Beyond Closure Models: Learning Chaotic Systems via Physics-Informed Neural Operators

Chuwei Wang
Caltech

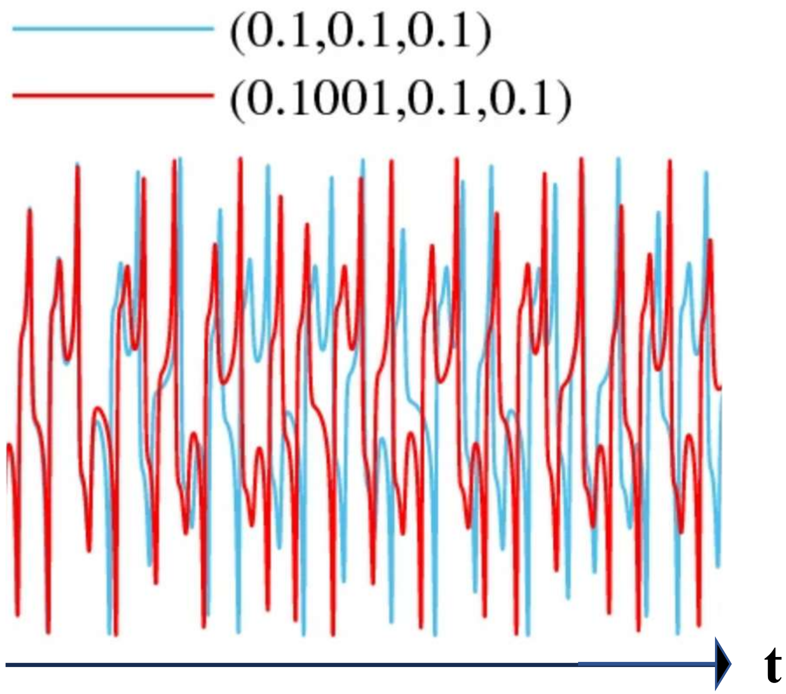
Joint work with Julius Berner, Zongyi Li, Di Zhou, Jiayun (Peter) Wang, Jane Bae, and Anima Anandkumar

2025.2.20

Chaotic Systems

Does the flap of a butterfly's wings in Brazil set off a tornado in Texas?

Small (numerical) error explodes along time!



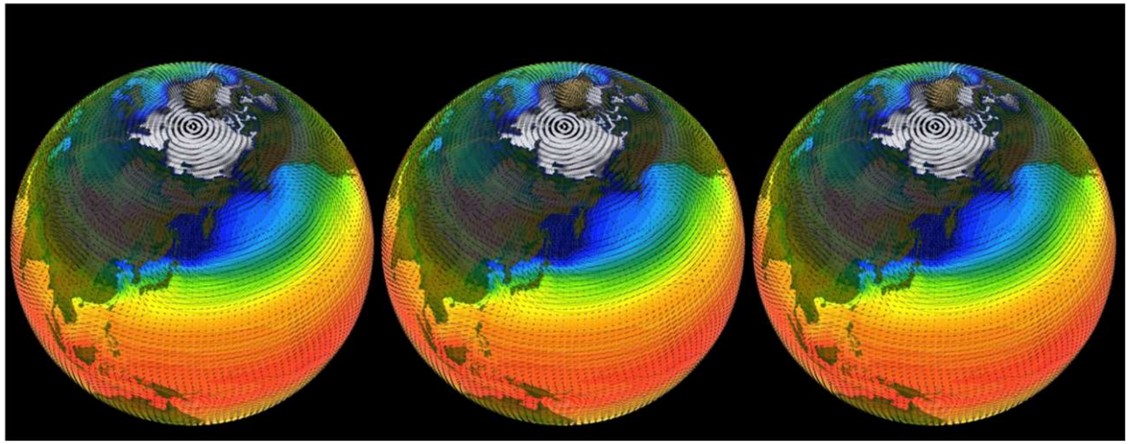
Edward Lorenz

Chaotic Systems

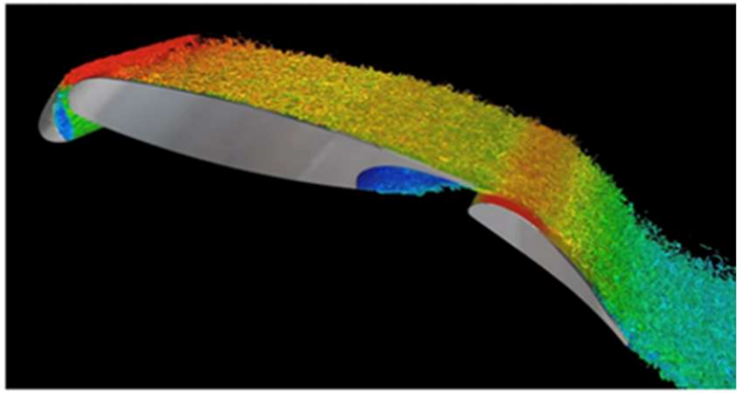
Does the flap of a butterfly's wings in Brazil set off a tornado in Texas?

Small (numerical) error explodes along time!

- Many physics systems are chaotic.
- Long-term behavior/ statistics is of great practical importance in applications.



Climate Modeling



Aircraft Design

Chaotic Systems

- Long-term behavior/ statistics is of great practical importance in applications.

