Cited 25 March 2021.
2. ISO 9001:2015 Quality management systems – Requirements
3. ISO/IEC 17025:2017 General requirements for the competence of testing and calibration laboratories
4. MATHMET https://www.euramet.org/european-metrology-networks/mathmet/ Cited 25 March 2021.
5. Medal-Care https://www.ptb.de/empir2019/medalcare/home/ Cited 25 March 2021.
6. Met4FoF https://www.ptb.de/empir2018/met4fof/information-communication/software/ Cited 25 March 2021.
7. PhysioNet https://physionet.org/content/ptb-xl/1.0.1/ Cited 25 March 2021.

Deep Drug Discovery

Djork-Arné Clevert

Key words: Machine learning, drug discovery

1 Abstract

Deep learning is currently the most popular technique of machine learning, and every other day we read new success stories about how it is dramatically changing a wide range of applications. In recent years, it has been increasingly used in drug discovery to predict potential drug targets, screen for active molecules or plan chemical retrosynthesis routes. In this talk I will present some of our recent publications and give a brief insight into how we use deep learning in our research at Bayer to (i) de novo design molecules under multi-parametric objectives [Méndez-Lucio(2020), Winter(2019a)], (ii) predict protein-ligand interactions using unsupervised learned representations and multitask regression [Winter(2019b), Kim(2020)], (iii) solve the inverse molecular problem for extended chemical fingerprints [Le(2020)] and molecular depictions [Clevert(2021)] or to (iv) predict molecular properties with multitask graph convolutional networks [Montanari(2020)].

References

[Clevert(2021)] Clevert DA (2021) Img2Mol - Accurate SMILES Recognition from Molecular Graphical Depictions DOI 10.26434/chemrxiv.14320907.v1

[Kim(2020)] Kim P (2020) Deep protein-ligand binding prediction using unsupervised learned representations. ChemRxiv

Djork-Arné Clevert
Bayer, Machine Learning Research, Müllerstraße 178, Berlin, Germany e-mail: djork-arne.clevert@bayer.com

[Le(2020)] Le T (2020) Neuraldecipher-reverse-engineering ecfp fingerprints to their molecular structures. Chemical Science 38(11):10,378–10,389

[Méndez-Lucio(2020)] Méndez-Lucio O (2020) De novo generation of hit-like molecules from gene expression signatures using artificial intelligence. Nature communications 11(1):1–10

[Montanari(2020)] Montanari F (2020) Modeling physico-chemical admet endpoints with multitask graph convolutional networks. Molecules 25(1):44

[Winter(2019a)] Winter R (2019a) Efficient multi-objective molecular optimization in a continuous latent space. Chemical Science 10(34):8016–8024

[Winter(2019b)] Winter R (2019b) Learning continuous and data-driven molecular descriptors by translating equivalent chemical representations. Chemical Science 10:1692–1701

Aggregation in Cell Culture – App development for cell clumping scoring

Edouard Duquesne[1]

During cell culture development processes large volumes of data are generated combining process parameters and analytical measurements to monitor and control the product quality.

Different types of images can also be generated during the process development activities at lab scale, to measure overall cell health such as Viable Cell Density (VCD). Typically, such information is obtained through automated feature extractions by commercial equipment. The direct analysis of the images generated by the standard image-based analytics systems such as ViCell are not common practices and their influence in the decision process remains low.

This study presents a method enabling the measurement and scoring of cell clumping from ViCell images. For each clone producer of antibody, the scoring of cell *clumpiness* is calculated via an automated workflow built in Jupyter Notebook/RShiny and provides an effective solution for the lab scientist to support their decision making in clump assessment.

To build this model, features were extracted from images using OpenCV and contour extraction. Examples of features are: length of contour, aspect ratio of the contour or area of contour. When the X matrix of features was created, the data were visualized in lower dimension using UMAP algorithm. A supervised model was then built to fit the labels using a five-fold cross validation and a test set of 33% of the initial dataset size.

As an example, for one cell culture run in Ambr15, up to 2500 images can be generated. Ranking these images manually would take about five hours and would be highly subjective to the operator. With the proposed model, the scoring takes 10 minutes, is automated and can be replicated.

Key words: Cell Aggregation, Machine learning, Image processing, Cell culture

[1] Edouard Duquesne
Edouard Duquesne, Sanofi, 13 Quai Jules Guesde, 94400 Vitry-sur-Seine, e-mail: edouard.duquesne@sanofi.com

Protein language modeling and transfer learning applied to predict TCR-epitope affinity

Gurpreet Singh[1], Matthias Bal[2], Nanda Kumar Sathiyamoorthy[1], Paul Smyth[2] and Ahmed Essaghir[2]

[1]GSK, Upper Providence, PA19426, United States

[2]GSK, 1330 Rixensart, Belgium

Key words: Protein language modeling, T-Cell receptor, Immune system, Masked language modeling, Transfer learning

Predicting the antigen affinity towards T Cell Receptors (TCRs) has been a long-standing challenge within the field of immuno- and bio-informatics. Knowing the binding affinity of T Cell Receptors towards their cognate epitopes will contribute to better understanding of T cell adaptive immune response and cross-reactivity, with implications on disease understanding, diagnosis, vaccine design and TCR therapeutic applications. Until recently, the only available immuno-informatics tools for T cell epitope screening are based on the HLA-epitope binding predictions (Hoof et al., 2009). However, modeling the direct interaction between TCR-epitope pairs is becoming nowadays applicable and many publications tackled this problem using different ML/AI approaches ranging from random forest (Gielis et al., 2019) to NLP deep learning architectures (Springer et al., 2020). The available TCR-epitope prediction models could be divided into two different categories: (i) classification, where the prediction is the binding probability of a TCR-epitope pair; (ii) regression, where the actual measurement value of the affinity of the TCR-epitope pair is predicted. The boom of TCR sequencing technology and the high-throughput screening of their epitopes contributed to populate public databases, such as EIDB (Vita et al., 2019) and VDJDB (Bagaev et al., 2020) that are queried to train mostly classification models. However, the scarcity of databases with TCR-epitope affinity measures made it difficult to build regression models. Luckily, single-cell immune profiling with DNA-barcoded dextramers of

HLA-epitopes coupled to next generation sequencing allows for the unravelling of both the TCR sequences and their cognate epitopes, and the determination of their binding affinity estimated for the DNA-barcode counts. This unprecedented amount of data opens the possibility to train ML/AI models to predict TCR-epitope affinity. In this work we trained a general protein language model on a large corpus of protein sequences and then fine-tune on 10X Genomics single-cell TCR-epitope binding counts data set to solve this task (10X Genomics, 2019). Epitope counts are log2 transformed and affinity is considered as the maximum value for each TCR. The final dataset consisting of 5,985 TCRs and 42 epitopes was clustered by applying GLIPH (grouping of lymphocyte interactions by paratope hotspots) algorithm on the TCRs and the train/test split sets were made by stratifying the bigger clusters (>= 5 unique CDR3b sequences) and independently splitting the smaller ones (Glanville et al., 2017). Our results show that this is a learnable and predictable task based on TCR and epitopes alone without any additional covariates. The performances of our models reached the following metrics: mae=0.794, r2=0.448 and rmse=1.029. Previous deep learning models that used a subset of the same data set reported an R^2 score below 0.2 (Fischer et al., 2020).

References

10X Genomics. (2019). *A New Way of Exploring Immunity - Linking Highly Multiplexed Antigen Recognition to Immune Repertoire and Phenotype*. 10x Genomics. https://www.10xgenomics.com/solutions/vdj/

Bagaev, D. V., Vroomans, R. M. A., Samir, J., Stervbo, U., Rius, C., Dolton, G., Greenshields-Watson, A., Attaf, M., Egorov, E. S., Zvyagin, I. V., Babel, N., Cole, D. K., Godkin, A. J., Sewell, A. K., Kesmir, C., Chudakov, D. M., Luciani, F., & Shugay, M. (2020). VDJdb in 2019: Database extension, new analysis infrastructure and a T-cell receptor motif compendium. *Nucleic Acids Research*, *48*(D1), D1057–D1062. https://doi.org/10.1093/nar/gkz874

Fischer, D. S., Wu, Y., Schubert, B., & Theis, F. J. (2020). Predicting antigen specificity of single T cells based on TCR CDR 3 regions . *Molecular Systems Biology*, *16*(8), e9416. https://doi.org/10.15252/msb.20199416

Gielis, S., Moris, P., Bittremieux, W., De Neuter, N., Ogunjimi, B., Laukens, K., & Meysman, P. (2019). Detection of Enriched T Cell Epitope Specificity in Full T Cell Receptor Sequence Repertoires. *Frontiers in Immunology*, *10*, 2820. https://doi.org/10.3389/fimmu.2019.02820

Glanville, J., Huang, H., Nau, A., Hatton, O., Wagar, L. E., Rubelt, F., Ji, X., Han, A., Krams, S. M., Pettus, C., Haas, N., Arlehamn, C. S. L., Sette, A., Boyd, S. D., Scriba, T. J., Martinez, O. M., & Davis, M. M. (2017). Identifying specificity groups in the T cell receptor repertoire. *Nature*, *547*(7661), 94–98. https://doi.org/10.1038/nature22976

Hoof, I., Peters, B., Sidney, J., Pedersen, L. E., Sette, A., Lund, O., Buus, S., & Nielsen, M. (2009). NetMHCpan, a method for MHC class i binding prediction beyond humans. *Immunogenetics*,

61(1), 1–13. https://doi.org/10.1007/s00251-008-0341-z

Springer, I., Besser, H., Tickotsky-Moskovitz, N., Dvorkin, S., & Louzoun, Y. (2020). Prediction of Specific TCR-Peptide Binding From Large Dictionaries of TCR-Peptide Pairs. *Frontiers in Immunology*, *11*. https://doi.org/10.3389/fimmu.2020.01803

Vita, R., Mahajan, S., Overton, J. A., Dhanda, S. K., Martini, S., Cantrell, J. R., Wheeler, D. K., Sette, A., & Peters, B. (2019). The Immune Epitope Database (IEDB): 2018 update. *Nucleic Acids Research*, *47*(D1), D339–D343. https://doi.org/10.1093/nar/gky1006

Uncertainty budget for acoustic characterization of Tissue Mimicking Materials (TMMs)

Piero Miloro[1], Raphaela Baesso[1], Aoife Ivory[1], Srinath Rajagopal[1], Bajram Zeqiri[1]

Key words: speed of sound, ultrasonic attenuation, uncertainty

1. Introduction

Accurate knowledge of the acoustic properties of TMMs is important for different medical ultrasound applications such as the development of new diagnostic or therapeutic strategies, one example being ultrasound-mediated hyperthermia. However, methods to measure these properties are not widely standardized and the International Electrotechnical Commission (IEC) has only recently published a technical specification on the matter (IEC TC/87 2019). In this study we present an uncertainty budget for the measurement of the longitudinal Speed of Sound (SoS) and Attenuation (Att) using a through transmission substitution over a frequency range 1 MHz-20 MHz. Results are applicable to liquid, soft and rigid TMMs.

2. Methods

The measurement technique has been described elsewhere (Baêsso et al. 2019). A broadband transducer and a broadband receiver are co-axially aligned and connected to a pulser-receiver. Initially, a water path transmission signal (reference signal) is acquired. The sample is inserted in the path and the through-transmission, front and back reflection signals acquired. Three samples of different thickness are measured at four transmitter-receiver separations. All the measurands (thickness, SoS and Att) are calculated from the acoustic signal (He 2000). Liquid samples are held in a cylindrical cell and sealed with using 12 µm polyethylene terephthalate acoustic windows. The uncertainty budget was evaluated for thickness, SoS and Att. For thickness, currently only Type A uncertainty is considered. Weighted averages and standard deviations are used over the range of frequencies. For SoS, the Type A uncertainty is obtained from the four measurements on the three samples. Type B comes from different sources. Oscilloscope dependent uncertainty sources are time base and sampling resolution. Uncertainty generated by the SoS in water depends on temperature (resolution 0.1 °C), while variation of SoS of the test material with temperature is taken from literature when not measured. The effect of thickness is based on its Type A uncertainty. Diffraction contributes in both reflection and transmitted mode. The former is equivalent to the experiments of (Khimunin 1975), where diffraction errors

[1] Ultrasound and Underwater Acoustics, National Physical Laboratory, Teddington, UK, e-mail: piero.miloro@npl.co.uk

are calculated for two identical transducers based on frequency and geometry only. The latter contribution is derived from (Zeqiri 1996) using the data from an axial scan carried out under the same combination of transmitter and receiver used for the measurements. For attenuation, Type A uncertainty is obtained from the three transmission losses and considering uncertainties in thickness and attenuation in water. Type B uncertainties depend on the oscilloscope (linearity and range to range corrections), Signal-to-Noise Ratio of the receiver and the linearity of the pulser-receiver amplifier. Other Type B sources include temperature dependence of the material, non-linear propagation, and diffraction correction for the transmitted pulse, obtained similar to the method described above. For both SoS and Att, Type A and Type B uncertainties are combined, and a factor k=2 is applied to obtain a coverage probability of 95%, under the assumption that the measurand can be characterized by a normal distribution.

3. Application of the method and discussion

The uncertainty budget has been calculated and applied to liquids, soft and bone TMMs. Examples are reported in Table 1. The uncertainty contributions for the soft TMM are reported in Table 2. The uncertainty budget considers the values of thickness, SoS and Att to be non-correlated, a further study would include the correlation between the measurands.

Table 1: SoS and Att for different materials at 5 MHz and 19 °C at 95 % confidence interval

Material	***SoS* (m/s)**	***Att* (dB/cm)**
Castor Oil	1523 ± 5	12.2 ± 1.2
IEC Agar TMM	1535 ± 2	2.56 ± 0.14
Delrin	2487 ± 2	24.7 ± 0.9

Table 2: Percentage contribution for the uncertainty components on a soft TMM at 5 MHz and 19 °C

SoS	*Type A*	*Time base*	*Resolution*	*SoS water*	*Thic kness*	*Temper ature*	*Diffr. Refl.*	*Diffr. Trans.*	*Expanded*
	0.02	<0.01	0.02	0.01	0.01	0.06	<0.01	<0.01	0.13
Att	*Type A*	*Lineari ty*	*Range to Range*	*SNR*	*Ampl ifier*	*Temper ature*	*Non-linear*	*Diffr.*	*Expanded*
	0.64	0.12	1.7	<0.01	0.57	1.46	1.13	0.23	5.5

References

1. Baêsso, Raphaela M., et al. 2019. “Ultrasonic Parameter Measurement as a Means of Assessing the Quality of Biodiesel Production.” Fuel 241: 155–63.
2. He, Ping. 2000. “Measurement of Acoustic Dispersion Using Both Transmitted and Reflected Pulses.” The Journal of the Acoustical Society of America 107(2): 801–7.
3. IEC TC/87. 2019. IEC TS 63081:2019 Ultrasonics – Methods for the Characterization of the Ultrasonic Properties of Materials.
4. Khimunin, A. S. 1975. “Diffraction Correction of Phase Velocity.” Acustica 32: 192–200.
5. Zeqiri, Bajram. 1996. “Validation of a Diffraction Correction Model for Through-transmission Substitution Measurements of Ultrasonic Absorption and Phase Velocity.” The Journal of the Acoustical Society of America 99(2): 996–1001.

Applying deep learning in metrology

an overview over some potentials and challenges

Jörg Martin[1*], Narbota Amanova[*], Lara Hoffmann[*], Franko Schmähling[*], Clemens Elster[*]

Key words: Artificial intelligence, deep learning, metrology, uncertainty

1. Motivation and background

In the recent years artificial neural networks have become an increasingly popular tool in a variety of areas from language processing, the prediction of customer preferences over medical analysis, physical and chemical modeling to autonomous vehicles - to name only a few. Often, the best results can be achieved with deep learning, that is with networks comprising many layers of neurons. In fact, the success of deep learning has been so extensive that the actual much broader term of artificial intelligence is nowadays often used as a synonym for deep learning.

In metrology the potential of deep neural networks lies in their flexibility and model agnostic nature which makes them a promising tool for cases where the underlying physical model is hard or even impossible to compute. In this talk an overview is given on the basic concepts behind deep learning, its potential applicability, and the challenges for metrology with a particular focus on the work in our working group “Data analysis and measurement uncertainty” at PTB.

2. Potentials and Challenges

Using deep learning for several years our working group has collected expertise in the application of deep neural networks in various fields, learned about obstacles that arise in their application and possible methods to overcome them. One application presented in this talk is the quality assessment of mammography [Kretz(2020a), Kretz(2020b)]. Deep learning can help to simplify and accelerate this procedure and further to equip the prediction with an according uncertainty. Another application mentioned will consider solving a large-scale inverse problem in optics [Hoffmann(2020), Hoffmann(2021)].

The flexibility that allows for such a diverse usage of deep learning comes with its drawbacks. Understanding why a neural network produces a certain prediction and whether, or in which cases, such a prediction can be trusted is one of the main challenges that has to be overcome to ease the uptake of deep learning. Various important concepts such as uncertainty and explainability are presented in this talk

[1] e-mail: joerg.martin@ptb.de

[*] Physikalisch-Technische-Bundesanstalt, Abbestr. 2, 10587 Berlin, Germany

together with the results of several methods studied and/or developed at PTB [Martin(2020a), Martin(2020b)]. We also compare current approaches to uncertainty quantification in deep learning within the context of simple regression problems.

References

1. L. Hoffmann and C. Elster (2020). Deep Neural Networks for Computational Optical Form Measurements, Journal of Sensors and Sensor Systems, 9 301--307. [DOI: 10.5194/jsss-9-301-2020]
2. L. Hoffmann, I. Fortmeier and C. Elster (2021). Uncertainty Quantification by Ensemble Learning for Computational Optical Form Measurements, preprint (https://arxiv.org/abs/2103.01259)
3. T. Kretz (2020). Development of model observers for quantitative assessment of mammography image quality, PhD Thesis. [DOI: 10.14279/depositonce-10552]
4. T. Kretz, K.-R. Müller, T. Schäffter and C. Elster (2020). Mammography Image Quality Assurance Using Deep Learning, IEEE Transactions on Biomedical Engineering, 2020. [DOI: 10.1109/TBME.2020.2983539].
5. J. Martin and C. Elster (2020). Inspecting adversarial examples using the fisher information, Neurocomputing, 382 80--86, 2020. [DOI: 10.1016/j.neucom.2019.11.052].
6. J. Martin and C. Elster (2020). Detecting unusual input to neural networks, Appl Intell, in press.

