【Grant-in-Aid for Transformative Research Areas（B）】

Theoretical and Methodological Foundation for Data-Driven Inference with Application to High-Dimensional Measurement

Principal Investigator	The University of Tokyo, Research Center for Advanced Science and Technology, Lecturer		
	TAKEISHI Naoya	Researcher Number : 20824030	
Project Information	Project Number : 25B402	Project Period (FY) : 2025-2027	
	Keywords : inverse problems, simulation, partial differential equations, machine learning, fluid flow measurement		

Purpose and Background of the Research

●Outline of the Research

Simulation is a fundamental tool for understanding physical phenomena through mathematical models. It is typically based on differential equations and solving them numerically may require enormous computing resources. Recently, machine-learning techniques including deep learning have been used to directly approximate solutions of differential equations or to learn surrogate models, for running simulations (or their approximations) faster, with growing applications across disciplines.

By comparing simulation with real observations, we can understand, predict, and control phenomena. To this end, we need to estimate the parameters of a simulator. Such inverse problems conventionally require optimization with many simulation runs. Data-driven inference, to learn solutions directly from simulation outcomes, may also be useful for such inverse problems, but applying it to high-dimensional problems is still challenging, and and solid theoretical guarantees remain unclear.

This Transformative Research Area studies algorithms and theories of solving data-driven inference. We will evaluate the validity of machine-learning-based inverse problem solvers. We will apply such solvers to high-dimensional measurements of flow fields. Experts in PDEs, machine learning, and fluid measurement will build a collaborative framework linking theory, methodology, and application (Fig. 1).

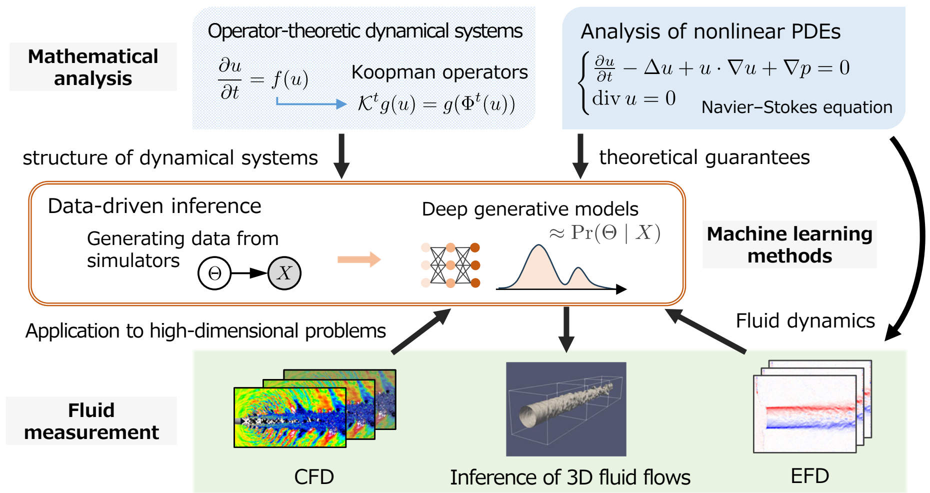

Figure 1. Structure of the Research Area: Theory, algorithms, and application of data-driven inference

●Data-Driven Inference

Data-driven inference typically proceeds as follows (Fig. 2):

① Run simulations with various parameters and store parameter–outcome pairs as data,
② Learn a model predicting the distribution of parameters given simulator's outcome, and
③ Infer the parameters corresponding to observations using the learned model.

When the parameters are high-dimensional, the data to be generated may become huge, and overly complex models may be needed, making computation impractical. We aim to streamline data generation and model architectures based on mathematical analysis of PDEs and the structure of the dynamics.

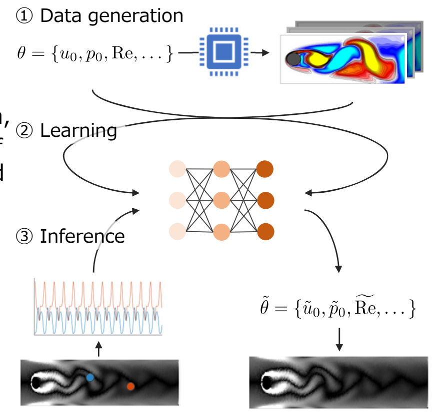

Figure 2. Procedures of data-driven inference

Expected Research Achievements

●Mathematical Analysis of PDEs

Analyzing nonlinear PDEs such as the Navier–Stokes equations, we will study the theoretical foundation of data-driven inference, particularly using operator-theoretic dynamical systems such as the Koopman operators. In operator-theoretic analysis, useful facts such as continuity and spectral properties are not yet guaranteed for complex nonlinear systems or infinite-dimensional domains. We will address them with functional analysis, complex analysis, and harmonic analysis. We will also build global well-posedness theory for equations of complex fluid phenomena.

●Development of Machine Learning Methods

Current methods of data-driven inference often rely on general deep learning generative models, which may need large data and are hard to apply directly to high-dimensional inference. We will study ways to embed the specific problem structure into machine learning models and the data-generation process; for example, low-dimensional manifolds, stability, or periodicity of dynamics.

●Application to Fluid Measurement

We will apply data-driven inference to fluid measurement by studying various configurations such as observation types, coverage, noise, data size, and grid resolution. Data-driven inference will be useful to investigate the dynamics of intermittent events behind turbulent jet acoustics (Fig. 3).

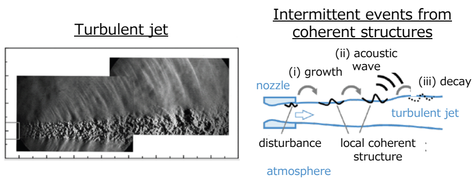

Figure 3. Fluid flows to be investigated

●Towards More Diverse Applications

High-dimensional data-driven inference can be used not only in fluid measurement but also in inference of multi-agent systems, seismic-wave tomography of subsurface structures, among many others. This Transformative Research Area can be a first step toward establishing such a large-scale interdisciplinary research collaboration.

Homepage Address, etc.	https://ailab.t.u-tokyo.ac.jp/datainfer