➢ Method 1: Fully-Resolved Simulations (FRS): simulate with sufficiently fine spatiotemporal meshes/grids. **Too Expensive!**

➢ Method 2: Coarse-grid simulations (CGS) **(?)**

Need special modifications of the dynamic

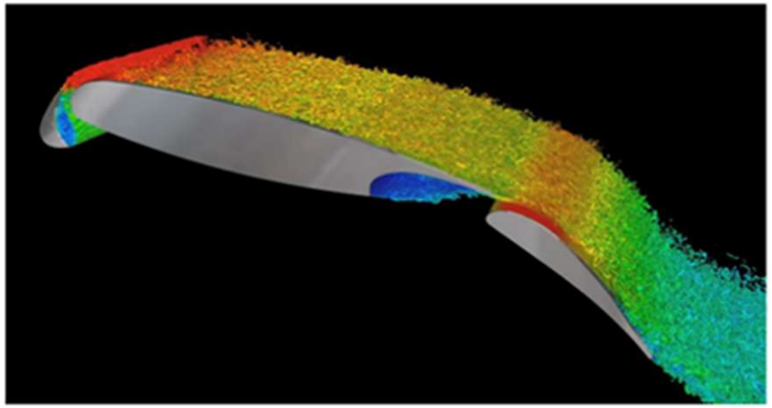
Coarse-graining, Reduced-order modeling,

Closure Modeling, etc.

(Terminologies from different areas)

Grid-point requirements (*Choi & Moin, PoF, 2012*)

DNS ~ $Re^{2.64}$ LES ~ $Re^{1.86}$ RANS ~ Re^{1}



(*ONERA*)

Aircraft wing section with $Re_c \sim 10^7$

Total number of grid points
- DNS ~ 10^{19} — Fully-Resolved
- LES ~ 10^{13} — Coarse-grained Dynamics
- RANS ~ 10^{7} — Coarse-grained Dynamics

[Remark]

- Long-term trajectory: meaningless and impossible.
- Long-term statistics: meaningful, possible.

How to give good estimations of long-term stats with CGS (limited computing resources)?

Outline

- (1) Problem formulation: long-term statistics & coarse-grid simulations.
- (2) Limitation of Closure Models: Non-uniqueness issue
- (3) Theoretical Perspective via Measure Flow:
  - Learning-based closures: impractical reliance on hi-fidelity data.
- (4) Coarse-graining with Neural Operator
- (5) Conclusion & Future Direction & Discussion

Main Results

- (1) Fundamental Shortcoming of Closure Modeling Scheme (explicit additive closures): The target mapping is **non-unique** (multi-map).

  Incorporating memory and randomness can not resolve the issue.
- (2) The amount of hi-fid data required for training a closure model suffices to estimate statistics well. **No more need to train a closure model!**
- (3) We need nonlinear interactions between different scales.

  Neural operators provide a solution. (Learn the solution operator)

  Relatively accurate per-step prediction is enough for long-term statistics.

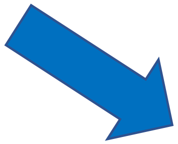
Outline

- (1) Problem formulation: long-term statistics & coarse-grid simulations.
- (2) Limitation of Closure Models: Non-uniqueness issue
- (3) Theoretical Perspective via Measure Flow:
  - Learning-based closures: impractical reliance on hi-fidelity data.
- (4) Coarse-graining with Neural Operator
- (5) Conclusion & Future Direction & Discussion

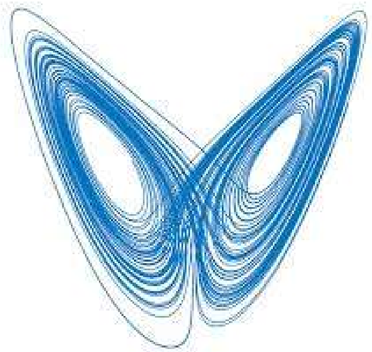
Problem Setting

[Dynamics, Attractor, and Statistics]

- A nonlinear (and chaotic) dynamics $\partial_t u = Au, u \in H$
- S_t: the semigroup (of the true dynamics) $u(\cdot, 0) \to u(\cdot, t)$
- 'Trajectory': $\{S_t u\}_{t \in R_{\geq 0}}$
- Attractor: $\Omega \subset H \; s.t. \lim_{t\to\infty} dist(S(t)u, \Omega) = 0 \;, \forall u \in \mathrm{H}.$
- Invariant measure $\mu^* \coloneqq \lim_{T\to\infty} \frac{1}{T} \int_0^T \delta_{S(t)u} \, dt$, (independent of the initial u)
  - Measure in function space
  - Supported on Ω
- **Statistics**: For a functional O, the stat $\langle O \rangle \coloneqq E_{u \sim \mu^*} O(u) = \int O(u) \mu^*(du)$



The Goal.



Lorenz Attractor

Coarse-grid Simulations

- Filtering operator $F: u \to \bar{u}$, e.g. spatial down sampling, Fourier-mode truncation.
  (Simulations with coarse grids)

Nonlinear dynamics: $\partial_t u = Au, u \in H$ (H: function space of interest)

➢ **[Coarse-graining]** (CG): $\partial_t v = \tilde{A}\, v, v \in F(H)$.

Target: design a vector field (operator) in the reduced space.

Coarse-grid Simulations

- Filtering operator $F: u \to \bar{u}$, e.g. spatial down sampling, Fourier-mode truncation.
  (Simulations with coarse grids)

Nonlinear dynamics: $\partial_t u = Au, u \in H$ (H: function space of interest)

➢ **[Coarse-graining]** (CG): $\partial_t v = \tilde{A}\, v,\, v \in F(H)$.

➢ Dynamics of $\bar{u}$: $\partial_t \bar{u} = A\bar{u} + (FA - AF)\boldsymbol{u}$.

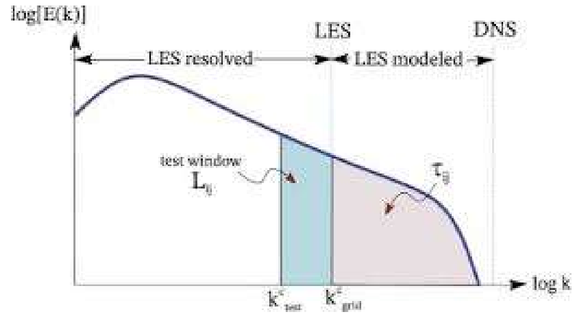
(nonlinear system $\mathcal{A}$ and $\mathcal{F}$ does not commute)

➢ Simulating on low-res grids (the space of $F(H)$) no access to $\boldsymbol{u}$.