Limitations of uncertainty propagation – Measurement uncertainty for the routine determination of *aqua regia* extractable metals in soil

Stephen L R Ellison

Key words: Measurement uncertainty, ISO 21748

The application of the 'law of propagation of uncertainty' as described in the GUM [JCGM(2008)] can be problematic in routine testing. For simplicity in application, the calculations given in test standards often fall far short of a complete measurement model. Conditions of measurement are usually held within relatively narrow ranges so that effects on the result are small, but the particular effect of a change within the permitted range is rarely known exactly and can vary from one test item to another. There is rarely a reliable, or even known, mathematical model for the effects of changes in measurement conditions. In addition, a particular laboratory implementation may show a consistent bias which is acceptable for the intended use but not negligible.

This presentation will describe the evaluation of measurement uncertainty for the routine determination of acid-extractable toxic metals in soil using a combination of acid extraction and atomic emission spectrometry. The process is a laboratory-specific modification of the *aqua regia* extraction procedure of ISO11466 [ISO(1995)]. The example illustrates the general approach to measurement uncertainty evaluation taken by ISO 21748 [ISO(2017)], which uses information on precision and trueness of a routine test procedure to provide an indication of the measurement uncertainty to be expected from the procedure. The example illustrates the experimental determination of sensitivity coefficients that cannot readily be derived from a mathematical model, and discusses the issues arising in the event of an appreciable bias which is not corrected for when within permitted limits.

Acknowledgements This project has received funding from the EMPIR programme co-financed by the Participating States and from the European Union's Horizon 2020 research and innovation programme.

S. Ellison
LGC Limited, Queens Road, Teddington, Middlesex, UK, e-mail: s.ellison@lgcgroup.com

References

[ISO(1995)] (1995) ISO 11466:1995: Soil quality – Extraction of trace elements soluble in aqua regia. ISO, International Organization for Standardization, Geneva, Switzerland

[ISO(2017)] (2017) ISO 21748:2017 – Guidance for the use of repeatability, reproducibility and trueness estimates in measurement uncertainty evaluation. ISO, International Organization for Standardization, Geneva, Switzerland

[JCGM(2008)] JCGM (2008) Guide to the Expression of Uncertainty in Measurement, JCGM 100:2008, GUM 1995 with minor corrections. BIPM

Simulation guided design of a TEM applicator for *in vitro* RF hyperthermia

Ioannis Androulakis[1], Riccardo Ferrero[2], Alessandra Manzin[3], Gerard van Rhoon[4]

Key words: RF hyperthermia, heating devices, EM modelling, thermal modelling

1. Introduction

Hyperthermia is one of the most known and potent sensitizers of radiotherapy (Oei[2020]); its sensitizing effect has been demonstrated *in vitro* and *in vivo* as well as in clinical studies. Recently, there have been attempts to better quantify and model this thermal sensitization effect (van Leeuwen[2017]), however there is not yet a generally accepted model that quantifies all aspects of combining hyperthermia and radiotherapy. Further *in vitro* studies are therefore crucial in better understanding the effects of hyperthermia enhanced radiotherapy treatments (Oei[2020]). In this effort, there is a need for affordable, easy to setup and robust radiofrequency (RF) applicators that produce a reproducible and well-focused temperature elevation in three dimensional (3D) cell cultures. Therefore, in this study we designed an RF hyperthermia applicator specifically for this purpose, by means of 3D electromagnetic (EM) and thermal computational modelling.

2. Materials and methods

The design of the RF applicator is based on a coaxial Transverse Electro-Magnetic (TEM) design (Lagendijk[1987]), consisting of an open-ended coaxial line with a hollow inner conductor. The advantage of a coaxial TEM applicator is that the peak of the power deposition is always focused towards the same point on the central cylinder axis, independently of the shape or material properties of the sample. The sharpness and shape of the peak is however dependent on design-specific characteristics of the applicator: the opening between the inner and outer conductor, the coaxial radii, the presence and measures of an inner reflector, the properties of

[1] Ioannis Androulakis
Department of Radiotherapy, Erasmus MC Cancer Institute, Rotterdam, Netherlands, email: i.androulakis@erasmusmc.nl
[2] Riccardo Ferrero
Istituto Nazionale di Ricerca Metrologica (INRIM), Torino, Italy
[3] Alessandra Manzin
Istituto Nazionale di Ricerca Metrologica (INRIM), Torino, Italy
[4] Gerard van Rhoon
Department of Radiotherapy, Erasmus MC Cancer Institute, Rotterdam, Netherlands

the applicator materials and the operating frequency. For the optimization of those parameters 3D electromagnetic and thermal simulation were performed .

3. Results and Conclusion

With the aim of depositing sufficient power to heat a 1 cm^3 sample up to 45 ºC, after a detailed parametric analysis we settled on the design of Figure 1.a, which operates at 433 MHz. In Figures 1.b and 1.c we demonstrate how the optimized applicator is expected to operate in an *in vitro* experiment on cancer cell cultures. Our results clearly illustrate that the applicator design can efficiently and selectively heat small *in vitro* samples up to the required temperature.

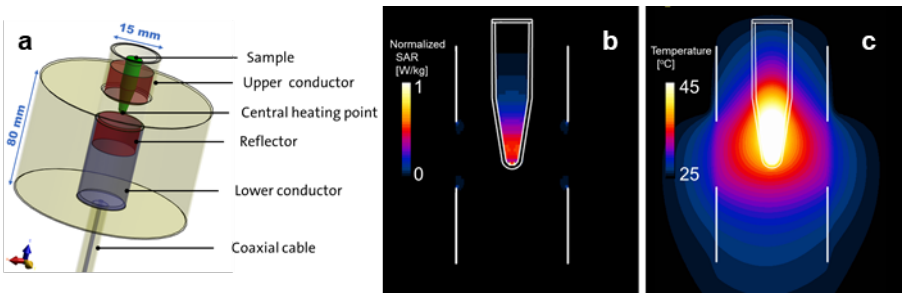


Figure 1: (a) The 3D model of the TEM applicator design with a sample vial positioned in the heating focus area , (b) the normalized specific absorption rate (SAR) distribution in the central horizontal slice of the cylindrical volume and (c) the temperature distribution for the same slice after 20 min of heating with 1.8 W of input power.

References

1. Lagendijk, J.J.W.: A New Coaxial TEM Radiofrequency/Microwave Applicator for Non Invasive Deep-Body Hyperthermia. In: Journal of Microwave Power, 18:4, 367-375 (1983).
2. van Leeuwen, C. M., Crezee, J., Oei, A. L. ,Franken, N. A. P., Stalpers, L. J. A., Bel, A., Kok, H. P. : 3D radiobiological evaluation of combined radiotherapy and hyperthermia treatments. In: International Journal of Hyperthermia, 33:2, 160-169 (2017).
3. Oei, A. L., Kok, H. P., Oei, S. B., Horsman, M. R., Stalpers, L. J. A., Franken, N. A. P., Crezee, J. : Molecular and biological rationale of hyperthermia as radio-and chemosensitizer. In: Advanced drug delivery reviews, 163, 84-97 (2020).

A Gaussian Process approach to uncertainty evaluation for machine learning

James Donlevy, Kavya Jagan and Andrew Thompson

Key words: Gaussian Processes, machine learning, uncertainty evaluation

Introduction

Machine learning offers the potential to widen the reach of metrology even to applications where underlying physical models are either not well understood or difficult to compute. However, the data-driven nature of machine learning brings new challenges around uncertainty evaluation. In this work, we investigate the use of Gaussian Processes (GPs) for uncertainty evaluation in a machine learning context, and for regression problems in particular.

Gaussian Process (GP) regression [2] is a flexible way of modelling data that is not limited by having to specify a parametrized functional form. GPs are typically used to model a variable that has spatial (and/or temporal) dependence. They do so by regarding the values of the variable at different spatial locations as being correlated and describing the strength of the correlation in a flexible way as a function of spatial separation through a covariance kernel. The idea of GP regression is to learn the correlation behaviour from training data comprising measured data collected at known, fixed locations and this knowledge can be used for prediction/interpolation within an area of interest. A particular advantage of GP regression is that the uncertainties associated with predicted values are automatically provided, as they are

James Donlevy
National Physical Laboratory, Hampton Road, Teddington, TW11 0LW, UK e-mail: james.donlevy@npl.co.uk

Kavya Jagan
National Physical Laboratory, Hampton Road, Teddington, TW11 0LW, UKe-mail: kavya.jagan@npl.co.uk

Andrew Thompson
National Physical Laboratory, Hampton Road, Teddington, TW11 0LW, UKe-mail: andrew.thompson@npl.co.uk

a Bayesian approach [1]. The uncertainties for locations close to the training data generally tend to be smaller than for locations far from them.

We investigate the use of GPs in machine learning regression problems, and in particular we explore methods in which uncertainty is decomposed into aleatoric and epistemic uncertainty, and in which uncertainty about hyperparameters is taken into account.

Capturing the different sources of uncertainty

Since machine learning models are data driven, a model that performs well on a particular data set may not work on another. In order to ensure that models are generalisable, care must be taken when evaluating uncertainty to capture the different sources of uncertainty. In machine learning, two important sources of uncertainty can be distinguished: uncertainty in the data and uncertainty about the model. The first of these, sometimes referred to as *aleatoric* uncertainty, is typically random in nature and cannot be removed by specifying a better model. The second is dependent upon the coverage of the training data, and is sometimes referred to as *epistemic* uncertainty. The two combine to give the overall *predictive* uncertainty.

GPs are a Bayesian technique and, as a result, they provide a probability distribution, and hence uncertainty information, associated with the predictions. The diagonal entries of the covariance matrix of the predictions provided by the GP regression are predictive variances. We then subtract the aleatoric contribution, which can be learnt from the training data, from the overall predictive variance to obtain the epistemic variance and hence epistemic uncertainty.

Taking into account uncertainty on hyperparameters

The correlation structure in training data is captured by the covariance kernel of GPs that depends on hyperparameters. In order to make predictions from GPs, these hyperparameters are learnt from the training data. They are then treated as fixed and predictions are made based on their values. Uncertainties associated with the hyperparameter estimates are not usually propagated through to predictions.

We use a Bayesian hierarchical model to capture the lack of perfect knowledge of the hyperparameters. The resulting posterior distribution cannot be defined analytically, and so we use a sampling scheme such as the Metropolis-Hastings (MH) algorithm [1] to generate samples from the distribution of the predictions.

Acknowledgements This work is supported by the UK Government's Department of Business, Energy and Industrial Strategy (BEIS).

References

1. A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin. *Bayesian Data Analysis*. Chapman & Hall/CRC, Boca Raton, Fl., second edition, 2004.
2. C. E. Rasmussen and C. K. I. Williams. *Gaussian Processes for Machine Learning*. MIT Press, Cambridge, Mass., 2006.

The role and use of measurement uncertainty in addressing specification requirements: medical temperature examples

John Greenwood[1] and Maurice Cox[2]

Key words: measurement uncertainty, specification, conformity

1. Summary

Measurement guides and standards are an essential source of information in most areas of testing. However, the terminology used to describe technical requirements in terms of measurement results is often inadequate in situations where conformity decisions must be made that are consistent with the Guide to the expression of uncertainty in measurement, or with the requirements of standards used to support accreditation such as ISO/IEC 17025 and ISO 15189. This example illustrates the issue in terms of two highly regarded guidance documents from the healthcare sector.

2. Introduction

Guidance documents, requirements and standards are found in all areas of testing and measurement. They are written by experts in the technical field with the aim of ensuring consistent and comparable results in the particular activity for which specifications are provided. Unfortunately, it is often not possible to establish whether these requirements are being met in a metrologically robust fashion.

In the worst cases this difficulty occurs because there is no requirement for the evaluation (or even awareness) of measurement uncertainty and requirements are expressed in qualitative terms, such as ‘accuracy’. This practice creates an immediate problem since the Guide to the expression of uncertainty in measurement (GUM) [BIPM et al (2008)] and standards such as ISO/IEC 17025 and ISO 15189 are generally concerned with uncertainty, not accuracy. These are fundamentally different concepts and, as is stated in the International vocabulary of metrology (VIM) [BIPM et al (2012)] and (in annotations to the html version), there is no established methodology for assigning a numerical value to measurement accuracy. Often the only approach that can be taken is to agree an ad hoc

[1] John Greenwood
United Kingdom Accreditation Service, UK, e-mail: john.greenwood@ukas.com
[2] Maurice Cox
National Physical Laboratory, UK, e-mail: maurice.cox@npl.co.uk

interpretation of the requirement, or to supplement the stated requirements with additional requirements such as limits in terms of measurement uncertainty [UKAS (2020)].

3. Examples

Here, two examples are developed in order to illustrate how such guidance and standards requirements might be interpreted. These have not been selected as 'bad' examples; rather, they are chosen because they show that even in otherwise authoritative and carefully drafted guidance there can still be issues of interpretation left to resolve. The aim is to support drafting of future revisions and also to indicate options when working with the existing versions.

The first example concerns the decontamination of medical devices by steam sterilization where the issue is to establish a clear and unambiguous rule to explain the role that measurement uncertainty must take in the decision process. The second example emerges during the mapping and monitoring of storage systems for blood products, where temperature is measured in order to decide whether an alarm condition is met. In this example, what at first may seem to be a clear 'accuracy' requirement is shown to be incomplete, requiring further information or assumptions before conformity can be decided. The discussion and approach in both examples are equally applicable to other guidance and standards documents and to measurements of other quantities.

4. Conclusions

In order to make conformity decisions that are consistent with the GUM and with requirements for accreditation, some additional information, clarification or interpretation is often required.

Such a situation can be avoided in future guidance and standards by (i) suitably defining the role of measurement uncertainty, (ii) avoiding the use of the term 'accuracy', and (iii) ensuring correction is made for any bias in the measurement.

Acknowledgements Health Services Scotland provided valuable input to the development of this example.

References

1. BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP, and OIML. Guide to the expression of uncertainty in measurement, JCGM 100:2008, GUM 1995 with minor corrections. BIPM, 2008.
2. BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP, and OIML. JCGM 200: International vocabulary of metrology Basic and general concepts and associated terms (VIM), 2012.
3. ISO 15189: Medical laboratories — requirements for quality and competence, 2012.
4. ISO/IEC 17025: General requirements for the competence of testing and calibration laboratories, 2017.
5. UKAS. Decision rules and statements of conformity. UKAS LAB 48, Ed 3, 2020.

Modelling of iron oxide nanocubes for magnetic hyperthermia application

Riccardo Ferrero[1], Gabriele Barrera[2], Hüseyn Sözeri[3], Federica Celegato[4], Marco Coïsson[5], Paola Tiberto[6], Alessandra Manzin[7]

Key words: Numerical modelling, Micromagnetic modelling, Thermal simulations, Magnetic nanoparticles, Magnetic hyperthermia

1. Introduction

Hyperthermia is an oncological therapy that can be used to sensitize tumour cells to radiotherapy and chemotherapy. Iron oxide nanoparticles (NPs) have demonstrated to be optimal heat mediators when excited by ac magnetic fields in the range 50 kHz - 1 MHz [Mahmoudi(2018)]. Their capability to release heat can be increased in the presence of hysteresis losses, which typically appear for NP sizes larger than 20 nm and can be properly tuned by varying NP dimension and shape. Additional effects that can influence the hysteresis losses are the magnetostatic interactions among NPs, which strongly depend on their aggregation state.

The presence of many factors makes complex to quantify their contribution to the NP heating performance, requiring a thorough experimental and modelling analysis of both magnetometric and calorimetric properties. Here, we focus the attention on iron oxide NPs with cubic shape and variable size, supporting measurement outcomes with a detailed micromagnetic and thermal modelling analysis.

2. Physics based models

The magnetic behaviour and the hysteresis losses of the iron oxide NPs were evaluated by means of an in-house GPU-parallelized 3D micromagnetic solver, based on the Landau-Lifshitz-Gilbert theory [Ferrero(2021)]. The numerical solver was previously validated by comparison to OOMMF, the micromagnetic code developed at NIST and recognized as a standard by the micromagnetics community.

To support calorimetric measurements, the heating efficiency of the NPs were estimated by means of thermal simulations, performed with an in-house 3D finite element code that solves the heat transfer equation. The code was used to simulate

[1] Riccardo Ferrero
Istituto Nazionale di Ricerca Metrologica, Torino (Italy), r.ferrero@inrim.it
[2,4,5,6,7] Gabriele Barrera, Federica Celegato, Marco Coïsson, Paola Tiberto, Alessandra Manzin
Istituto Nazionale di Ricerca Metrologica, Torino (Italy), g.barrera@inrim.it, f.celegato@inrim.it, m.coisson@inrim.it, p.tiberto@inrim.it, a.manzin@inrim.it
[3] Hüseyn Sözeri
TÜBİTAK UME, Gebze (Turkey), huseyin.sozeri@tubitak.gov.tr

thermal experiments performed on ferrofluids (NPs dispersed in water) contained in test vials and excited by ac magnetic fields [Manzin(2021)].

3. Results

A detailed study of cubic iron oxide NPs, with size in the range 30-200 nm, was performed to elucidate their potential use in magnetic hyperthermia. The NPs were prepared via hydrothermal route and then structurally and dimensionally characterized by XRD, SEM and TEM. Magnetometric measurements were performed and supported by micromagnetic simulations to analyse the hysteresis contribution versus NP size and aggregation state. Finally, calorimetric experimental and modelling characterizations were carried out at 100 kHz, evidencing good heating performance, with specific loss power around 100 W/g.

