The mathematical and statistical foundation of future data-driven engineering

Organisers: Mark Girolami, Omar Matar, Eric Moulines, Sebastian Reich, Aretha Teckentrup, Karen Willcox

January – June 2023

Background

We are in the midst of an unprecedented growth in the development and widespread availability of advanced technological infrastructure enabling measurement and modelling of physical, natural, and social systems. In his 1883 lectures on Electrical Units of Measurement, Lord Kelvin stated that when you can measure what you are speaking about, and express it in numbers, you know something about it: but when you cannot measure it, when you cannot express it in numbers, your knowledge is of a meagre and unsatisfactory kind. Continuing in this data-centric vein, contemporary science and engineering is being transformed at a fantastical rate due to our ability to measure and produce data as never before, driven by the increase in sensing technologies at all scales, large-scale scientific experiments, and the rapid growth of the web economy. Similar advances in the availability of computer-based modelling and simulation approaches to science and engineering are, as with data from measurement technologies, producing data at phenomenal rates giving rise to the so called Data Sciences and its mathematical underpinning. Data-driven formalisms, such as, Computational Statistics, Machine Learning, Artificial Intelligence, in particular have had an enormous impact on businesses and society due to their success in analysing and predicting human preferences from data. The Engineering sciences stand to benefit from this data revolution in similar ways, but realising this vision requires thoughtful, cross-disciplinary, and concerted efforts across the engineering and mathematical sciences.

Many **algorithmic developments** in science and engineering were first driven by technological advances within a specific application area. Deep neural networks are a well known example, but equally striking examples are provided by the finite element method, a direct result of the need to solve complex elasticity and structural analysis problems in civil and aeronautical engineering, and the ensemble Kalman filter (EnKF), simultaneously proposed in different areas of the geosciences and oceanography thanks to the emerging availability of Monte Carlo and ensemble prediction methods. A theoretical understanding of the methodologies emerged at later stages, which, in turn, triggered a secondary wave of wide-spread use in a wide and diverse range of application areas, some far removed from their origin. For example, the theoretical underpinning of the EnKF and its relation to the classical Kalman filter was at first unclear. More recently, a firmer theoretical understanding of the mathematical properties of the EnKF has emerged, and the EnKF has now found applications as a general tool for Bayesian inference and to machine learning in particular. Furthermore, fascinating connections to optimal control, McKean-Vlasov mean field equations, and Schrödinger bridges have also been established.

The tremendous acceleration of technological advances and an explosion in the available data bring the need for **transferable computational tools** and their **theoretical foundation** to the forefront more than ever. The six-month programme delivered a major stepping stone towards addressing these challenges against the background of specific engineering applications from

- Aerospace engineering,
- Chemical engineering,
- Computing Engineering
- Electrical Engineering
- Environmental and geotechnical engineering
- Mechanical Engineering.

More specifically, the programme contributed to a **synthesis** of approaches to engineering design and operation which leverages

- mathematical (mechanistic-based) and
- statistical (data-driven)

modeling frameworks. Whereas purely data-driven approaches are suitable in application areas where little is known about the processes generating the data, many others, such as the engineering sciences, natural sciences, finance and medicine, require the **seamless integration of mathematical models with observed data**. These mathematical models underpin our understanding of complex engineered systems, and provide crucial information complementing the observed data, which needs to be incorporated in our analysis. With the growth in computational power and its wide availability through cloud-computing technology, models can be designed to tackle increasingly complicated problems and to have direct impact on a widening range of engineering activities.

A further step is taken under the digital twin concept, where we do not only expect a model to represent a given phenomenon under fixed circumstances but, in addition, to **enable dynamical updates**, or adjustments of the model, in accordance with an evolving set of data and environmental conditions; this leads to challenging questions for

• online data assimilation and

• online model adaptation and model reduction.

The emerging digital twin concept hence goes far beyond traditional computational engineering and building a rigorous mathematical and statistical foundation has been at the core of the programme including efficient and robust

- optimisation and control, and
- decision-making

under uncertainties.

The overarching objective of the research programme has been to bring applied mathematicians, probabilists, statisticians, and engineers together to identify core challenges and formulate a common methodology, and to establish long-lasting interactions that will lead to advances in theory and methodologies for data-driven engineering.

Programme Outline

The programme attracted a total of around 150 international participants, with varied backgrounds in mathematics, statistics and engineering. Visits ranged in length from 1-2 weeks, especially during the focused deep-dive study periods detailed below, to several months and the entire duration of the programme.