➢ [**Closure Model**]: Evolve $\partial_t \bar{u} = A\bar{u} + clos(\bar{u}; \theta)$

CG Ansatz: $\tilde{A} = A + clos(\cdot, \theta)$

How to Design Closure Models?



Dynamic Smagorinsky Model

[(Handcraft) Classical Methods]

- Strong Physics Intuition.
- Mathematical Simplifications.
  - E.g. Assume clear scale-separations $u(x) = \overline{u}(x) + \epsilon\, w(x)$
  - Ideal/Limit cases: Kolmogorov microscales. ($+\infty$ Reynolds number, isotropic homogeneous).

[Data-driven Closure Models]

- Single-state Models
- History-aware Models
- Stochastic Closure Models

- Different Loss function, model architecture, ansatz …

Existing Data-driven Closure Models

- [CGS]: Evolve $\partial_t u = Au + clos(u;\theta)$
- Dynamics of $\bar{u}$: $\partial_t \bar{u} = A\bar{u} + (FA - AF)\boldsymbol{u}$.

Training data $\boldsymbol{u}_i$: comes from costly Fully-Resolved simulations ➡ input-output pairs

Supervised Learning

A priori Loss function $$J_{ap}(\theta;\mathfrak{D}) = \frac{1}{|\mathfrak{D}|}\sum_{i\in\mathfrak{D}} \|clos(\overline{u}_i;\theta) - (\mathcal{F}\mathcal{A} - \mathcal{A}\mathcal{F})u_i\|^2$$

Posteriori Loss function $$J_{post}(\theta;\mathfrak{D}) = J_{ap}(\theta) + \frac{1}{|\mathfrak{D}|}\sum_{i\in\mathfrak{D}} \|v_i(\cdot,\Delta t;\theta) - \mathcal{F}(S(\Delta t)u_i)\|^2$$