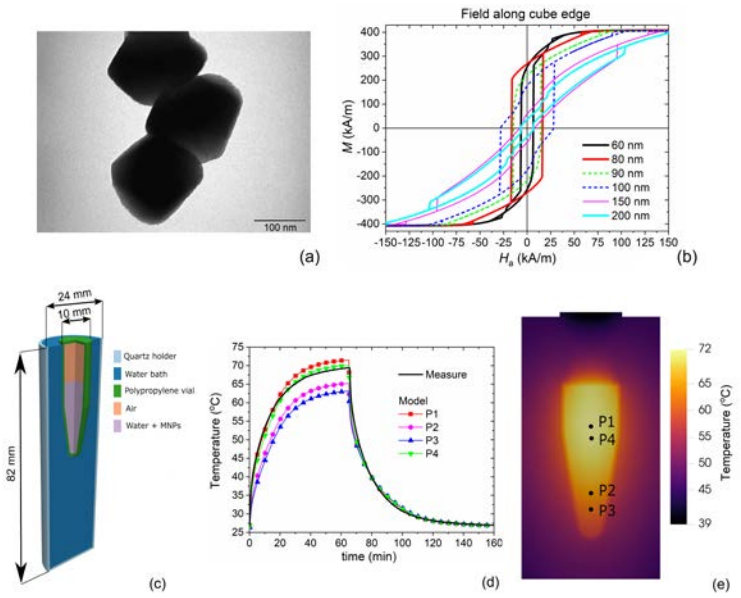


Figure 1: (a) TEM image of cubic iron oxide NPs with ~150 nm size. (b) Hysteresis loops calculated for different NP sizes. (c) Schematic of the calorimetric set-up. (d) Comparison between measured and modelled heating-cooling transients during calorimetric characterization. (e) Calculated spatial distribution of the temperature at the end of the heating interval.

References

1. Ferrero, R., Manzin, A.: Adaptive geometric integration applied to a 3D micromagnetic solver. J. Magn. Magn. Mat. **518**, 167409 (2021).
2. Mahmoudi, K., Bouras, A., Bozec, D., Ivkov, R., Hadjipanayis, C.: Magnetic hyperthermia therapy for the treatment of glioblastoma: a review of the therapy's history, efficacy and application in humans. Int. J. Hyperth. **34**, 1316–1328 (2018).
3. Manzin, A., Ferrero, R., Vicentini, M.: From micromagnetic to *in silico* modeling of magnetic nanodisks for hyperthermia applications. Adv. Theory. Simul. 2100013, 1--13 (2021).

Is There Consistency in ML Interpretability?

Ashish Sundar[1] and Philip Aston[2]

Key words: Interpretability, Explainability, Machine Learning, Random Forest, Boosted Forest, LIME, SHAP

1. Introduction

Interpretability methods aim to generate understanding into how a machine learning model infers its outputs. This is necessary as model complexity is constantly increasing to make use of complex relationships in data, but in doing so the model loses its interpretability.

It is supposed that if a set of features are correlated with the label, all interpretability methods should converge upon these set of features regardless of their differences and of the model used (assuming the model has converged). In this way, it should not matter what interpretability method or black box model is used.

In practice however, this is found not to be true. This work aims to analyse why this is.

2. Method

A sample of 100 patients' ECGs were taken from the PTBDB (Bousseljot et al, 1995), with 50 healthy patients and 50 patients suffering from myocardial infarction. The Symmetric Projection Attractor Reconstruction method (Aston et al, 2020) was applied to the ECGs to obtain attractor images from which 10 bin r and theta densities were derived to use as features for binary classification.

Three techniques were used to classify this dataset – random forests, gradient boosted forests and logistic regression until a high accuracy on the test set was obtained. Two popular model agnostic interpretability techniques (LIME (Lundberg et al, 2017) and SHAP (Riberio et al, 2016))

[1] A. Sundar
Data Science, National Physical Laboratory, Teddington, UK e-mail: ashish.sundar@npl.co.uk
[2] P. J. Aston
Data Science, National Physical Laboratory, Teddington, UK and Department of Mathematics,
University of Surrey, Guildford, UK e-mail: philip.aston@npl.co.uk

were then applied on the models and compared with the models' own feature importances and a Point-Biserial Correlation Coefficient test to analyse the differences, if any, that were present.

Other methods were also used for classification but they failed to converge and thus were discarded.

3. Results and Discussion

It was found that the interpretability methods' performance is limited by the models that they are used on. It was also found that because different models may place value on some features more than others, the weighting of the features differed somewhat between methods. However, it was found that the top n features were consistent across the interpretability methods and the point biserial correlation coefficient, where n = 2 in this problem. n is variable depending on the problem and the dataset, but it is supposed that the strongest features are chosen by the interpretability methods somewhat consistently, even if their relative strengths are different. Thus, it is proposed that n>0 upon convergence of the interpretability method on the model, and the convergence of the underlying model on the dataset.

It was also found that using different interpretability methods on the same model can yield different weightings of features. While this may come as a surprise, the internals of different interpretability methods may bias them toward certain patterns. Thus it is recommended that care is taken when using interpretability methods to explain black box models.

Quite surprisingly, it was also found that boosting / bagging forest models' own feature importance reporting was not entirely accurate.

References

1. PJ Aston, MI Christie, YH Huang and M Nandi. Beyond HRV: attractor reconstruction using the entire cardiovascular waveform data for novel feature extraction. *Phys. Meas.* 39, 024001, 2018.
2. R Bousseljot, D Kreiseler and A Schnabel: Nutzung der EKG-Signaldatenbank CARDIODAT der PTB über das Internet. *Biomedizinische Technik, Band 40, Ergänzungsband 1* (1995) S 317- https://www.physionet.org/content/ptbdb/1.0.0/
3. Ribeiro, M Tulio, S Singh, and C Guestrin. "" Why should I trust you?" Explaining the predictions of any classifier." *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*. 2016.
4. SM Lundberg and SI Lee: A Unified Approach to Interpreting Model Predictions. *Advances in Neural Information Processing Systems*, 2017

Evaluation of measurement uncertainty in totalization of volume measurements in drinking water supply networks

Álvaro Silva Ribeiro[1]

Key words: Measurement uncertainty, water supply networks, water net-balance

1. Introduction and measurement models

Clean water and sanitation are one of the 17 sustainable development goals (SDG) of the United Nations' 2030 agenda [UN (2015)], being related to economic growth, sustainable cities and communities, responsible consumption and production, and climate action. Water utilities rely on water supply networks based on hydrological and hydraulic elements, having a growing demand which increases the risk of scarcity [Reig (2013)]. Measurement of flow, volume, level and velocity are key to provide reliable information to management and to regulators and trust in trade [Borgomeo (2015)]. Major concerns are related to the impact of hidden water losses [Alegre (2016)], water net balance being a major tool for decision making and its uncertainty analysis required to assure the quality of data [BIPM (2008)].

The quantity of (liquid) volume is the measurand of interest in water supply networks, being measured directly, often using totalizer-type flow meters (e.g. turbine flowmeter) or indirectly, using flowmeters to evaluate the rate of a fluid flow during a time interval. Indirect methods enable the measurement of the volume, V, from volumetric flow rate, Q, measured usually at constant time intervals, Δt.

$$V = \sum_{i=1}^{n} Q_i \Delta t, \tag{1}$$

This approach was used to estimate the net-balance of a water supply system [Ribeiro (2019)] with various measurement locations for inflow and outflow,

$$V_{net} = \sum V_x^{(in)} - \sum V_x^{(out)}. \tag{2}$$

2. Uncertainty propagation and interpretation of results

The study uses measurement data of flow rate to obtain the total volume, assuming that the relative standard uncertainties of the individual measurements were approximately equal and three scenarios: constant relative standard uncertainty in

[1] Álvaro Silva Ribeiro
Laboratório Nacional de Engenharia Civil, Av. do Brasil, 101, 1700-066 Lisboa, Portugal, e-mail: asribeiro@lnec.pt

variable flow rate; constant relative standard uncertainty in stationary flow rate; and sampling effect in stationary flow rate.

For each scenario simplified equations were found, allowing to evaluate the measurement uncertainty for total volume estimates without requiring to calculate derivatives and using previous knowledge of the volume measured at each time interval of sampling. Based on these equations, an equation was developed to evaluate the water volume net-balance in an infrastructure, allowing also to conclude that net-balance standard uncertainty neglecting the effect of correlation gives an upper bound, and that the increase of the number of locations used in the net balance will increase the measurement uncertainty of the estimated net volume.

Given the usual complexity of metering systems in infrastructures such as those promoting the management of resources such as water supply, it is very useful to find solutions that can simplify the application of uncertainty calculation considering, on one hand, the importance of integrating this component into the instrumentation metrological management process and, on the other hand, by the relevance that this information may have in the decision-making process.

Acknowledgements

This paper was based on the EMPIR EMUE example A2.2.1, having as co-Authors: M. G. Cox, J. A. Sousa, A. van der Veen, M. Reader-Harris, L. L. Martins, D. Loureiro, M. C. Almeida, M. A. Silva, R. Brito, A. C. Soares.
The project 17NRM05 EMUE has received funding from the EMPIR programme co-financed by the Participating States and from the European Union's Horizon 2020 research and innovation programme".

References

1. Transforming our World: The 2030 Agenda for Sustainable Development. A/RES/70/1. United Nations, New York, 25 – 27 Sept. 2015.
2. Reig, P., Shiao, T., Gassert, F. – AQUEDUCT Water Risk Framework, World Resources Institute, Jan. 2013. Washington DC (USA).
3. Borgomeo, E. – Climate change and water resources: risk-based approaches for decision-making, Thesis presented for the degree of Doctor of Philosophy at the University of Oxford, Christ Church College, Oxford, July 2015.
4. Alegre, H., Baptista, J.F., Cabrera, E., Cubillo, F., Duarte, P., Hirner, W., Parena, R. (2016) – Performance indicators for water supply services, 3rd ed., IWA Publishing, ISBN 9781780406329, London
5. 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.
6. Ribeiro, A. S., Loureiro, D., Almeida, M. C., Cox, M. G., Sousa, J. A., Silva, M. A., Martins, L. L., Brito, R., Soares, A. C. – Uncertainty evaluation of totalization of flow and volume measurements in drinking water supply networks. FLOMEKO 2019. Lisbon, Portugal.

Optimisation of data acquisition for cardiac Magnetic Resonance Fingerprinting

Constance Gatefait[1] and Stephen Ellison[2] and Stephen Nyangoma[3] and Christoph Kolbitsch[4]

Key words: quantitative imaging; uncertainties; magnetic resonance imaging; cardiac imaging

Introduction:

Magnetic Resonance Fingerprinting (MRF) is a quantitative MRI technique providing multiple parameters (T1 and T2 relaxation times), used for medical diagnostics[3,4]. In contrast to classic MRI, acquisition parameters are varying over time to create tissue-dependent temporal signal curves called “fingerprints”. These curves are then matched to a dictionary, simulating the signal evolution for various tissue types (e.g. blood or muscle)[4]. Acquisition parameters encode the fingerprint of each tissue type into the MRF signal, hence their choice strongly determines the accuracy and reliability of the obtained information. In this study we present an assessment of different acquisition parameters for continuous cardiac MRF and their impact on T1 and T2 estimation.

Method:

MRF allows for several degrees of freedom in its design by varying parameters such as the repetition time (TR) and the flip angle (FA) pattern[1,2,3,5]. In this study, we keep TR constant and vary FA along a sinusoidal pattern (Figure 1)[2]. To evaluate and optimize the FA pattern a central composite experiment was designed. We aim to find the optimal combination between the maximum value and the duration of a sinusoidal FA pattern. The experimental design includes a total of 22 scans. Data was acquired with the parameters defined in the design in a standardized phantom made of 9 tubes[1]. Simulated fingerprints are then voxel-wised matched with the acquired data. Dot-products between theoretical fingerprints and the cMRF data are calculated. The highest dot-product value is considered the best match.

Results:

All relative errors are calculated based on results obtained with reference sequences. The best results ($T1_{err}$ = 8.12% ; $T2_{err}$ = 8.34%) were obtained with a maximum FA value of 61.2° and a lobe size of 79. Small maximum FA values, such as 10°, give

[1] Constance Gatefait
Constance Gatefait; Physikalisch-Technische Bundesanstalt, Braunschweig and Berlin, Germany; e-mail: constance.gatefait@ptb.de
[2] Stephen Ellison
Stephen Ellison; LGC Limited, Teddington, Middlesex, UK; e-mail: s.ellison@LGCGroup.com
[3] Stephen Nyangoma
Stephen Nyangoma; LGC Limited, Teddington, Middlesex ,UK; e-mail: stephen.nyangoma@LGCGroup.com
[4] Christoph Kolbitsch
Christoph Kolbitsch; Physikalisch-Technische Bundesanstalt, Braunschweig and Berlin, Germany; e-mail: christoph.kolbitsch@ptb.de

correct results for T1 estimation ($T1_{err}$ = 9.81%) but are not optimal for T2 estimation ($T2_{err}$ > 100%).

Conclusion:

This work presented the results obtained for a parameter evaluation in MRF and the assessment of uncertainties in quantitative mapping of T1 and T2.

Acknowledgements:

The results presented here have been developed in the framework of the 18HLT05 QUIERO Project. This project has received funding from the EMPIR programme co-financed by the Participating States and from the European Union's Horizon 2020 research and innovation programme.

Figures :

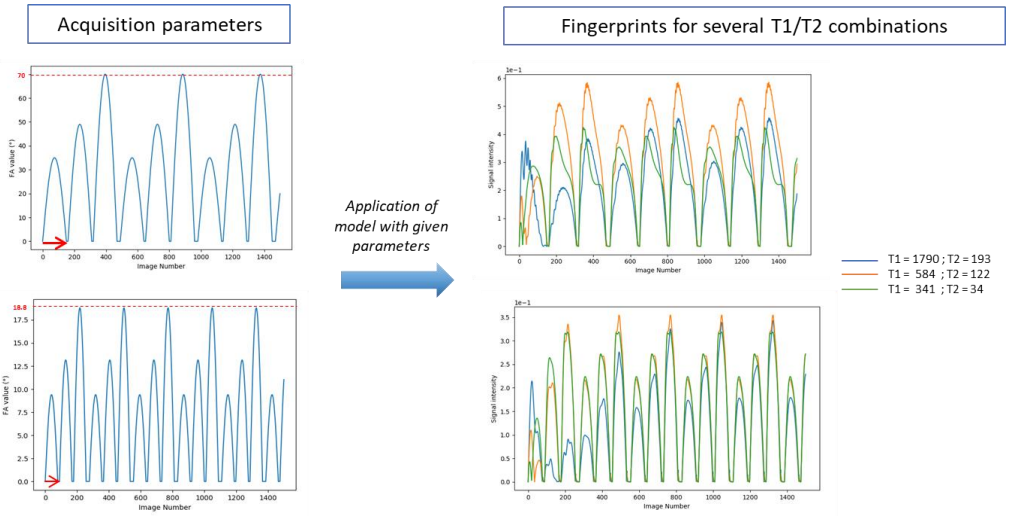


Figure 1 : MRF framework. Acquisition parameters are defined. The dotted red line represents here the max FA and the red arrow is the lobe duration. A physical model is then calculated with the given parameters. Unique fingerprints are then obtained for each T1/T2 combination.

References :

1. Captur et al. A medical device-grade T1 and ECV phantom for global T1 mapping quality assurance - the T1 Mapping and ECV Standardization in cardiovascular magnetic resonance (T1MES) program. J. Cardiovasc. Magn. Reson, 2016
2. Hamilton et al. Investigating and reducing the effects of confounding factors for robust T_1 and T_2 mapping with cardiac MR fingerprinting. MRI. 2018
3. Liu et al. Cardiac magnetic resonance fingerprinting: technical overview and initial results, JACC Cardiovasc.Imaging, 2018
4. Ma et al. Magnetic resonance fingerprinting. Nature 495, 2013
5. Sommer et al. Towards predicting the encoding capability of MR fingerprinting sequences. Magn Reson Imaging. 2017
6. Weigel, Extended phase graphs: Dephasing, RF pulses, and echoes - pure and simple. J.Magn. Reson. Imaging, 2015

The role of uncertainty in data-driven turbulence modelling

Andrea Ferrero, Angelo Iollo, Francesco Larocca and Francesca Romana Pennecchi

Key words: Machine learning, field inversion, uncertainty, turbulence modelling, RANS

1 Abstract

Turbulence modelling represents a critical aspect in the prediction of the flow field inside aerospace propulsion systems. Recently, high-fidelity simulations like Large Eddy Simulations (LES) or Direct Numerical Simulations (DNS) become possible thanks to the constant increase in computational power that has been achieved in the last decades [Sandberg and Michelassi(2019)]. However, these simulations remain prohibitive for performance prediction during a design process because of the large number of configurations which must be investigated. For this reason, high fidelity simulations can be exploited to generate trustworthy solutions on representative test cases in order to understand the phenomena which govern the flow field. Furthermore, it is possible to exploit these results to improve the accuracy of low order models which can then be used for design purposes. In particular, Reynolds-averaged Navier-Stokes (RANS) models represent an efficient way to compute the

Andrea Ferrero
Politecnico di Torino, Corso Duca degli Abruzzi 24, Torino, Italy e-mail: andrea_ferrero@polito.it

Angelo Iollo
INRIA, CNRS, IMB and Université de Bordeaux, 351 Cours de la Libération, Talence, France e-mail: angelo.iollo@inria.fr

Francesco Larocca
Politecnico di Torino, Corso Duca degli Abruzzi 24, Torino, Italy e-mail: francesco.larocca@polito.it

Francesca Romana Pennecchi
INRiM, Strada delle Cacce 91, Torino, Italy e-mail: f.pennecchi@inrim.it

average flow field but they can become quite inaccurate in the presence of separation or transition from laminar to turbulent flow.
In this framework, machine-learning strategies represent a possible approach to improve the predictive capability of existing RANS models starting from the high-fidelity data obtained from LES or experiments [Duraisamy et al(2019)]. Among the different algorithms, field inversion is a promising strategy. The approach, originally introduced by [Parish and Duraisamy(2016)], was exploited to improve RANS models for turbomachinery by [Ferrero et al(2020a)]. The method relies on two steps: an optimisation procedure (the field inversion) and a regression performed by machine learning. The first step requires the definition of an optimisation problem where the goal function is represented by the error between the numerical prediction and the reference data: this error is minimised by finding an optimal field of corrections which alter the source term of the turbulence model. The solution of the optimisation problem contains a lot of information: in each point of the computational domain the local correction and all the fluid variables are known. This makes it possible to exploit machine learning algorithms to identify a correlation between some local flow features and the correction field. This regression step allows to generalise the results and to use the data-augmented RANS model for general predictions.
Even if the first results of the field inversion strategy seem promising, several open questions remain. First of all, the reference data (experimental or from high-fidelity simulations) are affected by uncertainty and it propagates through the field inversion procedure up to the final data-augmented model [Ferrero et al(2020b)]. Furthermore, a significant modelling uncertainty is associated to the regression step: the selection of the flow features which should determine the local correction is not trivial. It is possible to follow some basic guidelines (nondimensional inputs, Galilean invariant inputs,...) but it is not clear how to demonstrate that the correlations captured by the regression analysis are based on a cause-effect principle.

