EDITORIAL

A computational mechanics special issue on: data-driven modeling and simulation—theory, methods, and applications

Wing Kam Liu[1] · George Karniadakis[2] · Shaoqiang Tang[3] · Julien Yvonnet[4]

1 Preface

There are more than a trillion sensors in the world today and according to some estimates there will be about 50 trillion cameras worldwide within the next 5 years, all collecting data either sporadically or around the clock. With such explosive growth of available data and computing resources, recent advances in machine learning and data analytics have yielded transformative results across diverse scientific disciplines, including image recognition, natural language processing, cognitive science, and genomics. However, in many engineering applications, quality and error-free data is not easy to obtain, e.g., for system dynamics characterized by bifurcations and instabilities, hysteresis, delayed responses, and often irreversible responses. Admittedly, as in all everyday applications, in engineering problems, the volume of data has increased substantially compared to even a decade ago but analyzing big data is expensive and time-consuming. *Data-driven methods, which have been enabled in the past decade by the availability of sensors, data storage, and computational resources, are taking center stage across many disciplines (physical and information) of science*. We now have highly scalable solutions for problems in object detection and recognition, machine translation, text-to-speech conversion, recommender systems, and information retrieval. All of these solutions attain state-of-the-art performance when trained with large amounts of data. However, purely data-driven approaches for machine learning present difficulties when the data is *scarce* and of *variable fidelity* relative to the complexity of the system. The vast majority of state-of-the art machine learning techniques (e.g., deep neural nets, convnets, RNNs, etc.) are lacking robustness and fail to provide any guarantees of convergence or quantify the error/uncertainty associated with their predictions. Hence, the ability to learn in a sample-efficient manner is a necessity in these data-limited domains. Less well understood is how to leverage the underlying physical laws and/or governing equations to extract patterns from small data generated from highly complex systems.

One example of open frontier in data-driven methods for mechanical science is the efficient and accurate description of heterogeneous material behavior that strongly depends on complex microstructure. This special issue will explore using mechanistic data-science multiscale finite element and numerical methods for material homogenization and concurrent multiscale analysis and design. One such approach, among many other reduced ordered methods, is the recently developed Self-Consistent Clustering Analysis (SCA) concurrent homogenization, which was developed to directly generating material laws on-the-fly, using an efficient two-stage solution to compute microscale material response from a statistically Representative Volume Element (RVE). The first stage, known as the offline or training stage, uses data science theories such as k-means clustering and self-organizing maps to "compress" the RVE. Next, the "prediction" stage solves the Lippmann–Schwinger equation to determine the response of each compressed RVE (CRVE) to arbitrary applied load with any constitutive relationship. The CRVE may then be considered a material point in the larger concurrent simulation. The SCA theory integrates multiscale mechanics of materials and data science theories to efficiently generate accurate material laws with

✉ Wing Kam Liu
w-liu@northwestern.edu

George Karniadakis
George_Karniadakis@brown.edu

Shaoqiang Tang
maotang@pku.edu.cn

Julien Yvonnet
julien.yvonnet@univ-mlv.fr

1   Northwestern University, Evanston, USA

2   Brown University, Providence, USA

3   Peking University, Beijing, China

4   Université Paris-Est, Champs-sur-Marne, France

a drastic reduction of computational cost over conventional approaches. Prediction comparisons with direct numerical simulations and experiments of nonlinear behavior for metal alloys, nano-polymer composites, and polymer matrix composites are encouraging. These use different constitutive laws within the CRVEs; in each case, computational expense is decreased substantially. This is just one example on the applications of data science methods and SCA to nonlinear behavior of advanced and additive manufacturing and jointing technologies, among many others.

To explore the future development and the adaptation of data-driven methods, new mathematical and computational paradigms and broad flexible frameworks are needed, which can lead to probabilistic predictions using the minimum amount of information that can be processed expeditiously and be sufficiently accurate for decision making under uncertainty. Integrating *multi-fidelity data* into large-scale simulations is necessary to speed up the computation but also to deal with the "hidden physics" not captured by the lack of resolution or the lack of proper constitutive laws or boundary conditions. Realizing the concept of "digital twin" requires advances in many fronts and probabilistic data-driven modeling approaches that quantify uncertainty as a key algorithmic component. Statistical learning can help formulate new concepts and promote machine learning methods that are appropriate for problems in the various fields of computational mechanics, where we also know the conservation laws of mass, momentum and energy but we need data to fully describe the system in terms of boundary conditions and uncertain constitutive laws. Data assimilation is not a new field and has been going on for years in the geophysics community but less so in computational mechanics. Deep learning provides multiple opportunities of fusing data and simulations in a seamless manner creating a new paradigm in the form of physics informed learning machines. In addition, the concepts of *active learning* and *transfer learning* are particularly useful and potentially cost-effective for the digital twin paradigm. Active learning will use uncertainties in the predictions to re-locate the sensors or add more sensors to increase accuracy so it enables in practice the long-standing aim of adaptive sampling in data gathering. Transfer learning is equally important as it exploits the knowledge gained to new but similar situations, hence requiring only a small amount of data and not attempting to learn from scratch.