Single-state model: $\boldsymbol{clos}: H(R^d) \to H(R^d)$

History-aware models: $\boldsymbol{clos}: H(R^3 \times [0,t')) \to H(R^3)$

$$u(x,t), x \in D, t \in [t_0 - t', t_0) \to clos(u)(x,t_0)$$

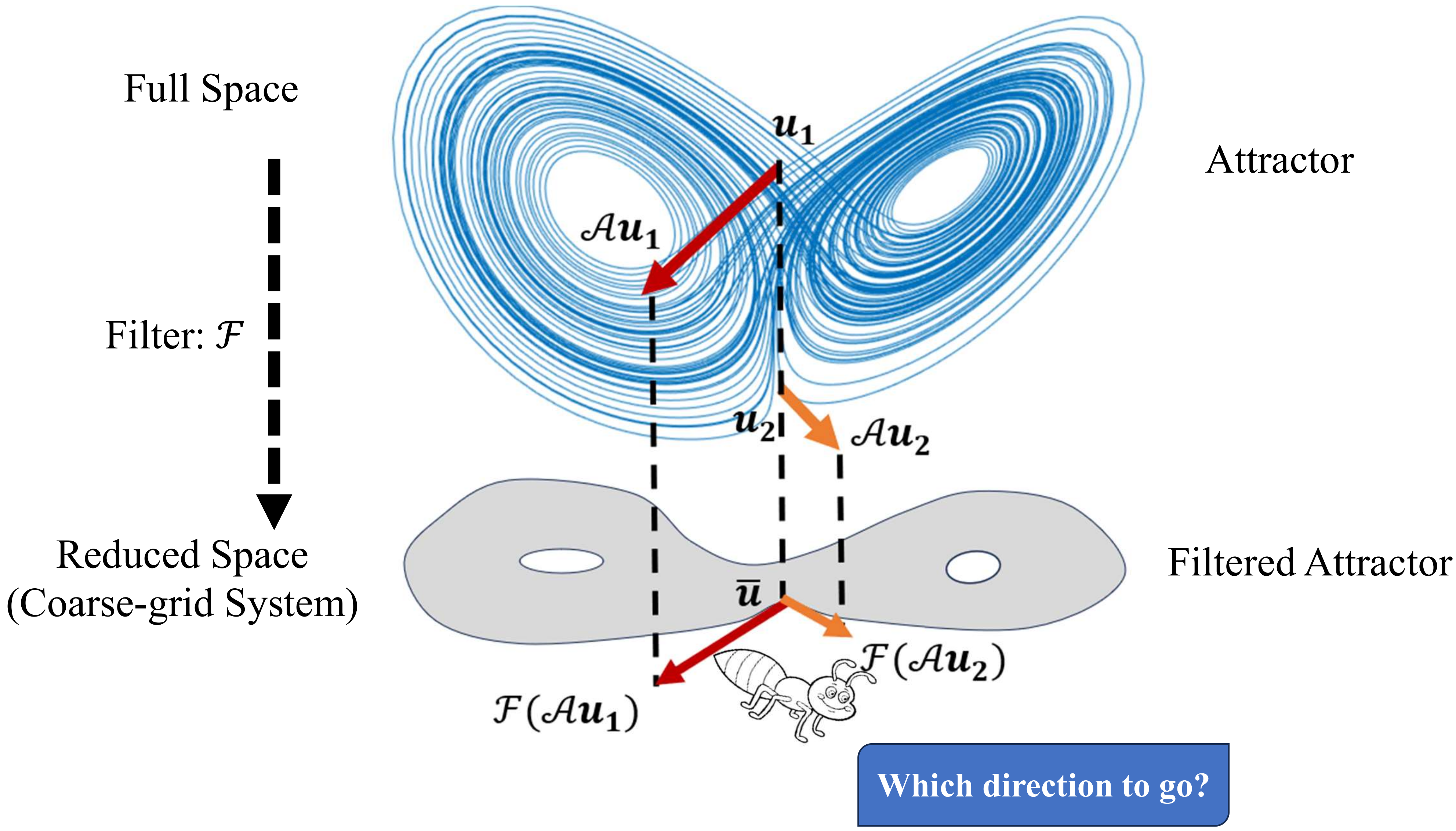
Stochastic closure models

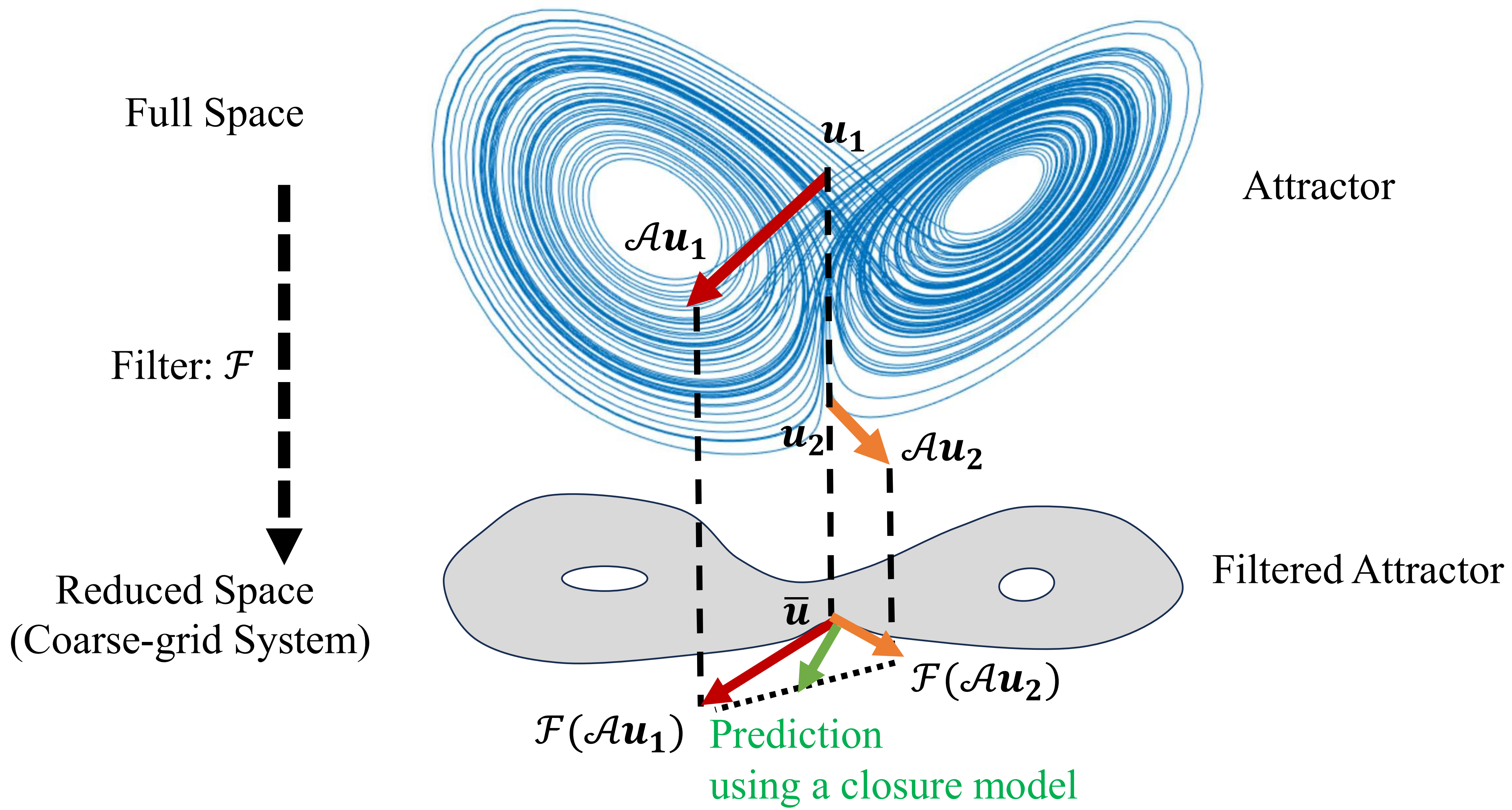
Outline

- (1) Problem formulation: long-term statistics & coarse-grid simulations.
- (2) Limitation of Closure Models: Non-uniqueness issue
- (3) Theoretical Perspective via Measure Flow:
  - Learning-based closures: impractical reliance on hi-fidelity data.
- (4) Coarse-graining with Neural Operator
- (5) Conclusion & Future Direction & Discussion

➢[**Closure Model**]: Evolve $\partial_t \bar{u} = A\bar{u} + clos(\bar{u};\theta)$

- [**Thm 1-1 (Single-state)**] The target mapping $clos(u)$ **is not well-defined** for all types of ansatz in the literature. (There are multiple possible outputs for the same input.)
- The approximation error has a lower bound independent of the model complexity.


Full Space
Filter: $\mathcal{F}$
Reduced Space
(Coarse-grid System)
u_1
$\mathcal{A}u_1$
u_2
$\mathcal{A}u_2$
Attractor
Filtered Attractor
$\bar{u}$
$\mathcal{F}(\mathcal{A}u_1)$
$\mathcal{F}(\mathcal{A}u_2)$
Which direction to go?