Regular activities

Outside of workshop weeks, activities such as seminars, tutorials, and discussions were organised. In some weeks, a specified focus for these activities was chosen in advance as a designated deep-dive study period. In the remaining weeks, the activities were chosen spontaneously based on current participants, their interests and synergies between them. An informal chat over coffee was organised at the beginning of each week to welcome and introduce new participants, and further engagement and networking opportunities were provided at a daily lunch with all the participants.

Deep-dive study periods: To create the desired long-lasting interactions between applied mathematicians, statisticians, and engineers, the programme included the following deep-dive study periods:

- Mathematics and statistics for low carbon energy systems, 30/01-10/02 2023, organiser: Chris Dent
- From physics-based to data-driven assessment of structures, 13/02-17/02 2023, organiser: Tim Rogers
- Optimal Control and Inference, 20/02-03/03 2023, organisers: Susana Gomes and Sebastian Reich

- Data Driven Optimisation, 11/04-21/04 2023, organiser: Antonio del Rio Chanona
- Data Driven (modeling and control for) Fluid Mechanics, 15/05-26/05 2023, organisers: Luca Magri and Karthik Duraisamy
- Generative Models for Inverse Problems, 19/6-23/6 2023, organisers: Alain Durmus and Eric Moulines

Each deep-dive period lasted between one and two weeks and provided an introduction to a particular research topic through explanatory lectures and informal discussions. The precise structure of the deep-dive periods varied, for example starting with a full day of tutorials and the remainder of the week consisting of discussion groups based on interest, or having one talk scheduled every day for the first week and working through different topics in this way. As is usual for an INI programme, plenty of room was left for discussion and spontaneous scheduling of activities.

As evidenced by the numbers of programme participants in residence, the deep-dive periods were successful in attracting participants to the programme. Feedback provided in the INI scientific questionnaires, as well as informal feedback from participants, indicates that the focus provided by the deep-dive periods enabled the concurrent participation of the participants with similar research interests, also in the cases where the participants did not already know each other. Much of the participants feedback also refers to the deep-dive periods when talking about new research directions, collabrations and application areas, including the following:

- "Significantly broadened perspectives on existing lines of energy systems and climate resilience work."
- "Exploring the link between control theory and filtering was very enriching."
- "This [deep-dive period] brings me new connections between stochastic control and inverse problems."

In a recent article in Plus Maths, Chris Dent highlights the deep-dive period on energy systems as a "a catalyst for various conversations, drawing people from mathematics, computer science and [other areas] into energy innovation. They all have much to contribute."

Seminars and tutorials: Outside of workshops and deep-dive periods, typically one or two activities were scheduled per week. Early career researchers and newly arrived progamme participants were encouraged to present their work. The seminars usually stimulated a lot of questions and scientific debate, and often led to follow-on discussions (one-to-one or in larger groups) or further talks providing more details on specific topics.

Workshops and events

The series of programme workshops kicked off with a tutorial workshop at INI, and also included workshops on Modelling, Analysis, and Inference for Digital Twins (held at the Alan Turing Institute, London), Computational Challenges and Emerging Tools (at INI), and Mathematical Foundations (held at ICMS, Edinburgh). In addition, the programme included an Open for business event on Digital Twins for Engineering Applications, as well as the public lectures by the distinguished Rothschild visiting fellow Robert Scheichl and the distinguished Kirk visiting fellow Claudia Schillings.

Workshops: In more detail, the workshops associated the programme were the following:

• Tutorial workshop, 16/01-20/01 2023, INI.

Organisers: Antonio del Rio Chanona (Imperial College London), Karthik Duraisamy (University of Michigan), Luca Magri (Imperial College London), Sean Meyn (University of Florida), Sebastian Reich (Universität Potsdam)

To set the scene and facilitate communication among the various research communities present, it was important to start with a series of tutorials. This workshop consisted of 6 short courses in key programme areas in applied mathematics, statistics and engineering, including data-driven optimisation, reinforcement learning, physics aware machine learning, data assimilation, model order reduction and data-driven structural assessment.

• Modelling, Analysis and Inference for Digital Twins, 28/03-30/03 2023, Turing Institute London.

Organisers: Françcois-Xavier Briol (University College London), Tim Dodwell (University of Exeter), Steven Niederer (King's College London), Chris Oates (Newcastle University), Carola-Bibiane Schönlieb (University of Cambridge), Karen Willcox (University of Texas at Austin)

This workshop focussed on foundational aspects of digital twins, bringing together world-leading experts to deliver a major stepping stone in addressing research challenges against the backdrop of key application areas. The three days of the workshop concentrated on applications, numerical aspects and statistical aspects of digital twins, respectively.

