Data Driven Fluid Mechanics

Combining First Principles and Machine Learning

MIGUEL A. MENDEZ
von Karman Institute for Fluid Dynamics, Belgium

ANDREA IANIRO
Universidad Carlos III de Madrid, Spain

BERND R. NOACK
Harbin Institute of Technology, China
Technische Universität Berlin, Germany

STEVEN L. BRUNTON
University of Washington, USA

Contents

Preface

This book is for scientists and engineers interested in data-driven and machine learning methods for fluid mechanics. Big data and machine learning are driving profound technological progress across nearly every industry, and they are rapidly shaping fluid mechanics' research. This revolution is driven by the ever-increasing amount of high-quality data, provided by rapidly improving experimental and numerical capabilities. Machine learning extracts knowledge from data without the need for first principles and introduces a new paradigm: use data to discover, rather than validate, new hypotheses and models. This revolution brings challenges and opportunities.

Data driven methods are an essential part of the methodological portfolio of fluid dynamicists, motivating students and practitioners to gathering practical knowledge from a diverse range of disciplines. These fields include computer science, statistics, optimization, signal processing, pattern recognition, nonlinear dynamics, and control. Fluid mechanics is historically a *big data* field and offers a fertile ground to develop and apply data-driven methods, while also providing valuable shortcuts, constraints, and interpretations based on its powerful connections to first principles physics. Thus, hybrid approaches that leverage both data-driven methods and first principles approaches, are the focus of active and exciting research. This book presents an overview and a pedagogical treatment of some of the data-driven and machine learning tools that are leading research advancements in model-order reduction, system identification, flow control, and data-driven turbulence closures.

About the Book and the VKI Lecture Series

This book originated from a one-week course from the von Karman Institute (VKI) for fluid dynamics (https://www.vki.ac.be/). The course was hosted by the Université libre de Bruxelles (ULB) from 24 to 28 February 2020, in the classic VKI lecture series format. These are one-week courses on specialized topics, selected by the VKI faculty and typically organized 8-12 times per year. These courses have gained a worldwide recognition and are among the most influential and distinguished European teaching forums, where pioneers in fluid mechanics have been training young talents for many decades.

The lecture series was co-organized by Miguel A. Mendez from the von Karman Institute (Belgium), Alessandro Parente from the Université libre de Bruxelles (Belgium), Andrea Ianiro from Universidad Carlos III de Madrid (Spain), Bernd R. Noack from Harbin Institute of Technology, Shenzhen (China) and TU Berlin (Germany) and Steven L. Brunton from University of Washington (US).

Online Material

The book is supported by supplementary material, including codes, experimental and numerical data, exercises, and the video lectures recorded from the course. All material is hosted on the course website:

https://www.datadrivenfluidmechanics.com/

The supplementary material covers more exercises, tutorials, and practicalities than could be included in this book while preserving its conciseness. Readers interested in gaining a working knowledge on the subject are encouraged and expected to download this material, study it along with the book, and test it on their own data. The large repertoire of computing tools implemented, together with the relevant datasets provided, offer a unique opportunity to learn by practicing with real experimental and numerical data.

The Audience

The book is intended for anyone interested in the use of data-driven methods for fluid mechanics. We believe that the book provides a unique balance between introductory material, practical hands-on tutorials, and state-of-the-art research. While keeping the approach pedagogical, the reader is exposed to topics at the frontiers of fluid mechanics research. Therefore, the book could be used to complement or support classes on data-driven science, applied mathematics, scientific computing, and fluid mechanics, as well as to serve as a reference for engineers and scientists working in these fields. Basic knowledge of data processing, numerical methods, and fluid mechanics is assumed.

The Book's Roadmap

Like the course from which it originates, this book results from the contribution of many authors. The use of machine learning methods in fluid mechanics is in its early days, and a large team of lecturers allowed the course attendees to learn from the expertise and perspectives of leading scientists in different fields.

Here we provide a roadmap of the book to guide the reader through its structure and link all the chapters into a coherent narrative. The book chapters can be clustered into six interconnected parts, slightly adapted from the VKI lecture series.

Part I: Motivation. This part includes the first three chapters, which introduce the motivation for data-driven techniques from three perspectives.