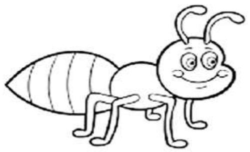


Full Space
Filter: $\mathcal{F}$
Reduced Space
(Coarse-grid System)
Attractor
$\boldsymbol{u}_1$
$\mathcal{A}\boldsymbol{u}_1$
$\boldsymbol{u}_2$
$\mathcal{A}\boldsymbol{u}_2$
Filtered Attractor
$\overline{\boldsymbol{u}}$
$\mathcal{F}(\mathcal{A}\boldsymbol{u}_2)$
$\mathcal{F}(\mathcal{A}\boldsymbol{u}_1)$
Prediction
using a closure model

➢[**Closure Model**]: Evolve $\partial_t \bar{u} = A\bar{u} + clos(\bar{u};\theta)$

- [**Thm 1-2 (History-aware)**] (For PDE systems) For any $\boldsymbol{u}$ and finite τ, there exist infinite u′ ∈ H such that $FS(t)u' = FS(t)u$ for all t ∈ [0,τ).

$\{\boldsymbol{F}(\boldsymbol{S}_t\boldsymbol{u})\}_{t<t'}$ are all the information I could use to decide which direction to move at moment $\boldsymbol{t'}$!



➢[**Closure Model**]: Evolve $\partial_t \bar{u} = A\bar{u} + clos(\bar{u};\theta))$

- [**Thm 1-3 (Stochastic closures)**] One cannot obtain the best approximation of $\mu*$ among distributions supported in the reduced space if there is additive randomness in the evolution of dynamics.

Parameters governing stochasticity tend to diminish after training.

Incorporating randomness in the closure model might be redundant.

Outline

- (1) Problem formulation: long-term statistics & coarse-grid simulations.
- (2) Limitation of Closure Models: Non-uniqueness issue
- (3) Theoretical Perspective via Measure Flow:
  - Learning-based closures: impractical reliance on hi-fidelity data.
- (4) Coarse-graining with Neural Operator
- (5) Conclusion & Future Direction & Discussion

Review: Goal = long-term stats

- Long-term trajectory: meaningless and impossible.
- Long-term statistics: meaningful, possible.

How to give good estimations of long-term stats with CGS (limited computing resources)?

[Rethinking]

- What is the best we can achieve with coarse-grid simul.?
- How to achieve that?
- Empirical results for existing methods still look good?
- CG Ansatz in closure modeling: $\tilde{A} = A + clos(\cdot, \theta)$. Alternative ansatz?

Review: Goal = long-term stats

- Long-term trajectory: meaningless [illegible] impossible.
- Long-term statistics: meaningful, p[illegible]ble.

How to give good estimations of long[illegible]m stats with CGS (limited computing resources)?

Check directly how the distribution /measure evolves.

Liouville equation / Fokker-Planck eqn. for measures in function spaces.

Check the stationary (Liouville / F-P) equation for the limit distribution!

Analyze the convergence of different CG ansatz.

Formulations

- Dynamics $\partial_t u = Au, u \in H$
- Coarse-graining: Filtering operator F. (linear, finite-rank)
- Reduced space: $F(H)$: finite-dim linear subspace of H.

- $H = F(H) \bigoplus F(H)^{\perp}$: $\boldsymbol{u} = v + w$
- ONB $\{\psi_i\}$: $F(H) = span\{\psi_i : i \leq d\}$.
- Canonical isometric isomorphism T: $\boldsymbol{u} \leftrightarrow c \in R^{\infty} \cap \ell^2$: $\boldsymbol{u} = \sum_i c_i \psi_i$.
- Denote $\hat{v} \coloneqq (c_1, c_2, \ldots c_d), \hat{w} \coloneqq (c_{d+1}, \ldots.)$.
- The nonlinear dynamics becomes: $\frac{dc}{dt} = f(c), f$ is an inf-dim vector field (function).
- **Example**: ψ_i: Fourier basis (or sin & cos to avoid complex number)
- Kuramoto–Sivashinsky : $\partial_t u + uDu + D^2 u + D^4 u = 0,\ D \coloneqq \partial_x$,
- . $f(c)_k = (-k^4 + k^2) c_k - ik \sum_{j+l=k} c_j c_l$.

Formulations

- $H = F(H) \bigoplus F(H)^{\perp}$: $\boldsymbol{u} = v + w$
- ONB $\{\psi_i\}$: $F(H) = span\{\psi_i : i \leq d\}$.
- Canonical isometric isomorphism T: $\boldsymbol{u} \leftrightarrow c \in R^{\infty} \cap \ell^2$: $\boldsymbol{u} = \sum_i c_i \psi_i$.
- Denote v $\coloneqq (c_1, c_2, \dots c_d), w \coloneqq (c_{d+1}, \dots .)$. (View v, $\hat{v}$ as the same)
- The nonlinear dynamics becomes: $\frac{dc}{dt} = f(c), f$ is an inf-dim vector field (function).
- $\begin{cases} \frac{dv}{dt} = f_r(v, w) \\ \frac{dw}{dt} = f_u(v, w) \end{cases}$, $f_r = (f_1, \dots f_d), f_u = (f_{d+1}, \dots)$
- Evolution: $\partial_t \rho(c, t) = -\nabla_c \cdot \big(\boldsymbol{f}(c)\rho(c, t)\big)$.
- Invariant measure μ^* (in H): $-\nabla_c \cdot \big(\boldsymbol{f}(c)\rho^*(c, t)\big)$=0

[Rethinking]

- What is the best we can achieve with coarse-grid simul.?

Proposition B.5. $\rho_1^* = \underset{\mu \in \mathscr{P}(\mathcal{F}(\mathcal{H}))}{\arg\min} \mathcal{W}_{\mathcal{H}}(\mu, \mu^*)$.

