



Steven L. Brunton · Maziar S. Hemati · Kunihiko Taira

## Special issue on machine learning and data-driven methods in fluid dynamics

Received: 17 June 2020 / Accepted: 23 June 2020  
© Springer-Verlag GmbH Germany, part of Springer Nature 2020

Machine learning (i.e., modern data-driven optimization and applied regression) is a rapidly growing field of research that is having a profound impact across many fields of science and engineering. In the past decade, machine learning has become a critical complement to existing experimental, computational, and theoretical aspects of fluid dynamics. In this short article, we are excited to introduce this special issue highlighting a number of promising avenues of ongoing research to integrate machine learning and data-driven techniques in the field of fluid dynamics. We will also attempt to provide a broader perspective, outlining recent successes, opportunities, and open challenges, while balancing optimism and skepticism.

In the field of fluid dynamics, there is an interesting parallel between the rise of machine learning in recent years and the rise of computational science decades earlier. Neither approach fundamentally changes the scientific questions being asked nor the higher-level objectives. Rather, both approaches provide sophisticated tools for analysis based on emerging technologies, enabling the community to address scientific questions at a greater scale and a broader scope than was previously possible. In the early years of computational fluid dynamics, there were voices of both extreme skepticism and open-ended optimism that these new approaches would supplant existing techniques. In reality, computational techniques have provided another valuable perspective for scientific inquiry, complementing more traditional approaches. It is therefore reasonable to believe that machine learning and data-intensive analysis will have a similar impact, complementing other well-established techniques to expand our collective capabilities.

Machine learning offers a wealth of techniques to discover patterns in high-dimensional data [5, 6], extending traditional modal expansions that have been a cornerstone of fluid dynamics for decades [26, 27]. Despite this great potential, it is important to recognize that these algorithms must be used properly, and that a single tool alone will not be equipped to address every task. The same factors that make data-driven and machine learning methods so appealing—namely, that they are relatively easy to use and do not require expert knowledge—can also serve as a potential downfall. Much as users of CFD software are cautioned against blind application of these numerical tools without proper knowledge, training, and verification/validation, we must adopt a similar philosophy for the use of data-driven and machine learning tools.

---

Communicated by Tim Colonius.

---

S. L. Brunton (✉)  
Department of Mechanical Engineering, University of Washington, Seattle, WA 98195, USA  
E-mail: sbrunton@uw.edu

M. S. Hemati  
Department of Aerospace Engineering and Mechanics, University of Minnesota, Minneapolis, MN 55455, USA  
E-mail: mhemati@umn.edu

K. Taira  
Department of Mechanical and Aerospace Engineering, University of California, Los Angeles, CA 90095, USA  
E-mail: ktaira@seas.ucla.edu

Many in the fluid mechanics community are asking why machine learning and data science techniques have received a surge in attention over the past few years. Indeed, many machine learning and data-driven methods were introduced and studied within fluid dynamics decades ago [6,21]. Yet, these past studies did not benefit from more recent advances in machine learning, which have been driven by the present confluence of big data, high-performance computing, advanced algorithms, and considerable investment by industry, leading to open-source tools and benchmark datasets. One of the most significant milestones in the modern machine learning era is the rise of deep learning [13] and the achievement of human-level performance in several highly challenging tasks, such as image recognition [15] and in control [18,22]. The success of these methods has been based primarily on large labeled data sets [9], open software for reproducible research, and powerful parallel computations in consumer GPU hardware. There have also been significant advancements from the data science community with eyes towards handling, analyzing, compressing, and transferring massively large data sets. Novel methods leveraging randomized numerical linear algebra, sparsity promoting norms, and efficient optimization techniques have become powerful enablers for the analysis of big data.

It is noteworthy that many of the enabling technologies in machine learning have deep roots in the computational fluid mechanics community [21]. Fluid dynamics researchers have pushed the frontiers of data storage and transfer capabilities, computational hardware, and scalable algorithms. As an example, consider that in the same year that snapshot proper orthogonal decomposition (POD) was introduced to the fluids community [23], the same algorithm was also introduced to the image sciences community to efficiently represent human faces [24], creating a major branch of modern data-intensive image analysis. In the past decade, machine learning has made great advances, driven by commercial successes in the fields of image sciences, natural language processing, and advertising. These approaches are now coming full-circle, and fluids researchers are increasingly leveraging these powerful techniques for the modeling and control of flow physics. What is exciting for the field of fluid mechanics is that these emerging approaches are sufficiently powerful to handle large data sets describing complex nonlinear dynamics that are commonly encountered in fluid flow analysis.