This special issue involves experts from diverse fields in computational mechanics and mathematics to contribute to different approaches to data-driven modeling and simulation, with emphasis on some of the aforementioned modern topics. Applications range from effective thermo-mechanical properties of nonlinear heterogeneous material, prediction of dynamic systems, like e.g. flow field and free surface motion in fluids, impact analysis, shock-to-detonation transition in energetic materials and uncertainty quantification such as high dimensional probability distributions identification and propagation, Bayesian inverse problems, and multi-fidelity modeling of stochastic processes. A large spectrum of most recent methodologies are investigated and reviewed, including Neural networks, Convolutional Neural Networks, Deep Generative Networks (DGN), Deep Material network (DMN), Locally Linear Embedding (LLE), Topological Data Analysis (TDA), General Equation for Non-Equilibrium Reversible-Irreversible Coupling (GENERIC), self-consistent clustering analysis (SCA), virtual clustering analysis (VCA), FEM clustering analysis (FCA), and parametric Gaussian processes (PGP).

The special issue has sixteen invited papers, broken down into four groups of contributions. The organization of these sixteen papers is as follows. The first group of papers is related to new methods in speeding up computational homogenization with the help of machine learning tools and for making materials characterization model-free. Hengyang Li, et al., summarized a class of clustering discretization methods for generation of material performance databases in machine learning and design optimization. Xiaoxin Lu, et al., proposed a data-driven computational homogenization method based on neural networks for the nonlinear anisotropic electrical response of graphene/polymer nanocomposites. Yinghao Nie, et al., discussed the principle of cluster minimum complementary energy of FEM-cluster-based reduced order method: fast updating the interaction matrix and predicting effective nonlinear properties of heterogeneous material. Lei Zhang, et al., studied the fast calculation of interaction tensors in clustering-based homogenization and extended VCA to solve finite strain problems. Hang Yang, et al., derived heterogeneous material laws via data-driven principal component expansions. Laurent Stainier, et al., proposed model-free data-driven methods in mechanics material data identification and solvers. Adrien Leygue, et al., proposed a non-parametric material state field extraction method from full field measurements.

The second group of three papers is related to Machine learning for uncertainty quantification in high dimensions and regression methods in big data. Thomas Y. Hou, et al., proposed to solve Bayesian inverse problems from the perspective of Deep Generative Networks. Maziar Raissi, et al., introduced parametric Gaussian process regression for big data. Yibo Yang and Paris Perdikaris developed conditional deep surrogate models for stochastic, high-dimensional and multi-fidelity systems.

The third group of four papers describes mechanistic machine learning strategies for system identification. Guorong Chen, et al., applied Deep Learning Neural Network to identify collision load conditions based on permanent plastic deformations of shell structures. Zeliang Liu, et al., addressed the transfer learning of deep material network for

seamless structure–property predictions. Kun Wang, et al., developed a cooperative game for automated learning of elasto-plasticity knowledge graphs and models with AI-guided experimentation.

The fourth group of three papers is concerned with data-driven algorithms for computational fluid dynamics. Beatriz Moya et al., proposed to learn sloshing dynamics by means of data. Saakaar Bhatnagar, et al., applied convolutional neural networks to aerodynamics flow, and Yaochi Wei, et al., performed integrated Lagrangian and Eulerian 3D micro-structure-explicit simulations for predicting macroscopic probabilistic shock-to-detonation transition thresholds of energetic materials.

Table of Content

1. Clustering discretization methods for generation of material performance databases in machine learning and design optimization
2. A data-driven computational homogenization method based on neural networks for the nonlinear anisotropic electrical response of graphene/polymer nanocomposites
3. Principle of cluster minimum complementary energy of FEM-cluster-based reduced order method: fast updating the interaction matrix and predicting effective nonlinear properties of heterogeneous material
4. Fast calculation of interaction tensors in clustering-based homogenization
5. Derivation of heterogeneous material laws via data-driven principal component expansions
6. Model-free data-driven methods in mechanics: material data identification and solvers
7. Solving Bayesian inverse problems from the perspective of deep generative networks
8. Parametric Gaussian process regression for big data
9. Conditional deep surrogate models for stochastic, high-dimensional, and multi-fidelity systems
10. Application of deep learning neural network to identify collision load conditions based on permanent plastic deformation of shell structures
11. Transfer learning of deep material network for seamless structure–property predictions
12. A cooperative game for automated learning of elasto-plasticity knowledge graphs and models with AI-guided experimentation
13. Non-parametric material state field extraction from full field measurements
14. Learning slosh dynamics by means of data
15. Prediction of aerodynamic flow fields using convolutional neural networks
16. Integrated Lagrangian and Eulerian 3D microstructure-explicit simulations for predicting macroscopic probabilistic SDT thresholds of energetic materials