- $\rho_1(v)$: marginal distribution of $\rho(v, w)$.
- Equivalent to $P_{\#}\mu^*$, P: orthonormal projection onto $F(H)$. (Typically $P = F$).
- **[Optimal CG dynamics]**
- $\begin{cases} \frac{dv}{dt} = f_r(v, w) \\ \frac{dw}{dt} = f_u(v, w) \end{cases}, \partial_t \rho(c, t) = -\nabla_c \cdot \big(\boldsymbol{f}(c)\rho(c, t)\big)$,

  check the evolution of $\rho_1(v, t) = \int \rho(v, w, t) dw$.
- Evolution: $\partial_t \rho_1(v, t) = -\nabla_{\mathbf{v}} \cdot \left(\rho_1(\mathbf{v}, t) \mathbb{E}_{\mathbf{w} \sim \rho(\mathbf{w}|\mathbf{v};t)}[f_r(\mathbf{v}, \mathbf{w})|\mathbf{v}]\right)$.
- CG dynamics: $\frac{d\mathbf{v}}{dt} = \mathbb{E}_{\mathbf{w} \sim \rho(\mathbf{w}|\mathbf{v};t)}[f_r(\mathbf{v}, \mathbf{w})|\mathbf{v}]$

J Langford and R Moser. "Optimal LES formulations for isotropic turbulenc'', 1999.

- **[Optimal CG dynamics]**
- $\begin{cases} \frac{dv}{dt} = f_r(v,w) \\ \frac{dw}{dt} = f_u(v,w) \end{cases}, \partial_t \rho(c,t) = -\nabla_c \cdot \big(\boldsymbol{f}(c)\rho(c,t)\big),$

  check the evolution of $\rho_1(v,t) = \int \rho(v,w,t)dw$.
- Evolution: $\partial_t \rho_1(v,t) = - \nabla_{\mathbf{v}} \cdot \big(\rho_1(\mathbf{v},t)\mathbb{E}_{\mathbf{w}\sim\rho(\mathbf{w}|\mathbf{v};t)}[f_r(\mathbf{v},\mathbf{w})|\mathbf{v}]\big).$
- CG dynamics: $\frac{d\mathbf{v}}{dt} = \mathbb{E}_{\mathbf{w}\sim\rho(\mathbf{w}|\mathbf{v};t)}[f_r(\mathbf{v},\mathbf{w})|\mathbf{v}]$

Unfortunately, not useful in practice: $\rho(w|v;t)$ depends on simul in H.
The only information can be used at state $v \in F(H)$, time t f
$\{v(x,\tau)\}_{\tau\le t}$ different from $\{\overline{u}(x,\tau)\}_{\tau\le t}$!

CG dynamics should have the form $\frac{dv}{dt} = f_*(v)$ (or $f_*(v,t)$).

Effective Optimal CG Dynamics

$$\frac{d\mathbf{v}}{dt} = \mathbb{E}_{\mathbf{w}\sim\rho^*(\mathbf{w}|\mathbf{v})}[f_r(\mathbf{v},\mathbf{w})|\mathbf{v}]$$

$$\partial_t v = \mathbb{E}_{u\sim\mu_*}[\mathcal{F}\mathcal{A}u|\mathcal{F}u = v]$$

$$clos(v) = \mathbb{E}_{u\sim\mu^*}[\mathcal{F}\mathcal{A}u|\mathcal{F}u = v] - \mathcal{A}v, \quad v \in \mathcal{F}(\mathcal{H})$$

What are date-driven closure models learning?

$$J_{ap}(\theta) = \mathbb{E}_{u\sim p_{data}}\|\mathcal{A}_\theta\mathcal{F}u - \mathcal{F}\mathcal{A}u\|^2$$
$$= \mathbb{E}_{(\mathbf{v},\mathbf{w})\sim p_{data}(\mathbf{v},\mathbf{w})}|f_r(\mathbf{v};\theta) - f_r(\mathbf{v},\mathbf{w})|^2$$

Resulting mapping: $\mathrm{clos(v)} = \mathbb{E}_{u\sim\hat{\boldsymbol{\mu}}_{\boldsymbol{data}}}[\mathcal{F}\mathcal{A}u|\mathcal{F}u = v] - \mathcal{A}v$.

Empirical measure of the hi-fid datapoints (functions/snapshots)

Theorem 2.2. *(Impractical Reliance on High-fidelity Data for Learning-based Closures)* *Let* $clos^*$ *be the optimal closure model that provides the best approximation of* μ^* *among all closure models. (i)* $clos^*$ *is dependent on* μ^**. (ii)Let* N_{stat} *be the number of fully-resolved data points (snapshots) required to achieve a sufficiently accurate estimation of long-term statistics, and* N_{clos} *the number required for a learning-based closure model to approximate* $clos^*$ *to a comparable accuracy. Then,* $N_{stat} = o(N_{clos})$.

➢Paradox: No need to train a closure model – the training data itself is enough!

➢Cannot generalize. Different domain, boundary condition, coefficient (e.g. Reynolds number).

[Rethinking]

- Empirical results for existing methods still look good?

Method	Optimal statistics	High-res. training data Snapshots	High-res. training data Trajs.	Complexity
Fully-resolved Simulation, e.g., DNS [35, 36]	✓	-		$Re^{3.52}$
Coarse-grid Simulation, e.g., LES [35, 36]	✗	-		$Re^{2.48}$
Single-state model [28]	✗	24000	8	$Re^{2.48}$
History-aware model [37]	✗	250000	50	$Re^{2.48}$
Latent Neural Stochastic Differential Equation (SDE) [34]	✗	179200	28	$\frac{1}{\delta t} Re^{1.86}$
Online Learning [38]	✗	-		$Re^{3.52}$
Physics-Informed Operator Learning (Ours)	✓	384	1	$Re^{1.86}$

Outline

- (1) Problem formulation: long-term statistics & coarse-grid simulations.
- (2) Limitation of Closure Models: Non-uniqueness issue
- (3) Theoretical Perspective via Measure Flow:
  - Learning-based closures: impractical reliance on hi-fidelity data.
- (4) Coarse-graining with Neural Operator
- (5) Conclusion & Future Direction & Discussion

What is going wrong with closure model?

