Colloquium Final Report

N. 614 – Machine learning methods for prediction and control of separated turbulent flows

Dates and location: **16/06/2021 - 18/06/2021, Paris, France**

Chairperson **Lionel Mathelin**

Co-Chairperson **Peter Schmid**

Conference fees

- Regular ONSITE fee **150.0 €**
- Regular ONLINE ATTENDANCE fee **70.0 €**

What other funding was obtained? **-**

What were the participants offered? **-**

Number of members of Euromech (reduced registration fee) **20**

Number of non-members of Euromech (full registration fee) **57**

Applicants (members)

- Miguel Beneitez
- Michele Alessandro Bucci
- Michele Buzzicotti
- Laurent Cordier
- Yohann Duguet
- Mohamed Elhawary
- Athanasios Giannenas
- George Haller
- Azeddine Kourta
- Adrian Lozano-Duran
- Francois Lusseyran
- Nicolas MAZELLIER
- Beverley McKeon
- Pierre-Yves Passaggia
- Luc Pastur
- Berengere Podvin
- Anne-Marie Schreyer
- Onofrio Semeraro
- DENIS SIPP
- Kunihiko Taira

Applicants (non members)

- Lionel Agostini
- Ateeb Ahmad
- Alexandre Allauzen
- Matthieu Ancellin
- Jane Bae
- Deniz Bezgin
- Damiano Capocci
- Johan Carlier
- Dylan Caverly
- Guy Y. Cornejo Maceda
- Francis De Voogt
- Nan DENG
- Florent Di Meglio
- Maximilian Dreisbach
- Menier Emmanuel
- Thibault Faney
- Ehsan Farzamnik
- Daniel Fernex
- Andrea Ferrero
- Lucas Fery
- Stéphane Février
- Pourya Forooghi
- Jonathan Freund
- Kai Fukami
- Thibaut Guegan
- Robin Heinonen
- Thierry HORSIN
- Xavier Jurado
- Thede Kiwitt
- Jiaqing Kou
- Tim Wilhelm Kroll
- Nishant Kumar
- Antoine Lechevallier
- Sangseung Lee
- Yiqing Li
- Qixin Lin
- Luca Magri
- Lionel Mathelin
- Rémi MOCHON
- Max Mustermann
- Frederic Nataf
- Demetrios Papageorgiou
- Romain Paris
- Thomas Philibert
- Fabio PINO
- Lorenzo Schena
- Richard Semaan
- XIE Shangyan
- Tarun Singh
- Vishal Srivastava
- Alexander Stroh
- Philibert Thomas
- Gilles Tissot
- Pedro Volpiani

- Olivier Wilk
- Jiasheng Yang
- Hadi Zolfaghari

Scientific Report

The steady rise of machine learning techniques, combined with the availability of affordable sensor arrays, has had a transformative impact in a large number of scientific fields. With the dramatic increase in data accessibility and computational power, traditional model-based approaches in engineering are giving way to a data-enhanced paradigm. Prediction and control of turbulent flows, a challenging area of engineering sciences, are no exception in this respect. Despite early attempts, the successful control of such complex systems by machine learning techniques raises specific issues such as weak observability or an exhaustive range of temporal and spatial scales. Moreover, the effective incorporation of knowledge about the physical system, such as symmetries, invariances or conservation laws, into the learning process is far from trivial.

Nonetheless, recent success of machine learning techniques in the prediction of chaotic dynamical systems and control of highly nonlinear flows has fueled a great many research efforts and has shown that progress in this field critically relies on an interdisciplinary skill set, ranging from applied mathematics and machine learning to physics, from computer science to experimental methods. The aim of the proposed workshop was to bring together control practitioners, fluid dynamicists and machine learning experts to critically review recent developments in the field and identify both opportunities and challenges in using machine learning techniques for high-dimensional physical systems. The workshop was meant to act as a forum for exchanging ideas and as an occasion to learn and discuss.

Altogether there were 77 participants from 12 countries and 34 presentations, including 6 keynotes. The list of participants and the full programme are available in a separate document. Most importantly, there was ample time for informal discussions among the participants during coffee breaks and lunches.

Specific topics addressed in the talks and discussed included:

• Data-driven modeling
Describing the behavior of a physical system, in order to predict its future state, or an associated quantity of interest, is of crucial importance in many situations. For instance, it is a building block for most applications in engineering. However, a reliable model describing the system at hand is not always available and data-based techniques have received a lot of attention recently. At the workshop, several efforts along this line have been discussed. The importance of relying on the right predictive variables, or features, to obtain an accurate, and most importantly, generalizable model was underlined and guidelines were provided. Also discussed is the role of past observations in obtaining good predictive capability models. An effort was discussed relying on the Mori-Zwanzig formalism to estimate the influence of unobserved variables onto measurement data in a principled way.
In a similar effort toward principled learning, prior expertise on the system can be included in the data-learning process, to improve the convergence rate of the learning algorithm and provide some bounds in its resulting behavior. For instance, penalization of the deviation from the solution of a Navier-Stokes solution can be considered for a model trained to reproduce flow fields. This Physics-informed machine learning approach is currently attracting a lot of efforts.

A faster and better learning can also be obtained by first providing empirical functions to train a neural network, prior to refining the training with actual data from the system. This transfer learning approach allows to save on the amount of data required for training and is an important ingredient in the common situation where data are scarce.