References

[Duraisamy et al(2019)Duraisamy, Iaccarino, and Xiao] Duraisamy K, Iaccarino G, Xiao H (2019) Turbulence modeling in the age of data. Annual Review of Fluid Mechanics 51:357–377

[Ferrero et al(2020a)Ferrero, Iollo, and Larocca] Ferrero A, Iollo A, Larocca F (2020a) Field inversion for data-augmented rans modelling in turbomachinery flows. Computers & Fluids 201:104,474

[Ferrero et al(2020b)Ferrero, Larocca, and Pennecchi] Ferrero A, Larocca F, Pennecchi FR (2020b) Uncertainty propagation in field inversion for turbulence modelling in turbomachinery. In: 2020 IEEE 7th International Workshop on Metrology for AeroSpace (MetroAeroSpace), IEEE, pp 303–308

[Parish and Duraisamy(2016)] Parish EJ, Duraisamy K (2016) A paradigm for data-driven predictive modeling using field inversion and machine learning. Journal of Computational Physics 305:758–774

[Sandberg and Michelassi(2019)] Sandberg RD, Michelassi V (2019) The current state of high-fidelity simulations for main gas path turbomachinery components and their industrial impact. Flow, Turbulence and Combustion 102(4):797–848

Generalization of least square method for straight line regression – A new approach

Jacek Puchalski[1]

Key words: weighted total least-squares, straight line regression, covariance matrix.

1. Introduction

Methods of using least square fit for linear regression are well-known and have been widely applied since the works of C.F. Gauss and A.M Legendre (dating to the eighteenth century). These methods are used in many engineering disciplines including; econometrics, social science, astronomy, geodetics, chemistry, climate science, biology but most important, in metrology. Mathematical methods in OLS version, minimize the sum of the square of distances from requested straight line of experimental points represented by *n*- dimensional vectors of data set $\boldsymbol{X}$ and $\boldsymbol{Y}$ mean vectors of coordinates $\boldsymbol{X}=[x_1,\dots,x_n]^T$ and $\boldsymbol{Y}=[y_1,\dots,y_n]^T$ of measurements points [2]. The straight regression line $\boldsymbol{Y_p} = a\,\boldsymbol{X_p} + \boldsymbol{b}$ where *a*- intercept, *b*- abscissa, $\boldsymbol{b}=b[\boldsymbol{1},\dots,\boldsymbol{1}]^{\boldsymbol{T}}$,where $\boldsymbol{X_p}=\left[x_{p1},\dots,x_{pn}\right]^T$, $\boldsymbol{Y_p}=\left[y_{p1},\dots,y_{pn}\right]^T$ are vectors of coordinates of projecting points, must be fitted closely and according to specific criteria, such as minimizing the objective function G. The proposed method resolves the problem of generalized least square method e.g. WTLS (Weighted Total Least Square). First the input data for the coordinates of the measurement points – vectors of $\boldsymbol{X,Y}$, and the full covariance matrix and additionally limits low and high of the interval range variability of parameter *a* which is proper to straight line regression and finally, the appropriate step Δa are stored in an Excel worksheet. The implementation of the numerical method is applied by a script running in MATLAB, which include, no more than 50 lines of code, and finally plots two one dimensional characteristics: of G(*a*) and G(*b*(*a*)).

2. A new approach

So, in general, when both coordinates are affected by errors and are correlated, 2*n*-dimensional vector $\boldsymbol{Z}=[\boldsymbol{X}^T,\boldsymbol{Y}^T]^T$, with covariance matrix $\boldsymbol{U_Z}$ with diagonal elements: $u^2(z_i)$, and other nondiagonal elements $\rho_{ij}u(z_i)u(z_j)$, where $\rho_{ij}-$ correlation coefficients, $u(z_i)$ and $u(z_j)$ uncertainties at measurement points z_i, $z_j$and *i*,*j*=1,..,2*n* and may be partitioned on four matrixes *n* x *n* sizes: autocorrelations for $\boldsymbol{X}$ e.g. $\boldsymbol{U_X}$, autocorrelations for $\boldsymbol{Y}$ e. g. $\boldsymbol{U_Y}$, cross-correlations $\boldsymbol{XY}$ e.g. $\boldsymbol{U_{XY}}$, $\boldsymbol{U_{XY}^T}$, as well as inverse of covariance matrix $\boldsymbol{U_Z^{-1}}$ can be partitioned on

[1] Jacek Puchalski
Certification Department, Central Office of Measures, Warsaw, Poland,
E-mail: jacek.puchalski@gum.gov.pl

four corresponding matrixes *n* x *n* sizes $\boldsymbol{V_{11}, V_{22}, V_3, V_3^T}$. The proposed method based on algorithm, that can be used in any lab, which finally plots numerical characteristic of objective function and answer the question, "is it possible to minimize objective function at given measurement data with given covariance matrix to estimate the confidence region of straight line regression? The algorithm allows us to numerically determine the covariance matrix $\boldsymbol{U_{ab}}$ and the confidence region for straight line regression [1]. The objective function $G(\boldsymbol{\Delta Z}, a, b) = \boldsymbol{\Delta Z^T U_z^{-1} \Delta Z}$, where $\boldsymbol{\Delta Z} = \boldsymbol{Z_p - Z}$, can be transformed into function of hyper quadratic dependence on $\boldsymbol{\Delta X}$, which has the minimum of $G(a, b)$, and it means that $G(\boldsymbol{\Delta Z}, a, b) \geq G(a, b)$ and

$$G(a, b) = [a\boldsymbol{X} + \boldsymbol{b} - \boldsymbol{Y}]^{\mathrm{T}} \boldsymbol{U_{Yeff}^{-1}} [a\boldsymbol{X} + \boldsymbol{b} - \mathbf{Y}]$$

The minimum of $G(a,b)$ function has the same form as for regression *y* on *x* [4] where effective inverse covariance matrix is: $\boldsymbol{U_{Yeff}^{-1}}(a) = (\boldsymbol{V_{11}V_{22}} - 0{,}25\, \boldsymbol{V_{33}V_{33}}) \cdot \boldsymbol{V^{-1}}(a)$ and $\boldsymbol{V_{33}} = \boldsymbol{V_3^T} + \boldsymbol{V_3}$ and a new matrix $\boldsymbol{V}(a) = \boldsymbol{V_{11}} + a\boldsymbol{V_{33}} + a^2\boldsymbol{V_{22}}$. Both matrixes $\boldsymbol{U_{Yeff}^{-1}}$ and $\boldsymbol{V}(a)$ should be positively definite. The formula for $G(a, b)$ can be minimalized for independent variable b in the same way as for regression *y* on *x* [4], and then $b = b(a)$ and the determination of minimum of G can be reduced to one-dimensional analysis of the numerically calculated function of $G(a)$ because of $G(a, b)) \geq G(a)$. A few interesting facts as a consequence of above minimalization are demonstrated.

3. Numerical examples

The method is tested with three numerical cases considered in previous publication [3]: 1) non correlated case $\boldsymbol{U_X}$, $\boldsymbol{U_Y}$ diagonal with heteroscedasticity, $\boldsymbol{U_{XY}} = 0$; 2) only correlation between x_i and y_i, so $\boldsymbol{U_X}$, $\boldsymbol{U_Y}$, $\boldsymbol{U_{XY}}$ are diagonal; 3) fully correlated all elements of matrices $\boldsymbol{U_X}$, $\boldsymbol{U_X}$ and $\boldsymbol{U_{XY}}$.The numerical method is convergent to the classical cases with analytical solutions for the *a* and *b* parameters and their covariance matrix $\boldsymbol{U_{ab}} = \boldsymbol{C\, U_Z\, C^T}$, where $\boldsymbol{C}$ – numerically determined matrix of sensitivity coefficients. The numerical method is very fast if the low and high level of very narrow interval range of *a* is selected. The method allows to solve any types of regression of two coordinates in Cartesian System with high accuracy of determining *a, b* parameters and it is convergent to the regression *y* on *x* and *x* on *y* where analytical solutions are available and such solutions are given [4].

References

1. 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. JCGM101:2008 Supplement 1. Propagation of distributions using a Monte Carlo method; JCGM102:2011 Supplement 2. Extension to any number of output quantities.
2. Draper R. D., Smith H. Applied Regression Analysis, 3rd Edition Willey New York (1998),
3. Amiri-Simkooei, A.R. et all: Estimation of straight-line parameters with fully correlated coordinates. Measurement 48 (2014) 378–386
4. Warsza, Z.L., Puchalski, J., Uncertainty bands of the regression line for correlated measured data of dependent variable Y, *25th Conference on Automation – Innovations and Future* Warsaw, Poland

Three dimensional MRF obtains highly repeatable and reproducible multi-parametric estimations in the healthy human brain

Guido Buonincontri[1,2] , Jan W. Kurzawski[1,3] , Joshua D Kaggie[4], Tomasz Matys[5], Ferdia A Gallagher[4], Matteo Cencini[1,2], Graziella Donatelli[2,5], Paolo Cecchi[5], Mirco Cosottini[2,5,6], Nicola Martini[7], Francesca Frijia[7], Domenico Montanaro[8], Pedro A. Gómez[2,9], Rolf F Schulte[10], Alessandra Retico[3], Michela Tosetti[1,2]

Key words: MR Fingerprinting, quantitative relaxometry

Abstract

Magnetic resonance fingerprinting (Ma et al., 2013) (MRF) is a novel multi-parametric MR imaging technique, able to provide robust quantification of tissue parameters (e.g. Proton Density PD, longitudinal relaxation time T1 and transverse relaxation time T2) in a relatively short scan time. Recent studies, focused on 2D implementations of MRF technique, have shown high repeatability and reproducibility of MRF measurements both in vitro(Jiang et al., 2017) and in the human brain (Buonincontri et al., 2019a; Körzdörfer et al., 2019). This study aims to evaluate repeatability and reproducibility in the human brain for a 3D MR Fingerprinting acquisition with whole head coverage and isotropic resolution. Therefore, we employed a travelling head study design, acquiring test-retest data for 12 healthy volunteers on 8 MRI scanners, all from the same vendor (3 systems at 3T field strength and 5 systems at 1.5T). Acquisition protocol included a 3D steady-state free precession MRF with spiral projection k-space acquisition (Cao et al., 2019; Gómez et al., 2020) and a 3D FSPGR anatomical scan at matching resolution. Subjects were removed from the scanner between test and re-test. To allow for a General Linear Model (GLM) analysis, each subject was scanned in multiple MR scanners. After registration of the MRF-derived PD T_1 and T_2 maps to an anatomical atlas, coefficients of variation (CVs) were computed to assess test/retest repeatability and inter-site reproducibility in each voxel, while GLM was used to determine the voxel-wise variability

[1]IRCCS Stella Maris, Pisa, Italy; [2]Imago7 Foundation, Pisa, Italy; [3]National Institute for Nuclear Physics (INFN), Pisa, Italy; [4]Department of Radiology, University of Cambridge and Cambridge University Hospitals NHS Foundation Trust, Cambridge, United Kingdom; [5]U.O. Neuroradiologia, Azienda Ospedaliera Universitaria Pisana (AOUP), Pisa, Italy; [6]University of Pisa, Department of Translational Research and New Technologies in Medicine and Surgery, Pisa, Italy; [7]U.O.C. Bioingegneria e Ing. Clinica, Fondazione Toscana Gabriele Monasterio, Pisa, Italy; [8]U.O.C. Risonanza Magnetica Specialistica e Neuroradiologia, Fondazione CNR/Regione Toscana G. Monasterio, Pisa-Massa, Italy; [9]Technical University of Munich, Munich, Germany; [10]GE Healthcare, Munich, Germany

e-mail: matteo.cencini@fsm.unipi.it

between all confounders. The covariates modelled as categorical variables in the GLM analysis were as follows: acquisition variability (test/retest data), field strength (1.5 T and 3 T), subject, and acquisition site.

Our analysis demonstrated a high repeatability (CVs 0.7-1.3% for T1, 2.0-7.8% for T_2, 1.4-2.5% for normalized PD) and reproducibility (CVs of 2.0-5.8% for T_1, 7.4-10.2% for T_2, 5.2-9.2% for normalized PD) in grey and white matter. Reproducibility values were in agreement with a similar analysis performed on 2D MRF data (Buonincontri et al., 2019b), while repeatability values were improved by approximately a factor of two. This could be explained by smaller registration errors and partial volume effect, as well as higher SNR efficiency. The data here presented matched or surpassed other quantitative assessments of repeatability and reproducibility in the literature using other common mapping techniques (Bauer et al., 2010; Deoni et al., 2008; Landman et al., 2011). Iterative reconstructions, the inclusion of B1+ or B1- in the model, and a full characterization of the systems (such as gradient delay and trajectory estimation) may further improve the robustness of these 3D MRF quantifications. In conclusion, three-dimensional MRF with spiral projection k-space trajectory obtains detailed parametric maps with highly repeatable and reproducible PD, T1 and T2 values.

References

1. Bauer, C. M., Jara, H., Killiany, R., & Alzheimer's Disease Neuroimaging Initiative. (2010). Whole brain quantitative T2 MRI across multiple scanners with dual echo FSE: applications to AD, MCI, and normal aging. Neuroimage, 52(2), 508-514.
2. Buonincontri, G., Biagi, L., Retico, A., Cecchi, P., Cosottini, M., Gallagher, F.A., Gómez, P.A., Graves, M.J., McLean, M.A., Riemer, F., Schulte, R.F., Tosetti, M., Zaccagna, F., Kaggie, J.D., 2019a. Multi-site repeatability and reproducibility of MR fingerprinting of the healthy brain at 1.5 and 3.0 T. Neuroimage. https://doi.org/10.1016/j.neuroimage.2019.03.047
3. Buonincontri, G., Biagi, L., Retico, A., Cecchi, P., Cosottini, M., Gallagher, F.A., Gómez, P.A., Graves, M.J., McLean, M.A., Riemer, F., Schulte, R.F., Tosetti, M., Zaccagna, F., Kaggie, J.D., 2019b. Multi-site repeatability and reproducibility of MR fingerprinting of the healthy brain at 1.5 and 3.0 T. Neuroimage 195, 362–372. https://doi.org/10.1016/J.NEUROIMAGE.2019.03.047
4. Cao, X., Ye, | Huihui, Liao, | Congyu, Li, Q., He, H., Zhong, | Jianhui, 2019. Fast 3D brain MR fingerprinting based on multi-axis spiral projection trajectory. Magn Reson Med 82, 289. https://doi.org/10.1002/mrm.27726
5. Deoni, S. C., Williams, S. C., Jezzard, P., Suckling, J., Murphy, D. G., & Jones, D. K. (2008). Standardized structural magnetic resonance imaging in multicentre studies using quantitative T1 and T2 imaging at 1.5 T. Neuroimage, 40(2), 662-671.
6. Gómez, P.A., Cencini, M., Golbabaee, M., Schulte, R.F., Pirkl, C., Horvath, I., Fallo, G., Peretti, L., Tosetti, M., Menze, B.H., Buonincontri, G., 2020. Rapid three-dimensional multiparametric MRI with quantitative transient-state imaging. Sci. Rep. https://doi.org/10.1038/s41598-020-70789-2
7. Jiang, Y., Ma, D., Keenan, K.E., Stupic, K.F., Gulani, V., Griswold, M.A., 2017. Repeatability of magnetic resonance fingerprinting T1 and T2 estimates assessed using the ISMRM/NIST MRI system phantom. Magn. Reson. Med. 78, 1452–1457. https://doi.org/10.1002/mrm.26509
8. Körzdörfer, G., Kirsch, R., Liu, K., Pfeuffer, J., Hensel, B., Jiang, Y., Ma, D., Gratz, M., Bär, P., Bogner, W., Springer, E., Cardoso, P.L., Umutlu, L., Trattnig, S., Griswold, M., Gulani, V., Nittka, M., 2019. Reproducibility and repeatability of MR fingerprinting relaxometry in the human brain. Radiology. https://doi.org/10.1148/radiol.2019182360
9. Landman, B. A., Huang, A. J., Gifford, A., Vikram, D. S., Lim, I. A. L., Farrell, J. A., ... & van Zijl, P. C. (2011). Multi-parametric neuroimaging reproducibility: a 3-T resource study. Neuroimage, 54(4), 2854-2866.
10. Ma, D., Gulani, V., Seiberlich, N., Liu, K., Sunshine, J.L., Duerk, J.L., Griswold, M.A., 2013. Magnetic resonance fingerprinting. Nature. https://doi.org/10.1038/nature11971

Forest embeddings for gene expression data modeling of tumor stage and survival in bladder cancer

Mauro Nascimben, Manolo Venturin and Lia Rimondini

Key words: data-driven biomarker research, statistical modeling, non-linear dimension reduction

In previous work, we introduced double discretization procedures to analyze gene expression data applicable to bladder cancer survival rate modeling [Nascimben et al, 2021]. However, despite good performance of the algorithms proposed (best three models achieved an accuracy of 0.98, 0.87 and 0.85 at cross-validation), dimensionality of the original dataset could not be reduced less than 5 components with neighbourhood components analysis.

In present work, tumor stage was added to the model enlarging the number of classes leading to a multiclass problem, and manifold learning techniques for dimensionality reduction were attempted in order to obtain a meaningful visualization of the dataset. Our numerical experiments evaluated which technique:

- could be integrated in our previous double discretization approach.
- offers better bi-component embedding of genetic profiles also in presence of unlabelled observations (partial learning).