Chapter 1, by B.R. Noack and coauthors, opens with a tour de force on machine learning tools for dimensionality reduction and flow control. These techniques are introduced to analyze, model, and control the well-known cylinder wake problem, building confidence and intuition about the challenges and opportunities for machine learning in fluid mechanics. **Chapter 2**, by J. Jiménez, takes a step back and gives both a historical and a data-science perspective. Most of the dimensionality reduction techniques presented in this book have been developed to identify patterns in the data, known as *coherent structures* in turbulent flows. But what are coherent structures? This question is addressed by discussing the relationship between data analysis and conceptual modelling and the extent to which artificial intelligence can contribute to these two aspects of the scientific method. **Chapter 3**, by S. Brunton, gives an overview of how machine learning tools are entering fluid mechanics. The chapter provides a short introduction to machine learning, its categories (e.g. supervised versus unsupervised learning), its subfields (regression and classification, dimensionality reduction and clustering) and the problems in fluid mechanics that can be addressed by these methods (e.g. feature extraction, turbulence modelling and flow control). This chapter contains a broad literature review, highlights the key challenges of the field, and gives perspectives for the future.

Part II: Methods from Signal Processing. This part brings the reader back to classic tools from signal processing, usually covered in curricula crossed by experimental fluid dynamicists, although with a large variety of depth. This part of the book is motivated by two reasons. First, tools from signal processing are, and will likely remain, the first 'off-the shelf' solutions for many practical problems. Examples include filtering, time-frequency analysis, and data compression using filter banks or wavelets, or the use of linear system identification and time series analysis via autoregressive methods. The second reason – and this is a central theme of the book – is that much can be gained by combining machine learning tools with methods from classic signal processing, as later discussed in Chapter 8. Therefore, **Chapter 4**, by M. A. Mendez, reviews the theory of linear time-invariant (LTI) systems along with their properties and the fundamental transforms used in their analysis: the Laplace, Fourier, and Z transforms. This chapter draws several parallels with more advanced techniques. For example, the use of the Laplace transform to reduce ordinary differential equations (ODEs) to algebraic equations parallels the use of Galerkin methods to reduce the Navier Stokes equation to a system of ODEs. Similarly, there is a link between the classical Z-transform and the modern dynamic mode decomposition (DMD). **Chapter 5**, by S. Discetti, complements the previous chapter by focusing on time-frequency analysis. The fundamental Gabor and continuous/discrete Wavelet transforms are introduced along with the related Heisenberg uncertainty principle and multiresolution analysis. The methods are illustrated on a time-series obtained from hot-wire anemometry in a turbulent boundary layer and from flow fields obtained via numerical simulations.

Part III: Data-Driven Decompositions. This part of the book consists of four chapters dedicated to a cornerstone (and rapidly growing sub-field) of fluid mechanics: modal analysis. This part is mostly concerned with methods for linear dimensionality reduction, originally introduced to identify, and "objectively" define, coherent structures in turbulent flows.

Chapter 6, by S. Dawson, is dedicated to the proper orthogonal decomposition (POD), the first and most popular tool introduced in the fluid mechanics community in the 1970s. The chapter reviews the link between POD with the singular value decomposition (SVD), its essential properties (e.g. optimality, relation to eigenvalue decomposition, and generalization to weighted inner products), its practical computation on discrete datasets, and its extension to continuous systems. This chapter closes with illustrative exercises that guide the reader to practical computation. **Chapter 7**, by P. Schmid, is dedicated to the dynamic mode decomposition (DMD), a powerful alternative to POD introduced by P. Schmid a decade ago. This chapter reviews the derivation of DMD and its roots in dynamical systems and Koopman operator theory. The main DMD algorithm is presented along with its "sparsity promoting" variant, and the chapter is enriched by three applications to experimental and numerical data, as well as a brief outlook at new extensions and generalizations.

Chapter 8, by M. A. Mendez, presents a generalized framework for deriving, computing, and interpreting *any* linear decomposition. Modal decompositions are analyzed in terms of matrix factorization and viewed as a special case of 2D discrete transforms. This framework is used to combine multiresolution analysis via filter banks with the classic POD, and derive the multiscale POD (mPOD). The mPOD is a recent decomposition that generalized the energy-based (POD-like) and the frequency-based (DMD-like) formalism. The chapter includes several exercises and tutorials, allowing the reader to test these decompositions on experimental data. Finally, this part on modal analysis closes with **Chapter 9**, by A. Ianiro, with an overview of good practices and applications of modal analysis. This chapter addresses essential questions on the statistical convergence of POD, the impact of random noise, and the possibility to extract phase information about the modes even if the data is not time-resolved. Moreover, the chapter presents interesting applications of the extended POD – in which decompositions of different datasets are correlated – to experimental and numerical data.