The use of machine-learning and data-science inspired approaches should be encouraged to solve problems in fluid dynamics, especially those that are difficult to solve with traditional methods. Many goals in fluid dynamics, such as analysis, modeling, sensing, estimation, design optimization, and control, may be posed as optimization problems. These problems are challenging because fluids are nonlinear and multiscale in space and time, resulting in high-dimensional and non-convex optimization landscapes. Fortunately, machine learning is improving our ability to tackle these traditionally intractable optimization problems. In addition to the critical tasks of benchmarking and validating new approaches with canonical problems, we must strive to use these emerging techniques to go beyond what is possible with existing techniques to reveal physical insights or improve analytical capabilities. Yet, it is important to understand that machine learning is not a panacea that will make domain researchers obsolete. Quite to the contrary, applying these techniques effectively requires considerable expertise in how to formulate the relevant optimization problem, in generating and curating the training data, in selecting or designing the relevant machine learning architecture, and in evaluating the results to extract correct physical insights.

Fluid dynamics, and physics in general, presents a unique set of challenges and opportunities for machine learning. Although fluids data are vast in some dimensions, it may be quite sparse in others. Let us also emphasize that data-driven techniques rely on the quality and the coverage of the training data, following the garbage-in garbage-out principle. Moreover, the coverage of the training data is critical to ensure robustness, as extrapolation is typically unreliable and may lead to catastrophic results. For these reasons, fluid flow data must be offered with reliability and accuracy to data-driven approaches. If fluid dynamics is to capitalize on these advances, we must embrace open-source software and open data, such as the Johns Hopkins Turbulence Database [20], the Stanford Center for Turbulence Research database [28], and the NOAA database [1]. The use of data-driven techniques for fluid dynamics should be solidly founded on the ability to conduct high-quality fluid mechanics research. The natural conclusion is that in the age of data-driven fluid dynamics [10,21], the performance of high-quality numerical simulations and experiments is ever more important.

## Summary of articles in this special issue

In light of the emerging developments and applications of machine learning and data-driven techniques in fluid mechanics, we have put together a collection of papers written by some of the active research groups in these areas. These invited papers are published in a double issue entitled *Machine Learning and Data-Driven Methods in Fluid Dynamics in Theoretical and Computational Fluid Dynamics*. Covered in these issues are a

wide range of topics, including analysis, modeling, numerical algorithms, sensing and estimation, and control for fluid flows with data-driven techniques. We summarize these papers below.

Several papers in this issue use machine learning for the reduced-order modeling of fluid dynamic phenomena. Loiseau [16] has used modal decomposition and sparse regression to obtain a nonlinear reduced-order model of a thermosiphon flow. The thermosiphon is an annular flow driven by a thermal instability, and it exhibits chaotic oscillations. Loiseau showed that these oscillations may be related to the simple Lorenz system, providing an interpretable nonlinear model for this complex flow. Hasegawa et al. [14] trained a machine-learned reduced order model (ML-ROM) based on a convolutional neural network autoencoder (CNN-AE) and a long short-term memory (LSTM) to capture the evolution of the laminar bluff body wakes. As their training data, flows over bodies of arbitrary body shapes are considered and are tested for unseen wakes. Their approach takes advantage of the low-dimensional nature of the latent space for laminar wake dynamics. Mendible et al. [17] developed a data-driven optimization procedure to segment and extract multiple traveling waves in fluid simulations. Typical dimensionality reduction techniques, such as POD and DMD, often fail to capture symmetries and invariants, such as translation invariance. This work uses sparse optimization to extend dimensionality reduction to include known symmetries.

Closure modeling is one of the most promising avenues of machine learning research in fluid dynamics. Balachandar et al. [2] use machine learning to model the hydrodynamic forces within particle-laden flows, providing closures that make a large step toward fully particle-resolved DNS. They show that by accounting for interactions with neighboring particles, their closure models are able to better predict the hydrodynamic forces within Euler–Lagrange simulations. These models would not have been possible without the synergistic combination of both physics and machine learning. In the work of Pawar et al. [19], an artificial neural network (multilayer perceptron) and a convolutional neural network are used to construct data-driven subgrid-scale closure models for two-dimensional turbulence. For the development of the models, they consider a number of spatial points and variables in learning the subgrid-scale stresses at a spatial point. They find that in comparison with the dynamic Smagorinsky model, the machine-learned model can accelerate computations with comparable accuracy. They also provide some cautionary notes on the use of data-driven models with importance of having checking mechanisms, especially for safety-critical applications.