Also presented were criteria to assess the quality of the new data in the sense of estimating whether they were redundant with an already seen dataset or not. This is particularly subtle in multiscale systems where what may appear as extrapolation in a traditional sense may not be the case due to scale invariance with turbulence. In response, a novel data scaling method for incompressible turbulent flows through sparse regression analysis that reveals the likeliness of data to have been "seen" was discussed. In case observational data come from images, such as in oceanography for instance, a number of factors may affect the quality, such as clouds preventing a good view of the sea surface from satelites. To approximate the state of the system, generative models can be trained and learn to mimic and explore the low-dimensinal manifold the data of interest often lie on. Several efforts along these lines were presented at the workshop, for instance predicting a three-dimensional turbulent field from corrupted field measurements. While many presented approaches were involving neural networks as an approximation model, several alternative choices were also discussed. For instance, genetic programming, where a model is learned by combining elementary functions and laws, was shown to be successful in learning behavior laws such as drag correlation functions or control policies.

Further, data-modeling can be part of a simulation workflow, where a computational model is making use of a data-based module for some critical part. An example was provided in the form of a data-accelerated Poisson solver using a convolutional neural network (CNN) coupled with a Navier-Stokes simulation code.

• Dimensionality reduced models

A widespread assumption in machine learning for systems as complex as turbulence flows as considered in this workshop is that the state of the system, or the quantity of interest, is of dimension much lower than the ambient dimension of the accessible variables. This situation can be exploited and a wide range of presentations at the workshop were explicitly relying on a reduced-order representation of the configuration at hand. For instance, an unsupervised technique such as Proper Orthogonal Decomposition (POD, also termed Principal Component Analysis in other scientific communities) can be used to learn a low-dimensional basis, complemented with a data assimilation technique to estimate the associated time-dependent coefficients of the basis elements. This combination of POD and (ensemble) Kalman filter was shown to provide good performance in predicting the future state of a turbulent flow. An alternative is to use an end-to-end learning of the low-dimensional manifold via auto-encoders. Several efforts in this direction were reported.

A somehow different approach is to learn a generative model using Latent Dirichlet Allocation (LDA). This technique learns a probabilistic approximation of the training data via a random combination of learned topics, which can be formulated as spatial fields in the fluid mechanics domain. Two efforts relying on LDA were discussed, with application in climate modeling and turbulent flow field reconstruction.

The popular Dynamic Mode Decomposition method (DMD) was also discussed and this tool was used for flow analysis by several authors. This formulation was extended to observables in terms of kernels and has shown good accuracy and stability properties as illustrated with the prediction of geophysical flows.

Dimensionality reduction can also be achieved through a clustering technique, so as to coarse-grain a dataset in a few representative instances. This representation

can then be used for analyzing the flow dynamics or for control purposes.

In contrast with projection-based approaches where a high-dimensional quantity is projected onto a low-dimensional approximation space, an alternative spectral submanifold (SSM) method was presented, relying on a restriction onto smooth invariant manifolds arising from nonlinear continuation of spectral subspaces.

• Closure modeling

Of particular interest for the scientific community involved in turbulent flows is turbulence modeling. Owing to the wide range of scales involved in a turbulent flow, simulation codes cannot resolve each degree of freedom in a computationally tractable manner and one has to rely on modeling to alleviate the simulation burden. In particular, very small scales are commonly seen as rather independent from boundary conditions and geometry, and hence constitute a good candidate for a closure model with a wide range of validity. While standard closure models were derived from first principles, data-driven models now emerge as a potential alternative. Several efforts towards this direction were reported during the workshop. A popular approach is to recalibrate legacy closure models with highly resolved data from experiment or Direct Numerical Simulation. Examples were given using the popular Spalart-Allmaras RANS model, the Reynolds Stress Model or PDF model, using supervised learning, reinforcement learning or a generative approach. One can also identify different regimes of turbulent flows, and their machine-learned specific closure models, in conjunction with a classifier. It results in region-specific closure models, potentially improving upon a global model used across the entire numerical domain. Also discussed was a field inversion approach where the correction to the resolved variables is learned so as to lead to the correct prediction of turbulent quantities.

Another typical closure model used in turbulent flow simulation is near a wall where stronger gradients typically call for a high mesh resolution, hence a high numerical cost. To circumvent this challenge, wall functions can be employed to locally model the behavior of the flow without resorting to a very large number of degrees of freedom. Illustrations of this line of research were given with data-driven learning of wall functions for Large Eddy Simulation (LES) and roughness functions with deep neural networks.

• Data-driven control

Beyond the modeling and discovery of governing equations, as mentioned above, many situations of practical interest involve some action onto the system under consideration, with the goal of achieving a given objective. Prototypical of this situation is the maximization of the lift-to-drag ratio in aerodynamics via a set of actuators affecting the flow field. Typical control strategies involve solving the governing equations, and their adjoint, hence constituting a formidable challenge for large-dimensional systems to be controlled in real-time. A different approach is to learn a control policy from experiments, mapping measurements to control actions. Several talks have reported efforts in this direction, using various methods for learning and approximating the control policy. Among these, Genetic Programming has been considered to learn an algebraic expression linking the sensor information to the action via episode-based experiments for a high Reynolds number flow around an airfoil, an Ahmed body flow and an open-cavity configuration.

An alternative method relying on reinforcement learning was also employed for low dimensional flows. It is not episode-based, hence allowing for a potentially high learning rate, but still requires algorithmic developments for a wider applicability.

Number of participants from each country

Country	Participants
United Kingdom	5
France	38
United States	8
Germany	9
Italy	4
Sweden	2
Canada	1
Spain	2
Belgium	2
Switzerland	2
Denmark	1
China	3
Total	**77**

Please send this report in electronic form to the Secretary General of EUROMECH, within one month after your Colloquium.