Part IV: Dynamical Systems. This part of the book consists of four chapters dedicated to various aspects of dynamical systems. **Chapter 10**, by S. Dawson, gives a brief overview of linear dynamical systems and linear control. This is one of the most developed disciplines in engineering, with applications across robotics, automation, aeronautics, and mechanical systems in general. Linear techniques provide a standard approach for closed-loop control and have been successfully used in fluid flows. This chapter illustrate the main concepts (state-space representation, controllability and observability, and optimal control) and tools (root locus, pole placement, PID controllers) focusing on a specific example from fluid mechanics, namely the stabilization of a wake flow. An overview of additional control techniques and a brief literature review for flow control are also provided.

Chapter 11, by S. Brunton, provides an overview of nonlinear dynamical systems.

The chapter introduces fundamental concepts such as flow maps, attracting sets and bifurcations, and gives a modern perspective on the field, with its current goals and open challenges. These include recent advances in the operator-theoretic views that seek to identify a linear representation of nonlinear systems and identify dynamical systems from data, further discussed in the following chapter. **Chapter 12**, also by S. Brunton, builds on the previous chapter and Part III of the book to introduce several advanced topics in model reduction and system identification. The chapter opens with a review of balanced model reduction goals for linear systems and builds the required mathematical background and the fundamentals of balanced POD (BPOD). Linear and nonlinear identification tools are introduced. Among the linear identification tools, the chapter presents the eigensystem realization algorithm (ERA) and the observer Kalman filter identification (OKID). Among the nonlinear identification tools, the chapter presents the sparse identification of nonlinear dynamics (SINDy) algorithm, which leverages the LASSO regression from statistics to identify nonlinear systems from data.

This part closes with **Chapter 13**, by P. Schmid, providing a modern account of stability analysis of fluid flows. The chapter begins with a brief review of the classic definition of stability (e.g., Lyapunov, asymptotic, and exponential stability) and moves towards a modern formulation of stability as an optimization problem: unstable modes are those along which the growth of disturbances is maximized. The chapter introduces a powerful, adjoint-based, iterative method to solve such an optimization and shows how to recover common stability and receptivity results from the general framework. Finally, an illustrative application to the problem of tonal noise is given.

Part V: Applications. This part of the book is dedicated to the application of data-driven and machine learning methods to fluid mechanics.

Chapter 14, by B.R. Noack and co-workers, is dedicated to reduced-order modeling. The chapter gives an overview of the classic POD-Galerkin approach, reviewing the main challenges in closure and stabilization as well as classic applications. It then moves to emerging cluster-based Markov models and their possible generalization. A detailed tutorial is also provided to offer the reader hands-on experience with reduced-order modeling.

Chapter 15, by K. Zdybal and co-workers, focuses on the use of data-driven models for studying reacting flows. The numerical simulation of these flows is extremely challenging because of the vast range of scales involved. This chapter gives a broad overview of how machine learning techniques can help reduce the computational burden. The key challenges of high dimensionality are discussed along with an overview of dimensionality reduction methods, ranging from classic principal component analysis (PCA) to local PCA, non-negative matrix factorization (NMF), and artificial neural network (ANN) autoencoders. The application of these tools to reduce dimensionality in the modelling of transport and chemical reactions is illustrated in a challenging test case.

Chapter 16, by S. Görtz and co-workers, is dedicated to the application of reduced-order modelling for multidisciplinary design optimization in aerodynamics. The design of an aircraft involves thousands of extremely expensive numerical simulations.

This chapter shows how linear and nonlinear dimensionality reduction tools can help speed up the process. POD, cluster POD, and Isomaps, combined with nonlinear regression, are discussed and demonstrated in industrially relevant cases such as the aero/structure optimization of an entire aircraft.

Chapter 17, by B.R. Noack and co-workers, is dedicated to flow control and how machine learning might revolutionize the field. The chapter gives first an overview of flow control, its purposes, goals, tools, and strategies. Then, two paradigms for flow control are introduced and compared. On the one hand, there are model-based approaches, rooted in first principles and our ability to derive models that predict how a system responds to inputs. On the other hand, there are model-free approaches rooted in powerful optimization strategies that can "learn" the best control laws from data, by simply interacting with the system. Cluster-based control and linear genetic programming are illustrated, and the chapter closes with a tutorial on an illustrative nonlinear benchmark problem.