Annotated medical data collection is costly and time consuming, for this reason any procedure that reduces the amount of labelled data needed should be carefully considered. In addition, tracking bladder cancer evolution over an euclidean plane could facilitate bio-statistical model interpretation by medical doctors.

Mauro Nascimben
Enginsoft SpA, Padova, Italy
Department of Health Sciences, University of Eastern Piedmont, Novara, Italy

Manolo Venturin
Enginsoft SpA, Padova, Italy

Lia Rimondini
Department of Health Sciences, University of Eastern Piedmont, Novara, Italy

Working procedure included creation of an optimized number of extra-trees (best random split at each node) to obtain a dissimilarity matrix by counting pairwise co-occurrences in the leaves. Subsequently, dissimilarity matrix was processed by two visualization techniques, t-distributed Stochastic Neighbor Embedding [Van Der Maaten, 2014] and Uniform Manifold Approximation and Projection [McInnes et al, 2018]. Notably, latter method is a promising technique applicable in the field of data-driven life science analysis, as shown recently on Omics (e.g. [Do and Canzar, 2021], [Gulati et al, 2020]). Finally, labels were re-calculated clustering the point clouds and goodness of the embedding evaluated for each experimental condition.

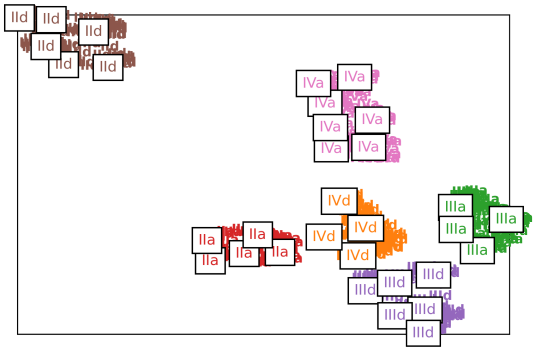


Fig. 1 Forest embedding and bi-component reduction of 25 genes for bladder cancer modeling (tumor stage II,III, IV and outcome "alive" or "dead"). In this example data was mapped to a normal distribution as pre-binning stage by means of a quantile function followed by CART discretization.

References

Do VH, Canzar S (2021) A generalization of t-sne and umap to single-cell multimodal omics. bioRxiv

Gulati S, Wang S, Parikh S, Liu B, Gokhale K, Manivannan M, Sahu S, Kim D, Parikh A (2020) Using machine learning to detect heterogeneity in single cell multi-omics datasets. In: Cancer Research, 16, vol 80

McInnes L, Healy J, Melville J (2018) Umap: Uniform manifold approximation and projection for dimension reduction. arXiv preprint arXiv:180203426

Nascimben M, Venturin M, Rimondini L (2021) Double stage discretization approaches for biomarker based bladder cancer survival modeling. Communications in Applied and Industrial Mathematics (submitted)

Van Der Maaten L (2014) Accelerating t-sne using tree-based algorithms. The Journal of Machine Learning Research 15(1):3221–3245

Tensor based Modelling of Human Motion in a Collaborative Human-Robot Approach

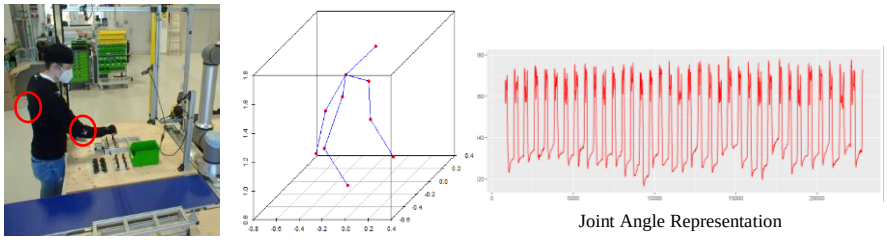
Philipp Wedenig[1], Lorena Gril[1] and Ulrike Kleb[1]

Key words: Human-Robot Collaboration, Human Motion Prediction, Tensor-on-Tensor Regression

1. Background and Data Base

Collaborative robotic systems are a key technology for future industrial applications. To guarantee the safety of humans, commercially available robotic systems rely on direct physical contact to the co-working person in order to detect hazardous situations, opposed to those systems equipped with predictive capabilities. Based on data of a motion tracking sensor system, the latter might be able to predict potential episodes, where the human and the robot might collide. Consequently, the robotic systems can react by adjusting the robots' speed or position to avoid unwanted physical contact. Common approaches make use of Artificial Neural Networks for predicting human motion, but have the disadvantage to need a high amount of training data [Martinez (2017)]. Due to the small training data set, the aim of this work is to perform human motion prediction of a repetitive assembly task using Tensor-on-Tensor regression [Lock (2018)].

For this purpose, human motion data of an assembling task, repeatedly performed by one person, was recorded by the OptiTrack motion capture system, leading to virtually equal cycles of positional data. To detect the human motion, infrared reflective markers have to be placed on corresponding joints of the human torso (marked as red dots at the left image in Figure 1). The OptiTrack system provides for each marker unique traceable Cartesian coordinates (x, y, z) over time, see Figure 1.



Infrared Reflective Markers | Cartesian coordinate Representation | Joint Angle Representation

Figure 1: Data Set Representation

[1]JOANNEUM RESEARCH Forschungsgesellschaft mbH - POLICIES,
Leonhardstraße 59, 8010 Graz, Austria
E-mail address: philipp.wedenig@joanneum.at

Subsequently, the dataset of joint positions was transformed into the joint angle space, which is computed from the positional joints considering a kinematic model, and applying the inverse kinematic to obtain the angles of joint points. The representation in the joint angle space considers the relative angle between adjacent joints, thus enabling the incorporation of physical limitations of the human body.

2. Modelling

To build the predictive model, a reference cycle of one complete assembling task was generated based on some selected cycles of the recorded data. To identify the individual cycles, Fast Fourier Transformation was applied on one selected representative angle of the joint angle space, returning a sufficiently smooth process over time. In our setting, the peaks of the transformed process (see right image in Figure 1) denote the end of one screwing movement, with four peaks representing one complete cycle. After stretching or compressing to equal length, the mean of all cycles was computed for each time step, resulting in the reference cycle.

As the data of the joint angle space is high dimensional and strongly correlated, a Tensor-on-Tensor regression approach [Lock (2018)] was used to get a predictive model based on the reference cycle. Estimates of coefficient tensors are generated with the R-package "MultiwayRegression" by using input data of k seconds of the past in order to predict the motion for l second in the future. The contracted tensor product for the linear prediction of an outcome array $\mathcal{Y}$ from the predictor array $\mathcal{X}$ is defined as

$$\mathcal{Y} = \langle \mathcal{X}, \mathcal{B} \rangle + \mathcal{E},$$

where $\mathcal{B}$ is the coefficient tensor and $\mathcal{E}$ the error term. The coefficient tensors are computed over a certain number of frames, depending on the frame rate. Applying the coefficient tensors to newly observed data, the model is able to predict the future motion pattern of a new cycle. In order to consider varying working speed, resulting in different length of the individual cycles, Dynamic Time Warping is applied to identify the best fitting coefficient tensor.

3. Results and Conclusion

First results of the Tensor-on-Tensor regression model are very promising for receiving multivariate predictions of highly correlated data in real-time. Further work will focus on optimizing the procedure to increase the prediction performance.

Acknowledgment and References

This work was funded by the Austrian Federal Ministry for Climate Action, Environment, Energy, Mobility, Innovation and Technology (BMK).

1. Martinez. J., Black, M. J., and Romero J. On human motion prediction using recurrent neural networks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), vol. 30, pp. 4674-4683, 2017.
2. Lock, E.: Tensor-on-Tensor Regression. Journal of Computationals and Graphical Statistics, vol. 27, pp. 638-647, 2018

Towards a Metrological Characterisation of Electric Properties Tomography

Alessandro Arduino[1], Sébastien Marmin[2], Haris Lulić [3], Alen Bošnjaković[3], Oriano Bottauscio[1], Vedran Karahodžić[3], Umberto Zanovello[1], Luca Zilberti[1]

Key words: Magnetic resonance imaging, electric properties tomography, quantitative imaging, uncertainty propagation

Electric properties tomography (EPT) is a quantitative imaging technique based on magnetic resonance imaging (MRI) that maps the distribution of the electric properties (EPs), conductivity σ and permittivity ε, within the human body in a non-invasive way. Different strategies have been proposed in the literature to implement the EPT technique [Liu(2017)]. Here, an open source tool for the application of three EPT implementations is presented and its usage for the metrological characterisation of the implemented strategies is described.

EPTlib is an open source, extensible C++ library collecting EPT methods. It is distributed through a Github repository that can be reached from the library website (https://eptlib.github.io/) and includes a terminal application for the immediate adoption of the implemented algorithms, which are Helmholtz-EPT (H-EPT) [Katscher(2009)], convection-reaction–EPT (CR-EPT) [Hafalir(2014)] and gradient-EPT (G-EPT) [Liu(2015)]. H-EPT and CR-EPT estimate σ by processing the transceive phase map acquired by the MRI scanner under the assumption of an almost homogeneous transmit sensitivity, like in the case of a 1.5 T birdcage coil (cf. Fig. 1). G-EPT, instead, estimates the EPs by elaborating the transceive phase and the transmit sensitivity maps associated to each channel of a parallel transmis-

[1]Istituto Nazionale di Ricerca Metrologica (INRiM), Torino, ITALY
[2]Laboratoire national de métrologie et d'essais (LNE), Trappes, FRANCE
[3]Institute of Metrology of Bosnia and Herzegovina (IMBIH), Sarajevo, BOSNIA AND HERZEGOVINA
e-mail: a.arduino@inrim.it

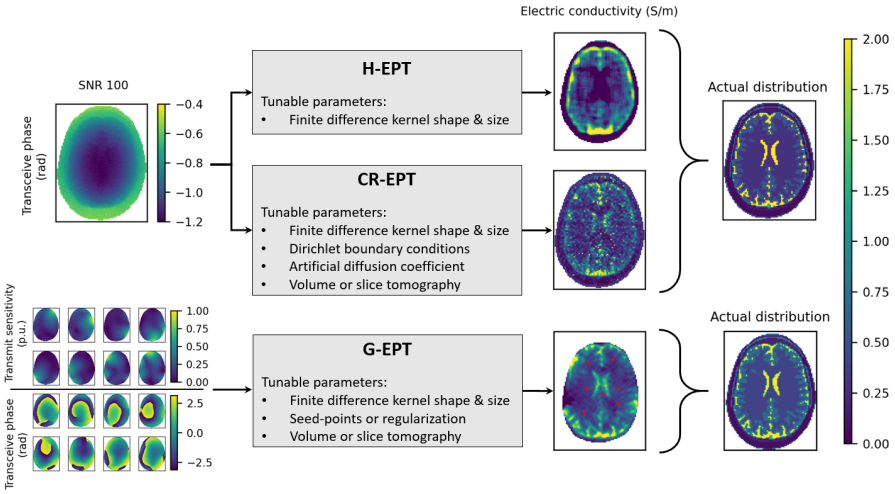


Fig. 1 Diagram of the analyzed EPT methods. H-EPT and CR-EPT elaborate the transceive phase of a 1.5 T birdcage coil to estimate σ at 64 MHz. G-EPT elaborates the transmit sensitivities and the transceive phases of a 7 T 8-channel head coil to estimate σ at 300 MHz. The input maps are corrupted by a spatially uncorrelated random error with SNR 100.

sion (pTx) system, like in the case of a 7 T 8-channel head coil (cf. Fig. 1). In this study, the required input maps are accurately simulated by methods for computational electromagnetism, which take into account the distribution of the EPs within a human head. The resulting maps are then corrupted with realistic noise sampled in accordance with the pulse sequence that would be used by the MRI scanner to acquire them. Figure 1 shows possible input and output maps of the EPT methods for a given selection of the tunable parameters.

During the talk, an *in silico* characterisation of the EPT methods, and the effect of their tunable parameters, will be presented in terms of bias and uncertainty propagation.

Acknowledgements The results here presented have been developed in the framework of the EMPIR Project 18HLT05 QUIERO. This project 18HLT05 QUIERO has received funding from the EMPIR programme co-financed by the Participating States and from the European Union's Horizon 2020 research and innovation programme.

References

Hafalir F, et al (2014) IEEE Trans Med Imaging 33:777–793
Katscher U, et al (2009) IEEE Trans Med Imaging 28:1365–1374
Liu J, et al (2015) Magn Reson Med 74:634–646
Liu J, et al (2017) IEEE Trans Biomed Eng 64:2515–2530

Convolutional neural network performance in the presence of physiological ECG noise

J. Venton, P.M. Harris, A. Sundar, N.A.S. Smith and P.J. Aston

Key words: electrocardiogram, physiological noise, deep learning, robustness

1 Introduction

An electrocardiogram (ECG) signal is a recording of the heart's electrical activity and has long been used to support the diagnosis of cardiovascular disorders. In recent years, deep learning methods of ECG classification have seen a rapid increase in success and popularity (Hong et al, 2020).

Robustness is a property of a machine learning model that describes how well it performs when input data is noisy. ECG signals are susceptible to physiological noise that originates from the subject, yet little work has been carried out to understand the robustness of deep networks to physiological ECG noise.

2 Method

A dataset of 2678 lead II ECG signals was used, with three diagnostic classes: 976 atrial fibrillation, 918 healthy and 784 ST depression (Venton et al, 2021). The dataset was filtered to remove noise to produce a clean dataset, then recorded physiological noise (Moody et al, 1984) was applied with a signal-to-noise ratio between 5 dB and 10 dB to create a noisy dataset (Figure 1).

Two ECG image transformations, Symmetric Projection Attractor Reconstruction (SPAR) (Aston et al, 2018) and scalogram, were applied to the datasets (Figure 1). Transfer learning was used to adapt a pretrained convolutional neural network (CNN) to classify the images corresponding to the clean and noisy datasets. Five fold cross validation was used and network performance was evaluated using the F1 score. To assess CNN robustness to physiological ECG noise, a network trained on the clean dataset was used to classify data from the noisy dataset and vice versa.

J. Venton
Data Science, National Physical Laboratory, Teddington, UK, e-mail: jenny.venton@npl.co.uk

P. M. Harris
Data Science, National Physical Laboratory, Teddington, UK e-mail: peter.harris@npl.co.uk

A. Sundar
Data Science, National Physical Laboratory, Teddington, UK e-mail: ashish.sundar@npl.co.uk

N. A. S. Smith
Data Science, National Physical Laboratory, Teddington, UK e-mail: nadia.smith@npl.co.uk

P. J. Aston
Data Science, National Physical Laboratory, Teddington, UK and Department of Mathematics, University of Surrey, Guildford, UK e-mail: philip.aston@npl.co.uk

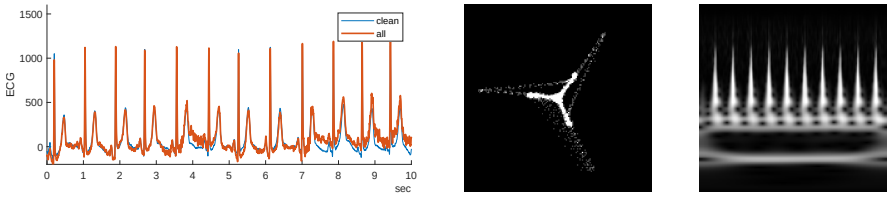


Fig. 1 Left: A typical ECG signal from the dataset, showing the cleaned signal (clean) and the signal after noise had been applied (all). Middle: A typical SPAR attractor generated using an ECG signal. Right: A typical scalogram generated using an ECG signal.

3 Results & Discussion

A summary of the F1 scores can be seen in Table 1. Performance was best for both image types when both training and test data were taken from the clean dataset and decreased slightly (< 0.05) when data were taken from the noisy dataset.

Table 1 Mean and standard deviation F1 scores across the five cross validation folds. Attractor and scalogram results shown for different clean or noisy combinations of the training and test data.

Training data:	clean	noisy	clean	noisy
Test data:	clean	noisy	noisy	clean
attractor	0.70 (0.02)	0.65 (0.02)	0.53 (0.05)	0.60 (0.04)
scalogram	0.79 (0.02)	0.77 (0.03)	0.65 (0.01)	0.74 (0.03)

Significantly, the F1 score penalty for using a network trained on noisy data to classify clean data was lower than the penalty for using a network trained on clean data to classify noisy data. This suggests that a CNN trained on clean ECG data image transforms has poor robustness to physiological ECG noise, and that robustness can be improved by including physiological ECG noise in the training data.

Acknowledgements This project 18HLT07 MedalCare has received funding from the EMPIR programme co-financed by the Participating States and from the European Union's Horizon 2020 research and innovation programme.

References

Aston P, Christie M, Huang Y, Nandi M (2018) Beyond hrv: Attractor reconstruction using the entire cardiovascular waveform data for novel feature extraction. Phys Meas 39

Hong S, Zhou Y, Shang J, Xiao C, Sun J (2020) Opportunities and challenges of deep learning methods for electrocardiogram data: A systematic review. Comp Biol Med 122

Moody G, Muldrow W, Mark R (1984) A noise stress test for arrhythmia detectors. Comput Cardiol 11

Venton J, Harris P, Sundar A, Smith N, Aston P (2021) Robustness of convolutional neural networks to physiological ecg noise. submitted to Philos T R Soc A

Modeling Pyroelectric Sensor Signals for Predicting Proximity

Franz Moser[1] and Ulrike Kleb[2]

Key words: Pyroelectric Signal, Proximity Sensor, Functional Regression

1. Background and Data

Ensuring human's safety during work is crucial in human-robot collaboration. Whereas many existing monitoring systems for avoiding collisions between humans and robots are derived from image-processing concepts, our approach utilizes the PyzoFlex® technology [Rendl(2012)] for a contactless distance measurement. The Pyzoflex® sensors are printed on flexible crystalline polymer foils and due to their pyroelectric activity they can detect an approaching heat source, e.g. a human. Technically speaking, a change in the heat source's position effects a temperature change inside the sensors polymer and this induces an electrical signal [Ferrari(2012)]. The aim of this work is to make use of the sensor's pyroelectricity for developing a soft sensor model working as a proximity sensor.