• Computational Challenges and Emerging Tools, 24/04-28/04 2023, INI. Organisers: Youssef Marzouk (MIT), Omar Matar (Imperial College London), Aretha Teckentrup (University of Edinburgh), Irina Tezaur (Sandia National Laboratories)

The particular focus of this workshop was on highlighting computational challenges in the application areas of the DDE programme (aeronautics, chemical engineering, and civil and environmental engineering) and showcasing promising emerging tools. Recurrent themes in the talks included reduced order models, modelling and control of complex systems, model calibration and learning, and scientific machine learning.

• Mathematical Foundations, 08/05-12/05 2023, ICMS Edinburgh.

Organisers: Jonas Latz (University of Manchester), Eric Moulines (Ecole Polytechnique), Andrew Stuart (Caltech), Aretha Teckentrup (University of Edinburgh)

This workshop focussed on algorithms and methodology, with a particular focus on the underlying mathematical and statistical foundations required to make computational tools efficient and transferable. The contributions evolved especially around methodologies to incorporate data into physical models, which is at the core of data-driven engineering. Here, the talks covered statistical theory and the theory of statistical computation, data assimilation, stochastic models in data science and the natural sciences, machine learning and with physical models, and rigorous computational strategies for data-driven engineering. One of the speakers, Dr Hanne Kekkonen from TU Delft, gave a captivating open lecture on computational imaging, especially targeted at students, researchers, and academics from the Maxwell Institute for Mathematical Sciences to engage them into the mathematics of data-driven engineering.

Open for Business event: Supported by the Newton Gateway to Mathematics, a one-day event on *Digital Twins for Engineering Applications – The Emerging Science and Technology* was held at the INI on 7 June 2023.

Over the past decade, digital twins have reached new levels of sophistication, thanks to the decreasing cost of computation, the availability of low-cost, high-fidelity sensors, and advances in wireless communication technologies. As the science of digital twinning develops, artificial intelligence (AI) will play an ever more important role in their design, development, and deployment. The aim of this event was to bring together the community to discuss the latest research and innovation in digital twinning for engineering, provide opportunity for networking and help foster cross-disciplinary connections.

Talks presented at the event discussed applications and innovations of digital twins in energy systems, manufacturing, and smart cities, among others. In the INI scientific questionnaires, participants highlighted the event as being particularly valuable in learning about new application areas of their research.

Kirk Lecture: Claudia Schillings gave her public lecture entitled Uncertainty quantification for Bayesian inverse problems: efficient methods in the small noise regime on 11th April 2023. In this talk, Professor Schillings discussed efficient sampling methods for Bayesian inference in the practically desirable but computationally challenging setting of very informative data.

Rothschild Lecture: Robert Scheichl gave his public lecture entitled *From Quantum Entanglement to Future Data-Driven Engineering* on 3rd May 2023. In this talk, Professor Scheichl discussed key challenges and recent progress at the interface of data science/machine learning and numerical analysis/scientific computing. An interview with Professor Scheichl is available here on the INI's Youtube channel.

Impact & Future Directions

The INI programme has provided a coordinated focus on what is now being described as Data-Centric Engineering. Since the programme started in January 2023, we have seen significant impact in terms of community and capacitybuilding in the UK as a direct outcome of the INI programme.¹ These provide significant foundations for building on, and growing the impact from the programme.

There is no doubt that in the latter half of 2023, the emergence of certain Generative AI technologies (Large Language Models, for example) has accelerated the interest in the potential of Data Driven methodologies in all aspects of engineering (material design, structure design, system and process design and control. The recently established Cambridge University Press journal on Data Centric Engineering has seen an increase in submissions at the interface between so called AI and Engineering Sciences.

The EPSRC has established a number of AI Research Hubs some of which will have direct impact on further development of the Mathematical, Statistical, and Computing areas of research that provide the foundations of Data Centric Engineering, and many of the fundamental areas that have been initiated in the programme will be built upon via some of these AI Hubs.

The leverage of this EPSRC investment with the EPSRC backing of the Digital Twin Network Plus will ensure that this future direction of basic research in the mathematical and computing sciences will further develop the foundations laid down during the INI programme of 2023.

¹The Royal Statistical Society has started a special section on Statistical Engineering. The Royal Academy of Engineering has initiated a study into how Data Science and Engineering can be brought together in Data Centric Engineering at the international level. The EPSRC have awarded £3M to the Alan Turing Institute to establish the UK research capability and build capacity via national research network in the research foundations of Digital Twinning.