Nonlinear dynamics: $\partial_t u = Au, u \in H$ (H: function space of interest)

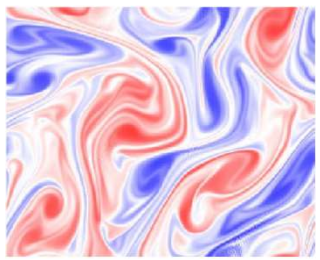
➤**[Coarse-graining]** (CG): $\partial_t v = \tilde{A}\, v, v \in F(H)$.

➤[**Closure Model**]: Evolve $\partial_t \bar{u} = A\bar{u} + clos(\bar{u}; \theta)$

CG Ansatz: $\boldsymbol{\tilde{A} = A + clos(\cdot, \theta)}$

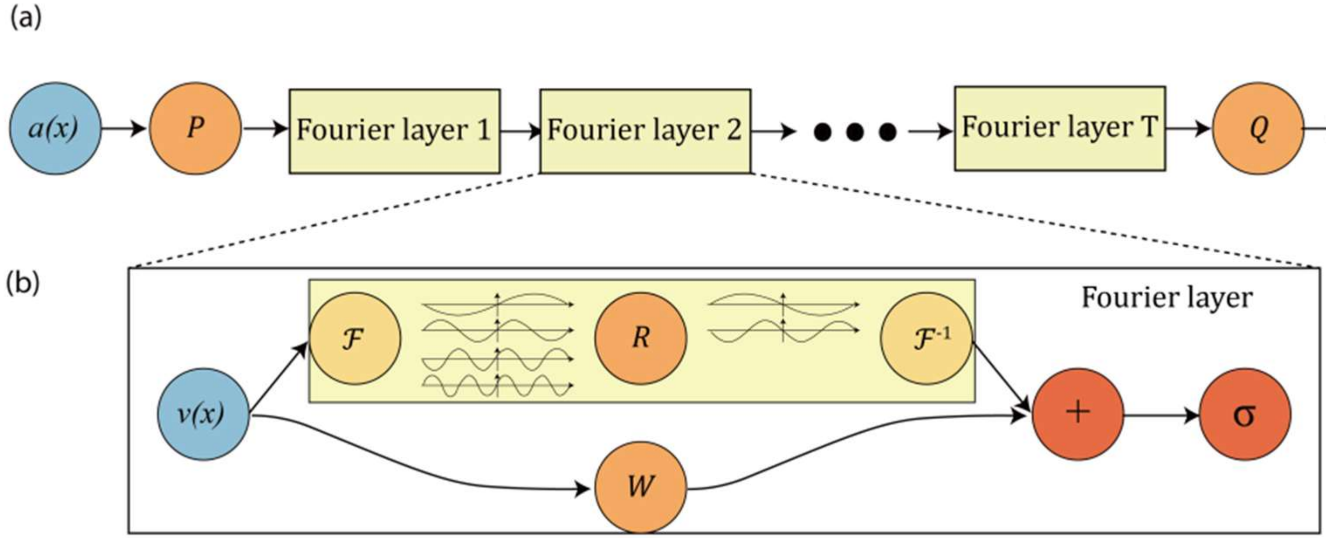
Separate roles: large-scale info/evolution + fine-scale info

- Clear scale-separations:
- Information from different scales are highly-entangled:



- **We need nonlinear interactions between scales!**

(Fourier) Neural Operator



Li et al. Fourier Neural Operator for Parametric Partial Differential Equations, 2021

(a) The full architecture of neural operator: start from input a. 1. Lift to a higher dimension channel space by a neural network P. 2. Apply four layers of integral operators and activation functions. 3. Project back to the target dimension by a neural network Q. Output u. **(b) Fourier layers**: Start from input v. On top: apply the Fourier transform $\mathcal{F}$; a linear transform R on the lower Fourier modes and filters out the higher modes; then apply the inverse Fourier transform $\mathcal{F}^{-1}$. On the bottom: apply a local linear transform W.

Figure 2: **top:** The architecture of the neural operators; **bottom:** Fourier layer.

$$\mathcal{G}_{FNO} := \mathcal{Q} \circ (W_L + \mathcal{K}_L) \circ \cdots \circ \sigma(W_1 + \mathcal{K}_1) \circ \mathcal{P},$$

The input can be either course-grid or fine-grid.
All viewed as discretization of an inf-resolution function.