The available information comprises several data sources of both motion parameters based on a DoE (varying in speed and temperature) and the corresponding sensor signals. Each experiment is carried out in an experimental test setup where a heat source is mounted on a specific rail vehicle. Moving the vehicle alternating both forward and backward generates sensor signals of a PyzoFlex® sensor array with 15 channels (**Figure 1**). As the signals are weak, show an electrical drift and are quite noisy, the data requires some pre-processing such as sensor calibration and signal filtering. For filtering, a discrete wavelet transformation [Berkel(2010)] was performed. Left side of **Figure 2** exemplarily depicts such smoothed signals over time.

2. Modeling and Results

The complete motion dynamics and the signals are synchronized in time and hence they can be considered as curves. This aspect suggests Functional Data Analysis (FDA) [Ramsay(2005)], which is suitable if the units of observation are curves and

[1] Franz Moser, JOANNEUM RESEARCH Forschungsgesellschaft mbH – POLICIES, e-mail: Franz.Moser@joanneum.at
[2] Ulrike Kleb, JOANNEUM RESEARCH Forschungsgesellschaft mbH – POLICIES, e-mail: Ulrike.Kleb@joanneum.at

functional dependencies arise. We developed two FDA modeling approaches using the R package FDboost [Brockhaus(2017)].

(a)  (b) 

Figure 1: (a) Heat source and the layout of the sensor array; **(b)** PyzoFlex sensor matrix

The first model predicts the signals based on the motion parameters, considering the known physical laws behind signal generation. This results in a function-on-scalar model and enables signal curve predictions based on new experimental setups. In the second step, a function-on-function approach, the time-dependent distance curve is modeled as a function of the sensor signal curves (**Figure 2**).

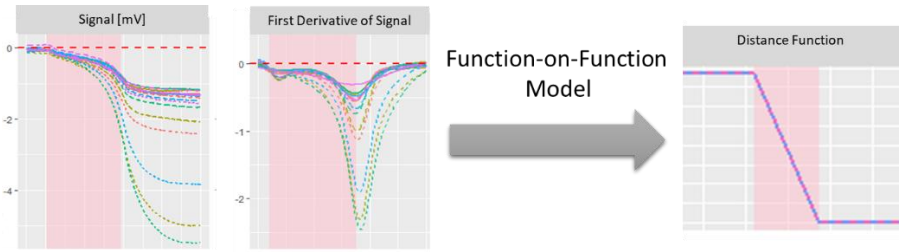


Figure 2: Electrical signals (left) and linear distance over time (right); Red boxes denote motion phases

Further work will focus on predicting the distance of the heat source in real-time. For future applications, this proximity soft sensor might be placed on a robot arm to ensure safety distance.

This work is part of a project funded by the Austrian Federal Ministry for Climate Action, Environment, Energy, Mobility, Innovation and Technology (BMK).

References

1. Berkel, van M Matthijs et al. (2010): Wavelets for feature detection. A theoretical background.
2. Brockhaus, Sarah & Rügamer, David & Greven, Sonja. (2017). Boosting Functional Regression Models with FDboost.
3. Ferrari, V. (8 2012). Printed thick-film piezoelectric and pyroelectric sensors. Printed Films: Materials Science and Applications in Sensors, Electronics and Photonics, 221-258. doi:10.1533/9780857096210.2.221
4. Ramsay, J. O., & Silverman, B. W. (2005). Functional Data Analysis. Springer.
5. Rendl, Christian & Greindl, Patrick & Haller, Michael & Zirkl, Martin & Stadlober, Barbara & Hartmann, Paul. (2012). PyzoFlex: Printed Piezoelectric Pressure Sensing Foil. UIST'12 - Proceedings of the 25th Annual ACM Symposium on User Interface Software and Technology.

Conformity assessment of lots in the framework of JCGM 106:2012

Rainer Göb, Steffen Uhlig, Bernard Colson

Key words: conformity assessment, statistical lot inspection, sampling uncertainty, producer's risk, consumer's risk

1 The general conformity assessment environment

ISO (2020) defines conformity assessment as an "activity to determine whether specified requirements relating to a product, process, system, person or body are fulfilled". JCGM (2012) establishes a framework for accounting for measurement uncertainty in conformity assessment. The focus of JCGM (2012) is on the conformity assessment of individual units of product based on measurements on a cardinal continuous scale. However, the scheme can also be applied to composite assessment targets like finite lots of product or manufacturing processes, and to the evaluation of characteristics in discrete cardinal or nominal scales.

2 Conformity assessment of finite lots

We consider the application of the JCGM scheme in the conformity assessment of finite lots of units subject to a dichotomous quality classification as conforming and nonconforming. A lot is classified as conforming if the actual proportion nonconforming does not exceed a prescribed upper tolerance limit, otherwise the lot is classified as nonconforming. The measurement on the lot is a statistical estimation of the proportion nonconforming based on attributes or variables sampling, and measurement uncertainty is sampling uncertainty. Following JCGM (2012), we analyse the effect of measurement uncertainty (sampling uncertainty) both in attributes sampling, and we calculate key conformity assessment parameters, in particular the

Rainer Göb
Department of Statistics, Sanderring 2, D-97070 Würzburg, Germany, e-mail: goeb@mathematik.uni-wuerzburg.de

Steffen Uhlig
QuoData GmbH, Prellerstraße 14, D-01309 Dresden, Germany, e-mail: uhlig@quodata.de

Bernard Colson
QuoData GmbH, Prellerstraße 14, D-01309 Dresden, Germany, e-mail: colson@quodata.de

producer's and consumer's risk. We suggest to integrate such parameters as a useful add-on into ISO acceptance sampling standards such as the ISO 2859 series.

References

1. ISO (2020) ISO/IEC 17000:2020 Conformity assessment – Vocabulary and general principles. International Organization for Standardization (ISO), Geneva.
2. JCGM (2012a) *JCGM 106:2012 Evaluation of Measurement Data – The role of measurement uncertainty in Conformity Assessment*. Joint Committee on Guides in Metrology (JCGM).

Multilevel Measurement for Business and Industrial Workforce Development: Part I

Jan Morrison[1] and William P. Fisher, Jr.[2]

Key words: probabilistic models, multilevel measurement, workforce development

1. Metrologically Applicable Models of Complex Situations

Technological change is occurring at a pace too rapid to allow prediction even in the near term of many of the kinds of skills that will be needed in the future. Workforce development has become a just-in-time process unfolding as new opportunities for innovation arise. Instead of preparing skill sets targeting specific job descriptions, the skills and jobs are created by the prospective incumbents themselves.

Stochastic models facilitate adapting to this dynamic state of affairs, guiding development of resilient information infrastructures incorporating measuring instruments traceable to metrological standards [Pendrill(2015)].

1.1 Measuring Workforce Implications of STEM Education

An example involves learning from variations in the structure, processes, and outcomes of science, technology, engineering, and mathematics (STEM) education conducted in a wide range of in-school, out-of-school, and after-school activities [Morrison(2018,2019)]. The conceptualization, design, management, analysis, reporting, and application of these kinds of measurements pose particularly daunting challenges to generalizing results produced from unique authentic local contexts.

Measuring, communicating, managing, and improving distributed process outcomes of these kinds demand complex, multilevel experimental designs, data aggregations, statistical summaries, instrument calibrations, and measurements.

[1] Jan Morrison
TIES: Teaching Institute for Excellence in STEM, Cleveland, OH, USA, e-mail: janmorrison@tiesteach.org
[2] William P. Fisher, Jr.
Living Capital Metrics LLC, Sausalito, CA, USA, e-mail: wpfisherjr@livingcapitalmetrics.com

Additional multidimensional and multifaceted complexities may be encountered, increasing the challenges that must be surmounted if the instrumentation produced is to be useful to end users in a distributed metrology network.

1.2 Managing STEM Learning Ecosystem Success

Multilevel probabilistic models incorporating explanatory variables predicting instrument calibrations and outcome measurements [Stenner(2013)] provide a context for reporting results in actionable, meaningful form to end users. The stability of the structural invariance designed into the form of the instrument is experimentally tested against both theoretical predictions and empirical data. Practical use of the instrument over time demands capacities for adaptively administering subsets of items measuring within expected uncertainty tolerances (reliabilities) without compromising the quantitative unit of comparison. Results may then support deployment of the instrument in ongoing quality improvement, accountability, and research applications reported in a quality-assured and metrologically traceable unit and uncertainty.

2. Developing *Caliper*

Caliper is an instance of this kind of measurement research and development involving over 90 STEM learning ecosystems comprised of employers, funders, local and state governments, in-school educators, and informal educators in workplaces, museums, zoos, parks, and elsewhere. These ecosystems have formed in five countries (Canada, Israel, Kenya, Mexico, and the United States) over the last six years. Measurement attention was focused on five pillars of ecosystem success (partnerships, measurements, administrative systems, teaching and learning practices, and workforce development connections) identified by stakeholders.

Caliper's design, calibrations, stakeholder comparisons, interpretation guidelines, and reports are described in a partner presentation by the same authors.

References

1. Morrison, J., Fisher, W.P. Connecting learning opportunities in STEM education. J. Phys. Conf. Series, **1065**, 022009 (2018)
2. Morrison, J., Fisher, W.P. Measuring for management in STEM learning ecosystems. J. Phys. Conf. Series, **1379**, 012042 (2019)
3. Pendrill, L., Fisher, W.P. Counting and quantification: Comparing psychometric and metrological perspectives on visual perceptions of number. Meas. **71**, 46-55 (2015)
4. Stenner, A.J., Fisher, W.P., Stone, M.H., Burdick, D.S. Causal Rasch models. Frontiers in Psychology: Quantitative Psychology and Measurement, **4**, 1-14 (2013)

Mathematical tools for a better analysis of the covariance in industrial data

Peggy Courtois[1], Christophe Dubois[2], and Jean-Michel Pou[3]

Key words: Covariance, International Vocabulary of Metrology, JCGM 106:2012, Mahalanobis Distance, Bayesian Measurement Refinement

1. Introduction

Every day, countless algorithms are developed and used to analyse and extract information to identify our consumable needs or create new ones. Slowly, industrials start to realise that, not only, data can be used to optimise their manufacturing lines, but when used properly, predict instrument drifts and help understand instrument measuring behaviour.

The performance of these algorithms depends on the quality of the data. In metrology, we all know that measurements are marred by uncertainties. By estimating the different sources of variation impacting the measurement, we have a closer look of the 'true value' of the measurands, and thus a better understanding of our instrument behaviour. The new supplement of the definition of the term *calibration* proposed by the International Vocabulary of Metrology[1] acknowledges the need to converge toward the true data.

An additional characteristic that most industrials ignore due to the complexity of the measurement analysis is the covariance. The covariance is defined as a measure of the joint variability of two variables (or two sets of data) in their linear relationship. The strength of the covariance between two sets of data can be significantly deteriorated as the dispersion of the measurements is great.

In this paper, we present a simple method, called *Bayesian Measurement Refinement* (BMR), inspired from the JCGM 106:2012. This approach retrieves the 'true values' of the measurands based on historical data. We will show that this approach is not limited to a single dimension and can easily be extended to N dimensions, taken into account the covariance in the historical data.

[1] Peggy Courtois
Deltamu, Centre d'Affaires du Zénith, 48 rue de Sarlève, 63800 Cournon d'Auvergne, France, e-mail: pcourtois@deltamu.com
[2] Christophe Dubois
Deltamu, Centre d'Affaires du Zénith, 48 rue de Sarlève, 63800 Cournon d'Auvergne, France, e-mail: cdubois@deltamu.com
[3] Jean-Michel Pou
Deltamu, Centre d'Affaires du Zénith, 48 rue de Sarlève, 63800 Cournon d'Auvergne, France, e-mail: jmpou@deltamu.com

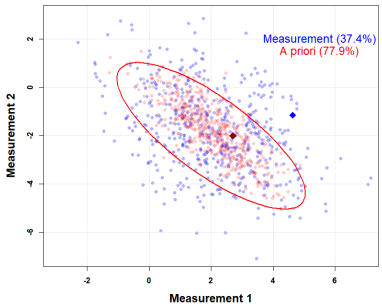
2. Bayesian Measurement Refinement (BMR)

Based on Bayes' theorem on conditional probability, the BMR approach uses historical data to determine the most likely value from a measurement and its uncertainty[2].

The method is built on the knowledge of the historical data, as known as *a priori*. The *a priori* can be either determined by previous knowledge or theory, either by standard measurements, or even by numerical simulation (FD-X 07-039). To estimate the 'true value' of a measurands, the uncertainty needs to be calculated. A strong hypothesis of the BMR approach is the normal distribution of the historical data (Central Limit Theorem).

Figure 1 shows a set of correlated data in 2 dimensions, where the a priori shown in red is correlated at 77.9% (r = 0.779), and the measurements in blue correlated at 37% (r = 0.37) only. By means of the BMR approach, we refined the measurements into refined values comprised in the red ellipse (99% of confidence interval). An example of a measurement (blue diamond) refined into a refined value (red diamond) is shown in Figure 1.

Figure 1: Comparison between the measurements (blue) with a correlation of 37%, and the a priori (red) correlated at 77.9%. The BMR approach is used on the measurement represented by the blue diamond, the refined value (red diamond) is shown in the ellipse delimiting the a priori data at 99%.



To support our work, we will use the BMR approach on simulated and real industrial data to bring the correlations out. Coupled with the distance of Mahalanobis (Hotelling, T2 control chart), we will then identify anomalies issued from a combination of several variables (multivariate approach). Future work would be to extend the Mahalanobis approach to Principal Component Analysis (PCA) to predict the shift of an instrument through time.

References

1. JCGM 200:2012, International vocabulary of metrology - Basic and general concepts and associated terms (VIM)
2. JCGM 106:2012, Evaluation of measurement data – The role of measurement uncertainty in conformity assessment

Multilevel Measurement for Business and Industrial Workforce Development: Part II

William P. Fisher, Jr[1] and Jan Morrison[2]

Key words: probabilistic models, multilevel measurement, workforce development

1. Caliper Pilot Study instrument development

In the context of Part I, variation in ecosystem success was mapped in terms of increasing stakeholder involvement, 161 assessment items specifying mapped locations were written, four parallel forms with 31 common items were devised, and an online administration system was created. A data set of 291 responses became available in mid-2020, and a second administration began in early March 2021.

This ongoing pilot study investigates the feasibility of developing a knowledge infrastructure that maps multiple levels of stakeholder engagement in the STEM Learning Ecosystems and measures ecosystem success in actionable, practical ways.

2. Method

Following a construct mapping design process [Fisher(2006),Wilson(2005)], item content was structured via a five-level theory of stakeholder empowerment [Ortiz(2017)] applied to all five domains (partnerships, measurements, administrative systems, teaching and learning, and workforce development) identified by stakeholders as essential to ecosystem success.

Conservatively estimating a measurement standard deviation of 1.0 logits, the 31 common items should result in a measurement separation reliability over 0.92

[1] William P. Fisher, Jr.
Living Capital Metrics LLC, Sausalito, CA, USA; BEAR Center, GSE, UC Berkeley, CA, USA; Research Institute of Sweden, Gothenburg, Sweden
e-mail: wpfisherjr@livingcapitalmetrics.com

[2] Jan Morrison
TIES: Teaching Institute for Excellence in STEM, Cleveland, OH, USA, e-mail: janmorrison@tiesteach.org

[Linacre(1993)]. The measurement model [Pendrill(2015),Wilson(2005)] has the form:

$$\ln(P_{nij} / (1 - P_{nij})) = B_n - D_i - K_j$$

which states the formal expectation that the log of the odds of success is a function of the difference between the experience B of stakeholder n and the difficulty D of item i relative to the difficulty K of disagreeing or agreeing at the level of category j. The basic form of this model, devised by Rasch in the 1950s, is widely used in education and health care measurement as a means of obtaining linear interval quantity values from ordinal observations. The idealized hypothesis is that nothing affects the outcome of the question and answer interaction but the difference between the individual ability and the item difficulty. Though never perfectly realized in practice, decades of research and applications have proven the value and utility of this model.

3. Calibration results

For the 161-item scale, measurement separation reliability ranged from 0.94 to 0.96 for the total sample of 291 respondents. Though there is always room for improvement, structural invariance was approximated, as measurements based solely on the workforce items alone correlated 0.80 with the overall measures, and measurements from the other three pillars' items correlated 0.83 to 0.92 with the overall measurements. None of the correlations are disattenuated; given the subscales' lower reliabilities, correlations approach 1.00 after accounting for uncertainty.

Figure 1 shows a typical example of the high correlation and linear relation obtained for measurements produced from entirely different sets of Caliper items. Theoretical groupings of items showed mean calibrations in the predicted order.

These and other results support the development of individualized point of use applications of measurements in formative quality improvement applications. Capacities for isolating and leveraging unique special strengths and weaknesses open onto new possibilities for customized management of partnerships, systems, and other ecosystem success drivers.

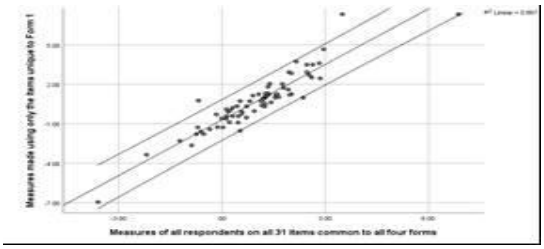


Figure 1: Same measurements from separate groups of items: 32 unique Form 1 items vs 31 common items

References

1. Fisher, W. P., Jr. Survey design recommendations. Rasch Meas. Trans. **20**, 1072-1074 (2006)
2. Linacre, J. M. Rasch-based generalizability theory. Rasch Meas. Trans. **7**, 283-284 (1993
3. Ortiz, D., Huber-Heim, K. From information to empowerment. The Internat. J. Manag. Educ., **15**, 318-331 (2017)
4. Pendrill, L., Fisher, W.P. Counting and quantification. Meas. **71**, 46-55 (2015).
5. Wilson, M R. Constructing measures. Erlbaum, Mahwah, NJ (2005)

A measure of the statistical homogeneity of turbulent simulations

A. Ferrero, F. Larocca, G. Scovazzi and M. Germano

Key words: Numerical simulation, Expected values, Statistical homogeneity

1 Abstract

A major problem in the numerical simulations of turbulent flows is the problem of the average, the extraction of expected values. The definition of independent solutions of the Navier-Stokes equations is theoretically undefined and the practical realization of an ensemble of realizations prohibitively expensive, so that ergodic averaging in time and eventual homogeneous directions is a common practice to extract *statistical* values. The increased computing power stimulates mixed approaches, where ergodic averaging is coupled to an ensemble of multiple realization [1, 2, 3] This problem is presently a matter of study in a joint cooperation between the Politecnico di Torino and the Duke University as regards the turbulent flow past a cylinder [4]. In the case of this benchmark flow we have two homogeneities, the time t and the spanwise spatial direction z, and we can average along one or both of them. Given the turbulent velocity field $u_i(x,y,z,t)$ we will introduce the following space and time *statistical* filtering operators $\mathscr{T}$ and $\mathscr{Z}$

$$\langle u_i \rangle_z = \frac{1}{Z}\int_0^Z u_i(x,y,z',t)dz' \quad ; \quad \langle u_i \rangle_t = \frac{1}{T}\int_0^T u_i(x,y,z,t')dt' \tag{1}$$

and we define $\mathscr{E}$ as the product of the two, $\mathscr{E} \equiv \mathscr{T}\mathscr{Z}$, where Z and T are the extents of the computational domain in the spanwise direction z and in time t.