The use of data science and machine learning has also stimulated the development of novel algorithms for numerical simulations of fluid flows. Foti et al. [11] leverage recent techniques in low rank decompositions and sparse sampling for efficient adaptive mesh refinement (AMR). Stevens and Colonius [25] have developed an enhanced shock capturing scheme that trains a neural network to improve the performance of a WENO method. The resulting WENO-NN combines favorable aspects of existing numerical methods with the expressive power of neural networks, resulting in performance gains in regions of high numerical viscosity.

Flowfield and parameter estimation are naturally data-centric problems and have also benefited from recent advances in machine learning and data-driven methods, as is highlighted in this special issue. Fukami et al. [12] surveyed the use of supervised machine learning techniques for flow estimation problems. Their work covers the use of multilayer perceptron, random forest, support vector regression, extreme learning machine, and convolutional neural networks on a number of canonical laminar and turbulent flows. Recommendations on how to perform the learning process and construct the networks for achieving accurate and robust estimations of flow fields from limited data are offered. In the work of Canuto et al. [7], data assimilation is used to develop cardiovascular models from clinical measurements. They use an ensemble Kalman filter (EnKF) to estimate model parameters for patient-specific models, in an effort to support clinicians to perform complex cardiovascular analysis on a reduced time frame for planning and diagnosis.

Recent trends in machine learning have provided powerful techniques for data-driven optimization and flow control. In this issue, Colvert and Kanso [8] have developed a sensing and control strategy that learns to orient the sensors in directions of maximal information. Although this flowtaxis is biologically inspired, there are many applications for such an adaptive sensing and control strategy. This work fits into a larger effort by the authors to efficiently and accurately classify wakes using techniques from machine learning. Bhattacharjee et al. [3] present a data-driven method for determining the optimal actuator location for controlling separated flow over an airfoil with minimal input energy. They also introduce complementary methods to extract controllable flow structures from flowfield data, helping identify physical mechanisms and instabilities that are exploited by the optimal actuation. Bieker et al. [4] have combined deep learning with model predictive control to manipulate the forces on bluff bodies. Instead of using a computationally expensive flow simulation or developing a reduced-order model with intrusive methods, they use a neural network to develop a surrogate model for bluff body flows with actuation, based on limited measurements of the forces on the bodies.

These articles collectively mark an exciting and growing avenue of research to bring machine learning and data science to bear on pressing problems in fluid dynamics. We believe that the applications of data-driven techniques are only in the early stage of their development and will continue to thrive over this decade. During that process, we expect that the fluid mechanics community will capitalize on the strengths of machine learning and data science, developing novel algorithms and gaining insights into the complex dynamics of fluid flows. The intersection of fluid mechanics and data science will continue to grow in importance, and we hope that this special issue provides inspiration for new research.

**Acknowledgements** We thank Tim Colonius, Kozo Fujii, Koji Fukagata, Petros Koumoutsakos, Nathan Kutz, Jean-Christophe Loiseau, Bernd Noack, and Peter Schmid for stimulating discussions on data-driven methods and their applications to fluid dynamics.