Chapter 18, by J. Rabault and A. Kuhnle complements the previous chapter with an overview of deep reinforcement learning (DRL) for active flow control. Reinforcement learning is one of the three paradigms of machine learning. Contrary to the other two (supervised and unsupervised learning), a reinforcement learning algorithm starts with no data and learns through experience, i.e., via trial and error. This framework is meant to tackle decision-making processes, such as teaching a computer to play chess or drive a car or, as the authors show, to control a fluid flow. This chapter introduces the main approaches of reinforcement learning (e.g. Q-learning versus policy gradient methods), the current research directions, and the recent applications to fluid mechanics problems. Guidelines to practically deploy DRL are given, along with a perspective for the future of the field.

Part VI: Perspectives. The book closes with **Chapter 19**, by J. Jiménez, with a fascinating perspective and important questions for the field. Combined with Chapter 2, this chapter explores how much the progressive synergy between machine learning and fluid dynamics, fostered by ever increasing computational capabilities, could promote the 'automation' of science and ultimately turn machines into colleagues. This, as masterfully illustrated with a simple case study, ultimately depends on whether 'blind' randomized trials can be integrated in the process of formulating hypothesis, eventually giving computers the ability to ask questions rather than just providing answers.

A note on the Notation

The reader will quickly realize that different chapters have (slightly) different notation. Among these, the same symbol is sometimes used for different purposes, and different symbols are sometimes used for the same quantities. This choice is deliberate. First, the book covers a wide range of disciplines, each with well-established notations. For example, the symbol $\boldsymbol{u}$ usually denotes the actuation in control theory and the velocity field in fluid mechanics. In reinforcement learning, the actuation is denoted by a_t and called the 'action' while the sensor measurement is denoted by s_t and called 'state' (while it is usually denoted by $\mathbf{y}$ in control theory). Resolving these ambiguities would make it difficult for readers to link the material in this book with the literature of the various intersected disciplines. Thus, each chapter represents the starting point towards more advanced and specialized literature, in which a standard notation has not yet been settled. Keeping the notation as close as possible to the cited literature helps the reader make essential connections. We hope that the reader will approach each chapter with the required flexibility, and we welcome comments, corrections and suggestions to benefit students for the next reprint.

List of Contributors

Abu-Zurayk, Mohammad
Institute of Aerodynamics and Flow Technology, German Aerospace Center (DLR), Germany

Aversano, Gianmarco
Université Libre de Bruxelles and Vrije Universiteit Brussels, Belgium
Université Paris Saclay, France

Bekemeyer, Philipp
Institute of Aerodynamics and Flow Technology, German Aerospace Center (DLR), Germany

Brunton, Steven
University of Washington, USA

Coussement, Axel
Université Libre de Bruxelles and Vrije Universiteit Brussels, Belgium
Université Paris Saclay, France

D'Alessio, Giuseppe
Université Libre de Bruxelles and Vrije Universiteit Brussels, Belgium
Université Paris Saclay, France

Dawson, Scott
Illinois Institute of Technology, USA

Discetti, Stefano
Universidad Carlos III de Madrid, Spain

Ehlert, Arthur
Technische Universität, Berlin, Germany
Industrial Analytics, Berlin, Germany

Fernex, Daniel
Technische Universität Braunschweig, Germany

Franz, Thomas
Institute of Aerodynamics and Flow Technology, German Aerospace Center (DLR), Germany

Görtz, Stefan
Institute of Aerodynamics and Flow Technology, German Aerospace Center (DLR), Germany

Ianiro, Andrea
Universidad Carlos III de Madrid, Spain

Jiménez, Javier
Universidad Politécnica Madrid, Spain

Kuhnle, Alex
University of Cambridge, United Kingdom

Lusseyran, François
Université Paris-Saclay, France

Maceda, Guy C.
Université Paris-Saclay, France

Malik, Mohammad, R.
Université Libre de Bruxelles and Vrije Universiteit Brussels, Belgium

Mendez, Miguel A.
von Karman Institute for Fluid Dynamics, Belgium

Morzynski, Marek
Poznań University of Technology, Poland

Nayeri, Christian N.
Technische Universität Berlin, Germany

Noack, Bernd R
Harbin Institute of Technology, Republic of China
Technische Universität Berlin, Germany

Parente, Alessandro
Université Libre de Bruxelles and Vrije Universiteit Brussels, Belgium

Rabault, Jean
Norwegian Meteorological Institute, Norway

University of Oslo, Norway

Ripepi, Matteo
Institute of Aerodynamics and Flow Technology, German Aerospace Center (DLR), Germany

Schmid, Peter
Imperial College London, United Kingdom

Semaan, Richard
Technische Universität Braunschweig, Germany

Sutherland, James C.
University of Utah, USA

Zdybal, Kamila
Université Libre de Bruxelles and Vrije Universiteit Brussels, Belgium