Let us first of all examine the homogeneity between space and time averages. Consistently with the filtering approach based on the generalized central moments [5] let us define the following turbulent stresses associated to the filters $\mathscr{T}$ and $\mathscr{Z}$

$$\tau_z(u_i,u_j) \equiv \langle u_i u_j \rangle_z - \langle u_i \rangle_z \langle u_j \rangle_z \quad ; \quad \tau_t(u_i,u_j) \equiv \langle u_i u_j \rangle_t - \langle u_i \rangle_t \langle u_j \rangle_t$$
$$\tau_z(\langle u_i \rangle_t, \langle u_j \rangle_t) \equiv \langle \langle u_i \rangle_t \langle u_j \rangle_t \rangle_z - \langle u_i \rangle_{zt} \langle u_j \rangle_{zt} \quad ; \quad \tau_t(\langle u_i \rangle_z, \langle u_j \rangle_z) \equiv \langle \langle u_i \rangle_z \langle u_j \rangle_z \rangle_t - \langle u_i \rangle_{tz} \langle u_j \rangle_{tz}$$

We note that due to

$$\mathscr{E} = \mathscr{Z}\mathscr{T} = \mathscr{T}\mathscr{Z} \tag{2}$$

A. Ferrero e-mail: andrea_ferrero@polito.it, F. Larocca e-mail: francesco.larocca@polito.it
Politecnico di Torino

G. Scovazzi e-mail: guglielmo.scovazzi@duke.edu, M. Germano e-mail: mg234@duke.edu
Duke University

we have the two identities

$$\begin{aligned}\tau_e(u_i,u_j) &\equiv \langle u_i u_j\rangle_e - \langle u_i\rangle_e\langle u_j\rangle_e \\ &= \tau_z(\langle u_i\rangle_t,\langle u_j\rangle_t) + \langle\tau_t(u_i,u_j)\rangle_z = \tau_t(\langle u_i\rangle_z,\langle u_j\rangle_z) + \langle\tau_z(u_i,u_j)\rangle_t \end{aligned} \quad (3)$$

and we can define two measures of statistical homogeneity, the first related to the time average and the second to the spanwise average, given by

$$M_t(x,y) = \frac{\langle\tau_t(u_i,u_i)\rangle_z}{R_{ii}} \quad ; \quad M_z(x,y) = \frac{\langle\tau_z(u_i,u_i)\rangle_t}{R_{ii}} \quad (4)$$

where $R_{ij} \equiv \tau_e(u_i,u_j)$ are the Reynolds stresses.

All that can be generalized. In the case of more *statistical* operators in time, homogeneous directions or ensemble solutions we can write

$$\mathscr{E} = \mathscr{E}_1\mathscr{E}_2 = \mathscr{E}_1\mathscr{E}_2\mathscr{E}_3 = \cdots \quad (5)$$

and we can read all that as a multiscale analysis of the variance, where $\mathscr{E}$ is the *statistical* operator and $\mathscr{E}_i$ are different nested partitions of the probability space. In this case we can write

$$\langle u_i\rangle_e = \langle\langle u_i\rangle_2\rangle_1 = \langle\langle\langle u_i\rangle_3\rangle_2\rangle_1 = \cdots \quad (6)$$

and

$$\begin{aligned}\tau_e(u_i,u_j) &= \tau_1(\langle u_i\rangle_2,\langle u_j\rangle_2) + \langle\tau_2(u_i,u_j)\rangle_e \\ &= \tau_1(\langle u_i\rangle_2,\langle u_j\rangle_2) + \langle\tau_2(\langle u_i\rangle_3,\langle u_j\rangle_3)\rangle_1 + \langle\langle\tau_3(u_i,u_j)\rangle_2\rangle_1 = \cdots \end{aligned} \quad (7)$$

In the presentation some numerical results related to a flow around a circular cylinder at Reynolds number 3900 will be provided. We remark finally that the problem of the average is preliminary to any other problem. The issues of the uncertainty and the validation of the results are strictly interconnected and dependent on that.

References

1. D. Carati, M. M. Rogers, A.A. Wray, *Statistical ensemble of large eddy simulations*, J. Fluid Mech., 455, 195-212, (2002)
2. D. Farrace, R. Panier, M. Schmitt, K Boulouchos, Y. M. Wright, *Analysis of averaging methods for large eddy simulations of diesel sprays*, SAE Int. J. Fuels Lubr. 8(3), (2015)
3. V. Makarashvili, E. Merzari, A. Obabko, A. Siegel, P. Fischer, *A performance analysis of ensemble averaging for high fidelity turbulence simulations at the strong scaling limit*, Computer Physics Communications 219, 236-245, (2017)
4. A. Ferrero, F. Larocca, G. Scovazzi, M. Germano, *A numerical study of the spanwise turbulence past a cylinder flow*, 17th ETC, 3-6 Sept. 2019, Torino, ITALY, (2019)
5. M. Germano, *Turbulence: the Filtering Approach*, J. Fluid Mech. 238, 325-336, (1992)

Forecasting the COVID-19 epidemic integrating symptom search behavior: an infodemiology study

Eugenio Alladio[1], Alessandro Rabiolo[2,3], Esteban Morales[4], Alessandro Marchese[3]

Keywords: Google Trends; SARS-CoV-2; Time Series; Predictive Models; Shiny Web-Application

Background: Previous studies have suggested associations between trends of web searches and COVID-19 traditional metrics. It remains unclear whether models incorporating trends of digital searches lead to better predictions.

Methods: An open-access web application was developed to evaluate Google Trends and traditional COVID-19 metrics via an interactive framework based on principal components analysis (PCA) and time series modelling. The app facilitates the analysis of symptom search behavior associated with COVID-19 disease in 188 countries. In this study, we selected data of eight countries as case studies to represent all continents. PCA was used to perform data dimensionality reduction, and three different time series models (Error Trend Seasonality, Autoregressive integrated moving average, and feed-forward neural network autoregression) were used to predict COVID-19 metrics in the upcoming 14 days. The models were compared in terms of prediction ability using the root-mean-square error (RMSE) of the first principal component (PC1). Predictive ability of models generated with both Google Trends data and conventional COVID-19 metrics were compared with those fitted with conventional COVID-19 metrics only. Confidence intervals can be evaluated, too.

Findings: The degree of correlation and the best time-lag varied as a function of the selected country and topic searched; in general, the optimal time-lag was within 15 days. Overall, predictions of PC1 based on both searched termed and COVID-19 traditional metrics performed better than those not including Google searches, but the improvement in prediction varied as a function of the selected country and timeframe. The best model varied as a function of country, time range, and period of

Eugenio Alladio[1]
Department of Chemistry, University of Turin, Turin, Italy
Alessandro Rabiolo[2,3]
Department of Ophthalmology, Gloucestershire Hospitals NHS Foundation Trust, Cheltenham, UK
Department of Ophthalmology, Vita-Salute University, IRCCS Ospedale San Raffaele Scientific Institute, Milan, Italy
Esteban Morales[4]
Jules Stein Eye Institute, David Geffen School of Medicine, UCLA, Los Angeles, USA
Alessandro Marchese3
Department of Ophthalmology, Vita-Salute University, IRCCS Ospedale San Raffaele Scientific Institute, Milan, Italy

time selected. Models based on a 7-day moving average led to considerably smaller RMSE values as opposed to those calculated with raw data.

Conclusions: The inclusion of digital online searches in statistical models may improve the prediction of the COVID-19 epidemic. Prepared website is available at predictpandemic.org.

References

1. Broy, M.: Software engineering -- from auxiliary to key technologies. In: Broy, M., Dener, E. (eds.) Software Pioneers, pp. 10-13. Springer, Heidelberg (2002)
2. Dod, J.: Effective substances. In: The Dictionary of Substances and Their Effects. Royal Society of Chemistry (1999) Available via DIALOG. http://www.rsc.org/dose/title of subordinate document. Cited 15 Jan 1999
3. Geddes, K.O., Czapor, S.R., Labahn, G.: Algorithms for Computer Algebra. Kluwer, Boston (1992)
4. Hamburger, C.: Quasimonotonicity, regularity and duality for nonlinear systems of partial differential equations. Ann. Mat. Pura. Appl. **169**, 321--354 (1995)
5. Slifka, M.K., Whitton, J.L.: Clinical implications of dysregulated cytokine production. J. Mol. Med. (2000) doi: 10.1007/s001090000086

Measurement system analysis of the CSLT measurement system

An experiment to defect diagnoses in deep drilled shafts

HO, Tak Cho Eric[12] , ZWETSLOOT Inez Maria[3]

Key words: concrete testing; ultrasonic; crosshole tomography; foundation testing; measurement accuracy and precision; CSLT; CSL; drilled shaft

Abstract

A new measurement system called Crosshole Sonic Logging Tomography (CSLT) provides information on the size, shape, and orientation of defects in a deep drilled shaft. The CSLT measurement system is not (yet) accredited in Hong Kong for foundation testing. Deep drilled shafts in Hong Kong are generally wide and deep. Existing measurement system analysis studies do not consider this large type of drilled shafts. The objective of our research is two fold: first we wish to quantify the measurement accuracy and precision of the CSLT method for large diameter drilled shaft (the most common shaft for public housing projects in Hong Kong). Second we study the repeatability and reproducibility of the CSLT measurement system. We construct a test pile shaft with known defects and perform experiments with a CSLT measurement systems to evaluate the system.

In the first stage of our experiment, we find that the CSLT is accurate in detecting shape, size, and location of large defects. We also find that

[1], FT Laboratories Limited., Fanling, New Territories, Hong Kong
e-mail: eric.ho@ft.com.hk

[2] Department of Advanced Design and Systems Engineering, City University of Hong Kong, Kowloon, Hong Kong
e-mail: eric.ho.engd@my.cityu.edu.hk

[3] Department of Advanced Design and Systems Engineering, City University of Hong Kong, Kowloon, Hong Kong
e-mail: i.m.zwetsloot@cityu.edu.hk

small defects close to the tube are difficult to detect. We concluded that the CSLT has satisfactory accuracy and precision for practical use.

In order to analyze the CSLT results we have developed feature extraction methodology to analyze the 2D images obtained from the CSLT software. We extract various features and compute performance measures.

In the second stage of our experiment, we quantify the discrepancy it may introduce by mean of an inter-operator comparison - two operators shall be collecting data with the same machine. After that, a repeatability experiment shall be carried out for the collection of multiple sets of data with the same machine. Finally, we shall evaluate the data and software reliability by means of comparing data collected by two machines of different types.

We believed the contribution of this study will help us to recommend the adoption of the CSLT for foundation testing in Hong Kong. We also believed this study will lead to a successful application of accreditation as an accredited foundation test method under the laboratory accreditation scheme in Hong Kong.

References

1. Amir, J.M. and Amir, E.I., 2009. Capabilities and limitations of cross hole ultrasonic testing of piles. In Contemporary Topics in Deep Foundations (pp. 536-543).
2. ASTM, 2016. Standard test method for integrity testing of concrete deep foundations by ultrasonic crosshole testing. Designation D6760-16.
3. Chernauskas, L.R. and Hajduk, E.L., 2009. The Use of Crosshole Tomography to Evaluate Drilled Shaft Repairs. In Contemporary Topics in Deep Foundations (pp. 560-567).
4. Hajduk, E.L., Tallent, M.R., Ledford, D.L. and Christopher, W.R., 2004. Crosshole sonic logging integrity testing for the new cooper river bridge. International Conference on Case Histories in Geotechnical Engineering.
5. Islam, M.S., Yasmin, T., Hannan, S.A. and Peela, P., 2015. Integrity evaluation of concrete by cross-hole sonic logging in Bangladesh. IABSE-JSCE Joint Conference on Advances in Bridge Engineering-III.
6. Jalinoos, F., Mekic, N. and Hanna, K., 2005. Drilled shaft foundation defects: Identification, imaging, and characterization. FHWA, Publication No. FHWACFL/TD-05-007.
7. Likins, G., Robinson, B. and Piscsalko, G., 2012. A brief overview of testing of deep foundations. Proceedings, Testing and Design Methods for Deep Foundations. Kanazawa, Japan, pp.97-104.
8. Likins, G., Rausche, F., Webster, K. and Klesney, A., 2007. Defect analysis for CSL testing. In Contemporary Issues in Deep Foundations (pp. 1-10).

Automated ML Toolbox for Cyclic Sensor Data

Tanja Dorst, Yannick Robin, Tizian Schneider and Andreas Schütze

Key words: machine learning, cyclic sensor data, software toolbox

1 Motivation

In this contribution, an easy-to-use software toolbox for machine learning is presented. The need for such a toolbox becomes obvious when building analytical models is too complex and data-based models are needed. Many software packages are already available but they are usually too complex for non-experienced users as the tools provide a wealth of tunable hyperparameters in order to cover as many problems as possible. Therefore, a software toolbox is introduced requiring no expert knowledge about data science that can be operated without any analytical model of the task at hand. The toolbox can be used by applying the data to a MATLAB interface which will automatically and without any hyperparameter tuning lead to an optimized machine learning model of the problem with performance comparable to a dedicated approach generated with available software regarding speed and accuracy. No tuning is necessary because complementary methods for feature extraction and selection (see Fig. 1) can cope with a large variety of problems regarding cyclic sensor data as already shown in [Schneider et al., 2018]. Furthermore, the methods are pre-configured to prevent standard machine learning problems like the course of dimensionality and overfitting thus allowing plug-and-play use.

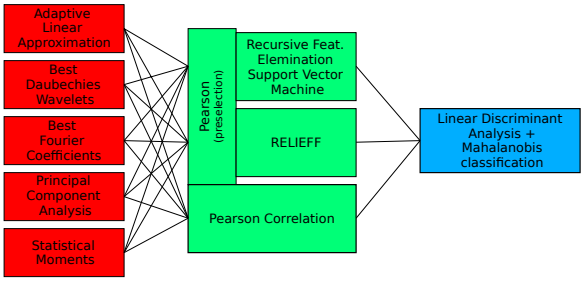


Fig. 1 Toolbox scheme with feature extraction (red), selection (green) and classification (blue).

Tanja Dorst, Tizian Schneider
ZeMA gGmbH, Saarbruecken, Germany e-mail: t.dorst@zema.de, t.schneider@zema.de

Yannick Robin, Andreas Schütze
Lab for Measurement Technology, Department of Mechatronics, Saarland University, Saarbruecken, Germany e-mail: y.robin@lmt.uni-saarland.de, schuetze@lmt.uni-saarland.de

2 Implementation

The toolbox is designed in an object-oriented way and structured around the concept of a so-called *SimpleTrainingStack*. The *SimpleTrainingStack* object can be filled with different modules (e.g. extractors, selectors, classifiers) which are then processed according to the LIFO principle (e.g. List. 1). The *MultisensorExtractor* is a wrapper, that allows the user to specify an extractor that is applied to all input sensors. The Pearson selector in combination with the number (here 500) limits the total number of features after preselection. *NumFeatRanking* first sorts the features according to their importance based on the specified selector (for this example RFESVM) and then performs an exhaustive search to find the optimal number of features (max. 100) according to the cross-validation error rate. In order to use the full toolbox, which means applying all 15 possible combinations (Fig. 1) of *SimpleTrainingStacks*, can be done as shown in line 6 in List. 1.

Listing 1 Example code for *SimpleTrainingStack* with Best Fourier Coefficient as extractor, Pearson correlation as preselector, RFESVM as selector and LDA as classifier and the full toolbox.

```
1 %Single Stack
2 obj = SimpleTrainingStack(...
3         {@MultisensorExtractor,@Pearson,@NumFeatRanking,@LDAMahalClassifier},...
4         {{@BFCExtractor},{500},{@RFESVM},{}});
5 %Full toolbox
6 obj = Factory.FullToolboxMultisens()
```

The training and actual prediction of either a single *SimpleTrainingStack* or the full toolbox can be performed as shown in List. 2. For the input data each sensor must be stored in a cell and each sensor has to be a two dimensional array with the individual recordings stored in rows. The target needs to be in a column vector.

Listing 2 Training with training data and application with test data.

```
1 obj.train(sensors_train,target_train)
2 prediction = obj.apply(sensors_test)
```

Applying the full toolbox to the task at hand will result in the best possible machine learning model based on the 15 available combinations. In order to obtain the information of which combination and features were used, it is possible to use the info function of the full toolbox object or explore the object itself. If the predefined methods do not fulfill the users need, additional modules can simply be added by following the interface structure of train and apply functions. The code for the toolbox can be found on GitHub (https://github.com/ZeMA-gGmbH/LMT-ML-Toolbox).

Acknowledgements The basic version of the software toolbox was developed within the research project MoSeS-Pro sponsored by the German Federal Ministry of Education and Research (funding code 16ES0419K), within the framework of the German Hightech Strategy.

References

[Schneider et al., 2018] Schneider, T., Helwig, N., and Schütze, A. (2018). Industrial condition monitoring with smart sensors using automated feature extraction and selection. *MST*, 29(9).