## References

1. <https://www.ncdc.noaa.gov/data-access>
2. Balachandar, S., Moore, W.C., Akiki, G., Liu, K.: Towards particle-resolved accuracy in Euler-Lagrange simulations of multiphase flow using machine learning and pairwise interaction extended point-particle (PIEP) approximation. *Theor. Comput. Fluid Dyn.* **34**(3) (2020). <https://doi.org/10.1007/s00162-020-00538-8>
3. Bhattacharjee, D., Klose, B., Jacobs, G.B., Hemati, M.S.: Data-driven selection of actuators for optimal control of airfoil separation. *Theor. Comput. Fluid Dyn.* **34**(3) (2020). <https://doi.org/10.1007/s00162-020-00526-y>
4. Bieker, K., Peitz, S., Brunton, S.L., Kutz, J.N., Dellnitz, M.: Deep model predictive flow control with limited sensor data and online learning. *Theor. Comput. Fluid Dyn.* **34**(3) (2020). <https://doi.org/10.1007/s00162-020-00520-4>
5. Brenner, M., Eldredge, J., Freund, J.: Perspective on machine learning for advancing fluid mechanics. *Phys. Rev. Fluids* **4**(10), 100501 (2019)
6. Brunton, S.L., Noack, B.R., Koumoutsakos, P.: Machine learning for fluid mechanics. *Annu. Rev. Fluid Mech.* **52**, 477–508 (2020)
7. Canuto, D., Pantoja, J.L., Han, J., Dutson, E.P., Eldredge, J.D.: An ensemble Kalman filter approach to parameter estimation for patient-specific cardiovascular flow modeling. *Theor. Comput. Fluid Dyn.* **34**(3) (2020). <https://doi.org/10.1007/s00162-020-00530-2>
8. Colvert, B., Kano, E.: Flow taxis in the wake of oscillating airfoils. *Theor. Comput. Fluid Dyn.* **34**(3) (2020)
9. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: Imagenet: A large-scale hierarchical image database. In: 2009 IEEE conference on computer vision and pattern recognition, pp. 248–255. IEEE (2009)
10. Duraisamy, K., Iaccarino, G., Xiao, H.: Turbulence modeling in the age of data. *Annu. Rev. Fluid Mech.* **51**, 357–377 (2019)
11. Foti, D., Giorno, S., Duraisamy, K.: An adaptive mesh refinement approach based on optimal sparse sensing. *Theor. Comput. Fluid Dyn.* **34**(3) (2020). <https://doi.org/10.1007/s00162-020-00522-2>
12. Fukami, K., Fukagata, K., Taira, K.: Assessment of supervised machine learning methods for fluid flows. *Theor. Comput. Fluid Dyn.* **34**(3) (2020). <https://doi.org/10.1007/s00162-020-00518-y>
13. Goodfellow, I., Bengio, Y., Courville, A.: *Deep Learning*. MIT Press, Cambridge (2016)
14. Hasegawa, K., Fukami, K., Murata, T., Fukagata, K.: Machine-learning-based reduced-order modeling for unsteady flows around bluff bodies of various shapes. *Theor. Comput. Fluid Dyn.* **34**(3) (2020). <https://doi.org/10.1007/s00162-020-00528-w>
15. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: *Advances in neural information processing systems*, pp. 1097–1105 (2012)
16. Loiseau, J.-C.: Data-driven modeling of the chaotic thermal convection in an annular thermosyphon. *Theor. Comput. Fluid Dyn.* **34**(3) (2020). <https://doi.org/10.1007/s00162-020-00536-w>
17. Mendible, A., Brunton, S.L., Aravkin, A.Y., Lowrie, W., Kutz, J.N.: Dimensionality reduction and reduced-order modeling for traveling wave physics. *Theor. Comput. Fluid Dyn.* **34**(3) (2020). <https://doi.org/10.1007/s00162-020-00529-9>
18. Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A.A., Veness, J., Bellemare, M.G., Graves, A., Riedmiller, M., Fidjeland, A.K., Ostrovski, G., et al.: Human-level control through deep reinforcement learning. *Nature* **518**(7540), 529 (2015)
19. Pawar, S., San, O., Rasheed, A., Vedula, P.: A priori analysis on deep learning of subgrid-scale parameterizations for Kraichnan turbulence. *Theor. Comput. Fluid Dyn.* **34**(3) (2020). <https://doi.org/10.1007/s00162-019-00512-z>
20. Perlman, E., Burns, R., Li, Y., Meneveau, C.: Data exploration of turbulence simulations using a database cluster. In: *ACM/IEEE Conference Supercomp.*, p. 23 (2007)
21. Pollard, A., Castillo, L., Danaïla, L., Glauser, M.: *Whither Turbulence and Big Data in the Twenty First Century?*. Springer, Berlin (2016)
22. Silver, D., Hubert, T., Schrittwieser, J., Antonoglou, I., Lai, M., Guez, A., Lanctot, M., Sifre, L., Kumaran, D., Graepel, T., et al.: A general reinforcement learning algorithm that masters chess, shogi, and go through self-play. *Science* **362**(6419), 1140–1144 (2018)
23. Sirovich, L.: Turbulence and the dynamics of coherent structures, parts I-III. *Quart. Appl. Math.* **XLV**(3), 561–590 (1987)
24. Sirovich, L., Kirby, M.: A low-dimensional procedure for the characterization of human faces. *J. Opt. Soc. Am. A* **4**(3), 519–524 (1987)
25. Stevens, B., Colonius, T.: Enhancement of shock-capturing methods via machine learning. *Theor. Comput. Fluid Dyn.* **34**(3) (2020). <https://doi.org/10.1007/s00162-020-00531-1>
26. Taira, K., Brunton, S.L., Dawson, S., Rowley, C.W., Colonius, T., McKeon, B.J., Schmidt, O.T., Gordeyev, S., Theofilis, V., Ukeiley, L.S.: Modal analysis of fluid flows: an overview. *AIAA J.* **55**(12), 4013–4041 (2017)

27. Taira, K., Hemati, M.S., Brunton, S.L., Sun, Y., Duraisamy, K., Bagheri, S., Dawson, S., Yeh, C.A.: Modal analysis of fluid flows: applications and outlook. *AIAA J.* **58**(3), 998–1022 (2020)
28. Wu, X., Moin, P.: A direct numerical simulation study on the mean velocity characteristics in turbulent pipe flow. *J. Fluid Mech.* **608**, 81–112 (2008)

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.