Center for Turbulence Research
Proceedings of the Summer Program 2024

Data-Driven Methods group: overview

Understanding, predicting, and controlling turbulent flows pose formidable challenges, primarily due to their multiscale nature and intricate nonlinear interactions across scales. Even nominally simple flow problems generate a rich spectrum of spatiotemporal features, making their characterization very challenging. Direct experimentation alone, while invaluable, is not able to capture the full complexity of turbulent flows. Sensing techniques in experiments can measure only projections, cross sections, or small fractions of the spatiotemporal flow field, which inevitably causes large information gaps. While numerical modeling offers unique advantages in exploring phenomena beyond the reach of physical measurements, running high-fidelity simulations effectively and confidently is a significant challenge. Even with advanced computational methods and supercomputing, much of the spatiotemporal spectrum remains inaccessible, particularly at high Reynolds numbers or in scenarios with complex geometries and coupled physics. These limitations underscore the need for a hybrid approach that consistently and systematically unifies simulation and experimentation for industrially relevant flows.

The Data-Driven Methods group of the 2024 Summer Program has combined the assimilation of large volumes of experimental data and high-fidelity simulations with learnable signal transformations leveraging symmetries and ad hoc numerical modeling. This approach has facilitated the identification of key physical mechanisms and has enabled the creation of accurate and efficient surrogate models. These methods are categorized into two subgroups: linear methods and nonlinear methods. The linear methods subgroup used resolvent analysis to gain physical understanding from limited flow field measurements, extracted linear operators that can best represent turbulent data, and proposed improvements to conventional techniques that drastically improve computation time. Linear methods lay the foundation for interpreting turbulent flows, offering efficiency and a well-understood reasoning framework. However, the inherent complexity of turbulent systems often necessitates nonlinear approaches, which excel at capturing multiscale interactions and intricate physical dynamics.

In classical linear approaches, the Navier-Stokes equations are linearized around a base state, such as the mean state for turbulent flows. Resolvent analysis identifies the most amplified linear modes, which, due to the nonnormal nature of the linear operator in wall-bounded flows, can coincide with energetic production mechanisms. Soria *et al.* apply this technique to an adverse pressure gradient turbulent boundary layer (APGTBL) on the verge of separation to demonstrate that both the most energetic structures and linearly amplified structures are found in the wake. Preskett *et al.* (Fundamentals group) apply this technique to experimental measurements of APGTBLs and find that the most energetic frequencies in the log layer coincide with the most amplified frequencies. This technique demonstrates great promise for APGTBLs through limited measurements of the mean flow field, which can be augmented with additional data or closure models.

Although linear approaches enable efficient predictive capabilities, a major limitation of their applicability is their difficulty dealing with complex geometries. In order to circumvent this challenge, Herrmann *et al.* use dynamic mode decomposition in concert with *a priori* knowledge of the nonlinearities to construct a data-driven resolvent operator from a reduced basis of modes. Theoretical challenges of linear approaches for turbulent flows is the choice of base state and whether the linearized Navier-Stokes equations can capture turbulent phenomena. Khoo *et al.* address these challenges by identifying data-driven linear operators through two different Ansätze using an intrusive and nonintrusive scheme. They identify an equation-free linear operator that can incorporate turbulent fluctuations.

Regardless of how the linear operators are constructed, analyses often rely on Krylov subspaces for linear stability theory, resolvent analysis, and other matrix operations. Furthermore, many classical methods in modal analysis like the proper orthogonal decomposition (POD) maximize L_2 norms and rely on Krylov methods via the singular value decomposition. Schmid *et al.* develop an algorithm that greatly accelerates these analyses with minimal, user-controlled distortion. Adjoint methods are used by Müller *et al.* to accelerate Bayesian optimization through adjoint-based gradient information and by Kontogiannis *et al.* to carry out Bayesian inference for model discovery.

Nonlinear approaches can be leveraged to discover physical relationships from the data through advanced scale-space decompositions. For example, Macedo *et al.* investigate the use of autoencoders to capture and retrieve flow field variations in response to geometric changes in internal flows, seamlessly transitioning from circular to square pipe configurations. The latent space of these simulations was found to be a linear function of the cross-sectional area and perimeter. Autoencoders are also studied by Kelshaw and Magri to identify a latent space for a general dynamical system. From this latent space, POD is generalized to manifolds to further identify key dynamics within the latent space.

The use of machine learning techniques enables the development of efficient, physics-informed surrogate models. Wang *et al.* assimilate experimental data through Fourier neural operators with loss functions that incorporate physical knowledge to accurately forecast wind-driven runback water films. In addition to translation equivariance—inherently present in convolutional neural networks—Bezgin *et al.* attempt to incorporate regular representations of additional symmetry groups, including discrete pointwise rotations and mirroring, to close the Reynolds-averaged Navier-Stokes equations. The development of these efficient reduced-order models could enable forecasting within practically actionable time frames. Nóvoa and Magri studied this issue via the noisy data of a hydrogen-based annular combustor and a turbulent wake by utilizing the regularized bias-aware ensemble Kalman filter to combine reduced-order models with flow measurement.

The Data-Driven Methods group's advances in linear methods allow physical inference from limited data, challenge conventional linearizations for turbulence, and accelerate existing methods. In a similar vein, advances using techniques from machine learning identify reductions in the turbulent flow data that maintain the essential flow dynamics and create efficient surrogate models that faithfully model the dominant traits of turbulent data. These developments, combined with technological progress in sensing and actuation, may lay the foundations for the next-generation design of near-real-time digital twins.

S. R. Gomez & D. Rossinelli