Intrinsic and Metrological Correlations on the Risks of False Conformity Decisions

Luciana Separovic[1], Ricardo J. N. Bettencourt da Silva[2] and Felipe R. Lourenço[3]

Key words: multiparameter drug, correlation components, total risk, decision making.

1. Introduction

The conformity assessment of a medical product with multiple active pharmaceutical ingredients may present experimental correlation between the measured values, which can affect the total risk of false decisions [Kuselman (2018)]. This correlation (called total correlation) can arise from the characteristics of the product (intrinsic) or from the measurement procedure (metrological). This work aimed to evaluate the impact of the intrinsic and metrological components of total correlation on the total risk of false conformity decisions.

2. Materials and Methods

The content of 12 batches of amiloride hydrochloride (AML) + hydrochlorothiazide (HCT) tablets was estimated by HPLC separately (independent measurements) and then together (correlated measurement). The correlations were calculated for each type of analysis, using a MS-Excel spreadsheet [Separovic (2020)]. The total risks were estimated in a Bayesian framework (global risks) by Monte Carlo Method, considering the intrinsic and metrological correlations and also considering only the total correlation.

[1] Luciana Separovic
Departamento de Farmácia, Faculdade de Ciências Farmacêuticas, Universidade de São Paulo; 05508-000; São Paulo, Brazil, e-mail: luseparovic@usp.br
[2] Ricardo J. N. Bettencourt da Silva
Centro de Química Estrutural, Faculdade de Ciências da Universidade de Lisboa; 1749-016; Lisbon; Portugal, e-mail: rjsilva@fc.ul.pt
[3] Felipe R. Lourenço
Departamento de Farmácia, Faculdade de Ciências Farmacêuticas, Universidade de São Paulo; 05508-000; São Paulo, Brazil, e-mail: feliperl@usp.br

3. Results and Discussion

A process historical scenario with mean values of 93.3% and 94.4%, and standard deviation 1.66% and 2.2% for components HCT and AML, respectively, was considered. The global risks for this scenario are shown in Table 1, which also reports the total risks determined from simulations where calculated intrinsic and metrological correlations are considered and where, for simplicity, the same correlation coefficient (total correlation) is used for both correlation components.

Table 1: Total consumer's and producer's risk for independent and correlated measurements considering quantified "total correlation" or "separate intrinsic and metrological correlations".

Measurement	***Independent***		***Correlated***	
Correlation	***Intr. and Metr.***	***Total***	***Intr. and Metr.***	***Total***
Total consumer risk (%)	0.69 (0.03)	1.18 (0.07)	0.91 (0.05)	1.03 (0.03)
Total producer risk (%)	5.4 (0.1)	5.3 (0.1)	4.8 (0.1)	4.70 (0.09)

Mean and (standard deviation) obtained from six replications of simulations.

Considering the independent measurement, the total producer risk was above 5%, while for the correlated measurement, the same risk was below 5%. This example illustrates how the metrological correlation affects the total correlation and the total risks of false decision. This observation proves that the simplified management of data correlation can produce unrealistic risk determinations.

4. Conclusions

The approach proposed in this paper, that considers the intrinsic and metrological correlation separately, allows a more realistic determination of the risk of false conformity decisions than assuming that all correlation components have the same value.

The relevance of metrological correlation for the risks of false decision is noticed from the significant difference ($p<0.05$) between the total risk values obtained by the independent or correlated measurements. Considering the specific case presented in the paper, the total risk was reduced by about 11% when adopting the correlated analytical procedure in comparison to the independent analytical procedure.

References

1. Kuselman, I., Pennecchi, F.R., da Silva, R.J.N.B., Hibbert, D.B., Anchutina, E.: Total risk of a false decision on conformity of an alloy due to measurement uncertainty and correlation of test results. Talanta. **189**, 666–674 (2018)
2. Separovic, L.; Bettencourt da Silva, R. J. N.; Lourenço, F. R.: Determination of Intrinsic and Metrological Correlations of Components of Product Impact on Risks of False Decisions in Conformity Assessment. Proceedings, **55**, 1-3 (2020)

A Local-Integral approach to Electric Properties Tomography

L. Zilberti, A. Arduino, U. Zanovello and O. Bottauscio[1]

Key words: Electric Properties Tomography (EPT), Magnetic Resonance Imaging (MRI), Quantitative Imaging, Green's theorem.

1. Introduction

Magnetic Resonance Imaging (MRI) is a long-standing diagnostic technique. New quantitative implementations of MRI are being developed, thanks to techniques that elaborate the MRI data to obtain images where the pixels represent the measurement of some physical parameter. One such technique is Electric Properties Tomography (EPT), which analyses the spatial distribution of the radiofrequency magnetic field employed during MRI to recognize the underlying spatial distribution of the dielectric properties (conductivity σ and permittivity ε) in the body that undergoes the exam, to be used as biomarkers for a number of pathologies [Leijsen(2021)].

Recently, many algorithms have been proposed to solve the EPT problem. Here we present a new approach, based on Green's integral identities, under development in the 18HLT05 QUIERO project. This project 18HLT05 QUIERO has received funding from the EMPIR programme co-financed by the Participating States and from the European Union's Horizon 2020 research and innovation programme.

2. Theory

Under the so-called "local homogeneity assumption" (LHA), the rotating part of the magnetic flux density B, used in MRI, satisfies the complex Helmholtz equation

$$\nabla^2 B = -\omega^2 \mu_0 \left(\varepsilon - j\frac{\sigma}{\omega} \right) B \tag{1}$$

[1] Luca Zilberti, Alessandro Arduino, Umberto Zanovello, Oriano Bottauscio
Istituto Nazionale di Ricerca Metrologica, Strada delle Cacce 91, Torino, Italy
e-mail: l.zilberti@inrim.it, a.arduino@inrim.it, u.zanovello@inrim.it, o.bottauscio@inrim.it,

where ω is the angular frequency, μ_0 is the magnetic permeability of vacuum and j is the imaginary unit. If the right-hand side of the equation (1) were a known term ζ, the equation could be interpreted as a Poisson equation. The latter, by virtue of Green's second identity, in a homogeneous region Ω admits the integral solution

$$B(P) = \oint_{\partial\Omega} \left[B \frac{d}{dn}\left(\frac{1}{4\pi r}\right) - \frac{1}{4\pi r}\frac{dB}{dn} \right] ds + \int_{\Omega} \frac{\zeta}{4\pi r} dv \tag{2}$$

where $\partial\Omega$ is the boundary of the region, oriented according to an inward normal unit vector n, while r is the distance between the fixed pole P and any integration point.

Actually, the magnetic flux density $B(P)$ can be measured during MRI. Hence, we invert the equation to get the permittivity ε and conductivity σ

$$\varepsilon - j\frac{\sigma}{\omega} = \frac{B(P) - \oint_{\partial\Omega} \left[B \frac{d}{dn}\left(\frac{1}{4\pi r}\right) - \frac{1}{4\pi r}\frac{dB}{dn} \right] ds}{\omega^2 \mu_0 \int_{\Omega} \frac{B}{4\pi r} dv} \tag{3}$$

Equation (3) can be applied for any point P where the distribution of B has been sampled (considering a small homogeneous region surrounding the point), obtaining an indirect measurement of the dielectric parameters.

3. Results and discussion

Figure 1 shows the results of a virtual experiment performed on simulated noiseless data at 64 MHz. The reconstruction matches the state of the art, showing a good capability to recognize the white and grey matter, but suffers some artefacts at their boundaries (where the LHA drops) and in the cerebrospinal fluid.

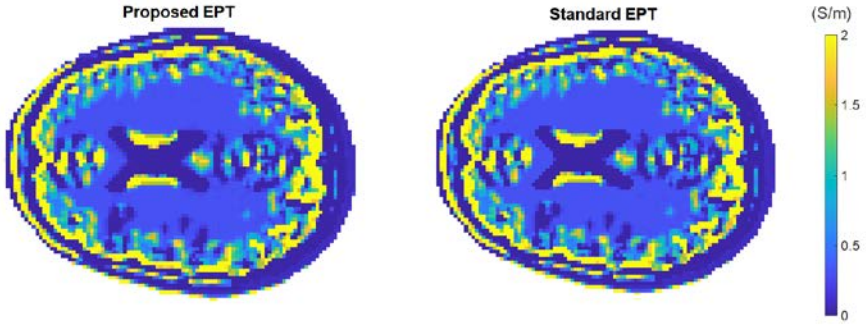


Figure 1: Conductivity in a head model: results of the proposed (left) and reference (right) methods.

References

1. Leijsen, R., Brink, W., van den Berg, C.A.T., Webb, A., Remis, R. Electrical properties tomography: a methodological review. Diagnostics **11**, (2021)

Understanding Neural Network Classifications: Local Interpretable Model-agnostic Explanations (LIME)

Hamza Zaheer[1], Philip J Aston[2], Jenny Venton[3]

Key words: LIME, Convolutional Neural Network, Average heatmaps, Interpretability.

1. Introduction

The Local Interpretable Model-agnostic Explanations (LIME) is an interpretability method that explains predictions of a machine learning model whether it is classification or regression [1]. LIME explains the predictions through the use of important predictors and the fitting of a simple interpretable model [2]. Electrocardiogram (ECG) signals are often classified using deep networks and networks that classify image transforms of an ECG have been explored [3].

The LIME method is used to generate heatmaps which showcase the areas or features of an image that are important for that particular classification, allowing for a better understanding of how the model reaches the classification decisions. Furthermore, average heatmaps are also generated for each category of classification to determine any underlying patterns that may be persistent for that particular classification. The aim of this study is to understand the basis for image classifications, thereby improving the interpretability of the ECG classification networks.

2. Methodology

The dataset being used was compromised of attractor images, generated using the Symmetric Projection Attractor Reconstruction (SPAR) method, these attractor images were ECG signals condensed into single images which can hold several minutes of ECG signal data [4]. The model used for classification was a pretrained convolutional neural network, trained using transfer learning to classify the attractor images. LIME heatmaps were generated using the test dataset only.

The network was used to classify each of the ECG input images into one of the three following categories: Atrial Fibrillation (AF), Normal (Normal) and ST Depression (STD). LIME was then used to generate a heatmap for the image based on that classification. A number of parameters can be varied with LIME. However, after testing various parameters the settings shown in *Table 1* below were chosen, as these seemed to provide the best insight into the classification decisions of the transfer learning model. The generated heatmaps were then stored, based on whether they had been correctly or incorrectly classified.

Since the real-world projects are comprised of large amounts of data it would be impractical to have to manually analyse each heatmap that is generated. Therefore, to determine patterns and trends in classification and misclassification, heatmaps for each of the categories were averaged [5].

At present, no specific method is available to generate average heatmaps therefore, a simpler technique was applied. This involved calculating the arithmetic mean of the heatmap images, allowing for easier comparisons between different categories of classification.

[1] Hamza Zaheer
Data Science, National Physical Laboratory, Teddington, UK, e-mail: hamza.zaheer@npl.co.uk
[2] Philip J Aston
Data Science, National Physical Laboratory, Teddington, UK, and Department of Mathematics, University of Surrey, Guildford, UK, e-mail: philip.aston@npl.co.uk
[3] Jenny Venton
Data Science, National Physical Laboratory, Teddington, UK, e-mail: jenny.venton@npl.co.uk

Table 1: Final set of parameter settings used for generating heatmaps.

Parameter	***Value***
Target number of features	400
Number of synthetic images	3000
Segmentation method	Superpixels
Type of simple model	Tree
Output upsampling method	Nearest
Size of mini-batch	256

3. Results – Averaging

The results provide much needed insight into the process of classification. *Figure 1* shows the AF and Normal average heatmaps along with their respective average attractor images. These heatmaps show which areas or features were important for the classifications and highlight patterns and trends relevant for each class.

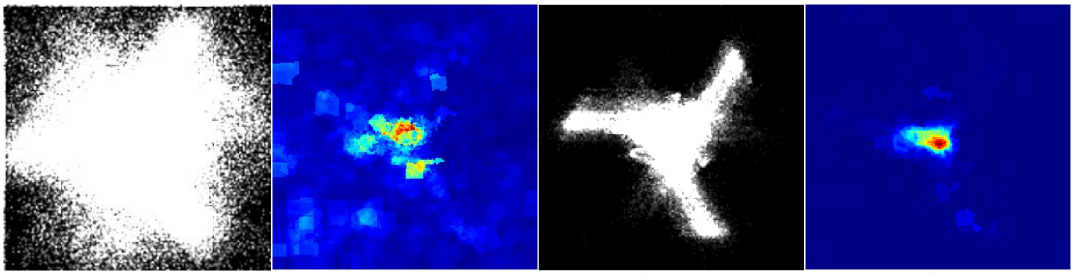


Figure 1: From L to R: Average image for AF category, average heatmap for AF category. Average image for Normal category, average heatmap for Normal category. Highest area of importance (Red) to lowest (Dark blue).

4. Discussion & Conclusion

In conclusion, the work is ongoing but the possibility of interpreting and understanding classification decisions of a neural network opens a wide door of possibilities. Not only does heatmapping allow for a better understanding of machine learning model decisions, the increased interpretability of these models will help build trust in their output when they are used as a decision support tool in a real-world setting.

Further work is being carried out with regards to different datasets, models, interpretability methods and research is being conducted to ensure that results achieved through these methods are robust, effective and can be applied to various practical issues.

References

[1] M. T. Rebeiro, S. Singh and C. Gestrin, “Why Should I Trust You?” Explaining the Predictions of Any Classifier,” Cornell University, 2016.

[2] Mathworks, “Understand Network Predictions Using LIME,” 2020. [Online]. Available: https://uk.mathworks.com/help/deeplearning/ug/understand-network-predictions-using-lime.html.

[3] S. Hong and Y. Zhou, “Opportunities and Challenges of Deep Learning Methods for Electrocardiogram Data: A Systematic Revew,” *Computers in Biology and Medicine,* vol. 122, 2020.

[4] P. J Aston and M. Christie, “Beyond HRV: attractor reconstruction using the entire cardiovascular waveform data for novel feature extraction,” *Physiological Measurement,* vol. 39, no. 2, 2018.

[5] S. Singhal, “All About Heatmaps,” December 2020. [Online]. Available: https://towardsdatascience.com/tagged/heatmap.

A Probability-box Based Approach for Measurement Problems

Tathagata Basu , Jochen Einbeck, Matthias C. M. Troffaes and Alistair Forbes

Key words: Delta method, Probability Box, Imprecise Probability

1 Introduction

Metrology has an important role in modern science and relies on the accuracy and repeatability of the measurements, which may be obtained through different expensive experiments and are often noisy due to uncertainty. However, most of the times, we rely on these measurements to obtain the measurand through a known functional form f. Therefore, we express a model by $y = f(\mu)$, where $\mu := (\mu_1, \ldots, \mu_m)$ are m different inputs and y is the measurand. Our objective is to quantify the underlying uncertainty around y. To do so, we find a confidence interval for $f(\mu)$, based on some estimates for μ. A common approach is the delta method, which is based on the multivariate normal approximation of a function. Let $\hat{X} := (\hat{X}_1, \ldots, \hat{X}_m)$ be an estimator of μ such that approximately $\hat{X} \sim N(\mu, \Sigma)$, where $\Sigma := \mathrm{Cov}(\hat{X})$. If f is differentiable, then by the Taylor expansion we have

$$f(\hat{X}) \approx f(\mu) + \nabla f(\mu)^T (\hat{X} - \mu). \tag{1}$$

If f is approximately linear around μ within the distributional range of $\hat{X}$, then we approximately have $f(\hat{X}) \sim N(f(\mu), \nabla f(\mu)^T \Sigma \nabla f(\mu))$. We can use this approxi-

Tathagata Basu
Durham University, UK e-mail: tathagatabasumaths@gmail.com
Jochen Einbeck
Durham University, UK e-mail: jochen.einbeck@durham.ac.uk
Matthias C. M. Troffaes
Durham University, UK e-mail: matthias.troffaes@durham.ac.uk
Alistair Forbes
National Physical Laboratory, London, UK e-mail: alistair.forbes@npl.co.uk

mate distribution to construct a 95% confidence interval for $f(\mu)$ to perform uncertainty quantification. For the variance term $\nabla f(\mu)^T \Sigma \nabla f(\mu)$, we have to use the sample variance matrix as we do not know Σ, and we have to use $\nabla f(\hat{X})$ as we do not know $\nabla f(\mu)$.

2 Modelling with Probability-boxes

To avoid those above mentioned issues, we propose the use of an imprecise probabilistic method for uncertainty quantification in metrology. This is yet to be explored in the field. We propose using p-boxes [1], which consist of a set of probability distributions bounded by lower and upper cumulative distribution functions so that

$$\mathscr{F} := \{F : \underline{F}(x) \leq F(x) \leq \overline{F}(x), \forall x \in \mathbb{R}\}. \tag{2}$$

This helps us to relax the distributional assumptions and thereby leads to more robust estimates. Additionally, uncertainty expressed as a p-box can be easily propagated through a range of standard operations.

3 Numerical Experiment

We illustrate our results by analysing the uncertainty associated with an end gauge calibration problem [2], where we try to determine the length (ℓ_M) of an end gauge (M) by comparing it with the length (ℓ_S) of a known standard (S) using the relation, $\ell_M = \frac{\ell_S(1+\alpha_S\theta_S)+d}{1+\alpha_M\theta_M}$. Here, α_M and θ_M (α_S and θ_S) are the thermal expansion coefficient and temperature deviation of M (S) and d is the difference between ℓ_M and ℓ_S. In practice, α_M and θ_M (α_S and θ_S) are often weakly correlated. We use p-boxes to characterise these variables. We inspect their behaviour against different dependence structures for uncertainty propagation and obtain a robust estimate of the measurand (y). Finally, we compare our results with the delta method.

References

[1] Ferson S, Kreinovich V, Grinzburg L, Myers D, Sentz K (2015) Constructing probability boxes and Dempster-Shafer structures. Sandia National Laboratories (Report SAND-2015-4166J), URL https://www.osti.gov/biblio/1427258

[2] JCGM 100:2008 (2008) Evaluation of measurement data – guide to the expression of uncertainty in measurement. URL https://www.bipm.org/en/publications/guides, accessed: 2019-05-24