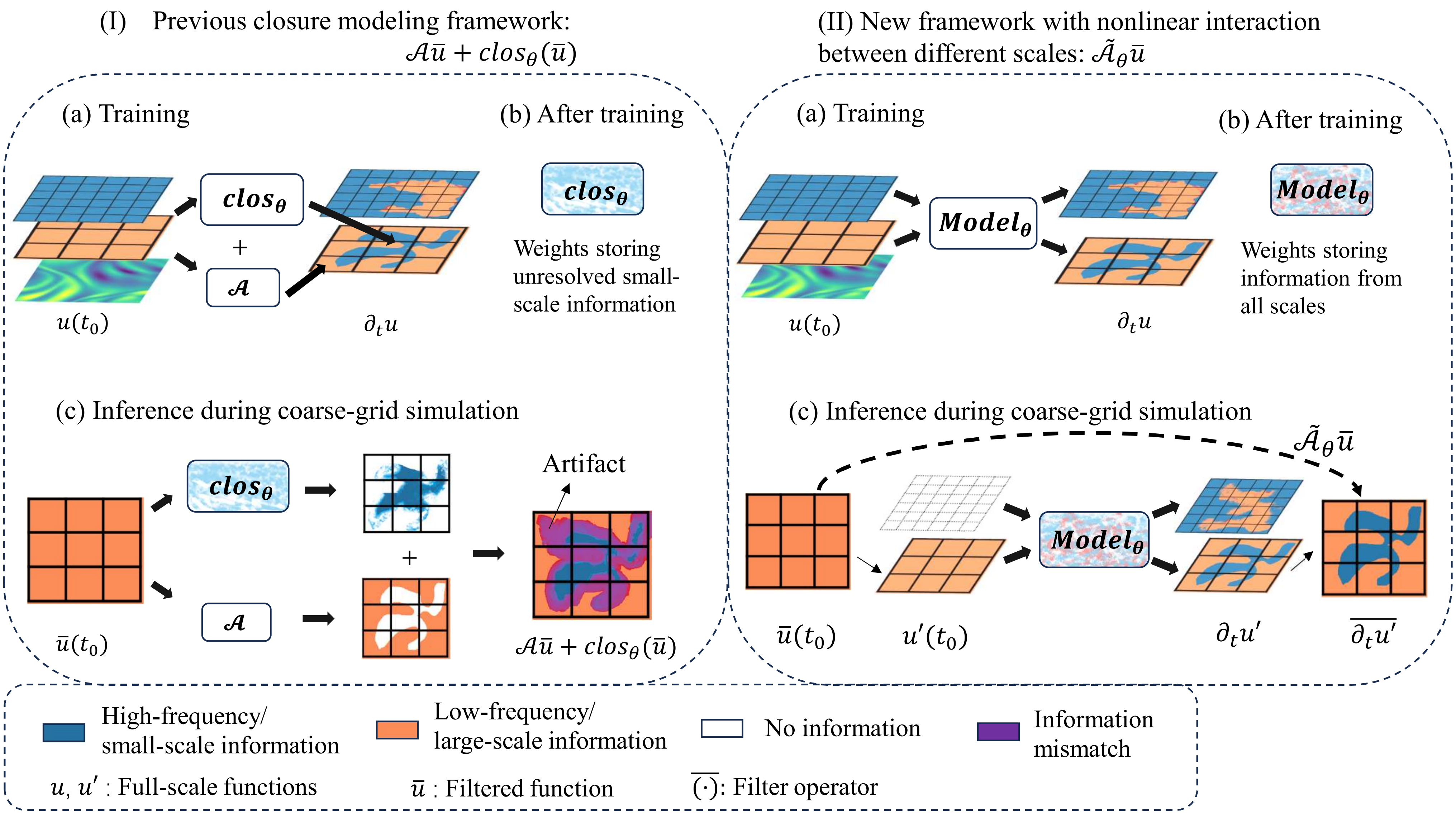
Neural Operator for Coarse-Graining

[Property]

- The input can be either course-grid or fine-grid.
- All viewed as discretization of an inf-resolution function.

[Method]

- Train a neural operator G_θ to approximate the solution operator (semigroup).
- $G_\theta u \approx \{S(t)u\}_{0<t\leq h}$, h: parameter.
- Taking coarse-grid input $v \in F(H)$, $G_\theta v$, $G_\theta(G_\theta v\,|_{t=h})$, $G_\theta(G_\theta(G_\theta v\,|_{t=h})|_{t=h})$ … … .



$$\text{Infinitesimal generator: } \mathcal{A} = \lim_{t \to 0} \frac{1}{t}(S_t - I)$$

Convergence Guarantee

Theorem 5.1 *For any $h > 0$, denote $\hat{\mu}_{h,\theta} := \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \delta_{\mathcal{G}_\theta^n v_0(x)}$, any $v_0(x)$ with $x \in D'$. For any $\epsilon > 0$, there exists $\delta > 0$ s.t. as long as $\|(\mathcal{G}_\theta u)(\cdot, h) - S(h)u\|_{\mathcal{H}} < \delta, \forall u \in \mathcal{H}$, we have $\mathcal{W}_{\mathcal{H}}(\hat{\mu}_{h,\theta}, \rho_1 *) < \epsilon$, where $\mathcal{W}_{\mathcal{H}}$ is Wasserstein distance with $\|\cdot\|$ being the cost function.*

- Large time steps – Faster convergence to the attractor.
- An imperfect neural operator is fine!
- Perusing small per-step error is enough.
- Good estimation for any long-term statistics!

Convergence Guarantee

Theorem 5.1 *For any $h > 0$, denote $\hat{\mu}_{h,\theta} := \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \delta_{\mathcal{G}_\theta^n v_0(x)}$, any $v_0(x)$ with $x \in D'$. For any $\epsilon > 0$, there exists $\delta > 0$ s.t. as long as $\|(\mathcal{G}_\theta u)(\cdot, h) - S(h)u\|_{\mathcal{H}} < \delta, \forall u \in \mathcal{H}$, we have $\mathcal{W}_{\mathcal{H}}(\hat{\mu}_{h,\theta}, \rho_1 *) < \epsilon$, where $\mathcal{W}_{\mathcal{H}}$ is Wasserstein distance with $\|\cdot\|$ being the cost function.*

- Important component of the proof:

[Property](Resolution-Invariance of FNO) For any $u_0 \in H$, there exists $u \in H$, such that $\overline{u} = \overline{u}_0$; $G(\overline{u}_0) = \overline{G(u)}$. The CGS is consistent with trajectory from u.

[Thm](Shadowing Lemma) In hyperbolic set of S_t, for any $\epsilon > 0$, there exists δ such that if a sequence u_n satisfies $|S(\Delta t)u_n - u_{n+1}| < \delta$ for all $n \in N$, then there exists v such that $|u_n - S(n\Delta t)v| < \epsilon$ for all n.

- Even if the simulation is deviating far away from original traj, it is consistently close to a traj and the error of resulting statistics have an upper-bound.

[Thm](Moore) For any $h > 0$, any initialization $u \in H$, $\lim_{N\to\infty} \frac{1}{N} \sum_n O(S_{nh}u) = \langle O \rangle$.

- Large time step does not harm performance.

Practical Algorithm: Physics-informed Neural Operator

- Limited fully-resolved data.
- The equation contains all the information!

- Supervised learning (fitting input-output data).

  Data pairs from coarse-grid simulation. (Inaccurate, cheap)

- Supervised learning (fitting input-output data).

  Data pairs from fully-resolved simulation. (Accurate, expensive, scarce)

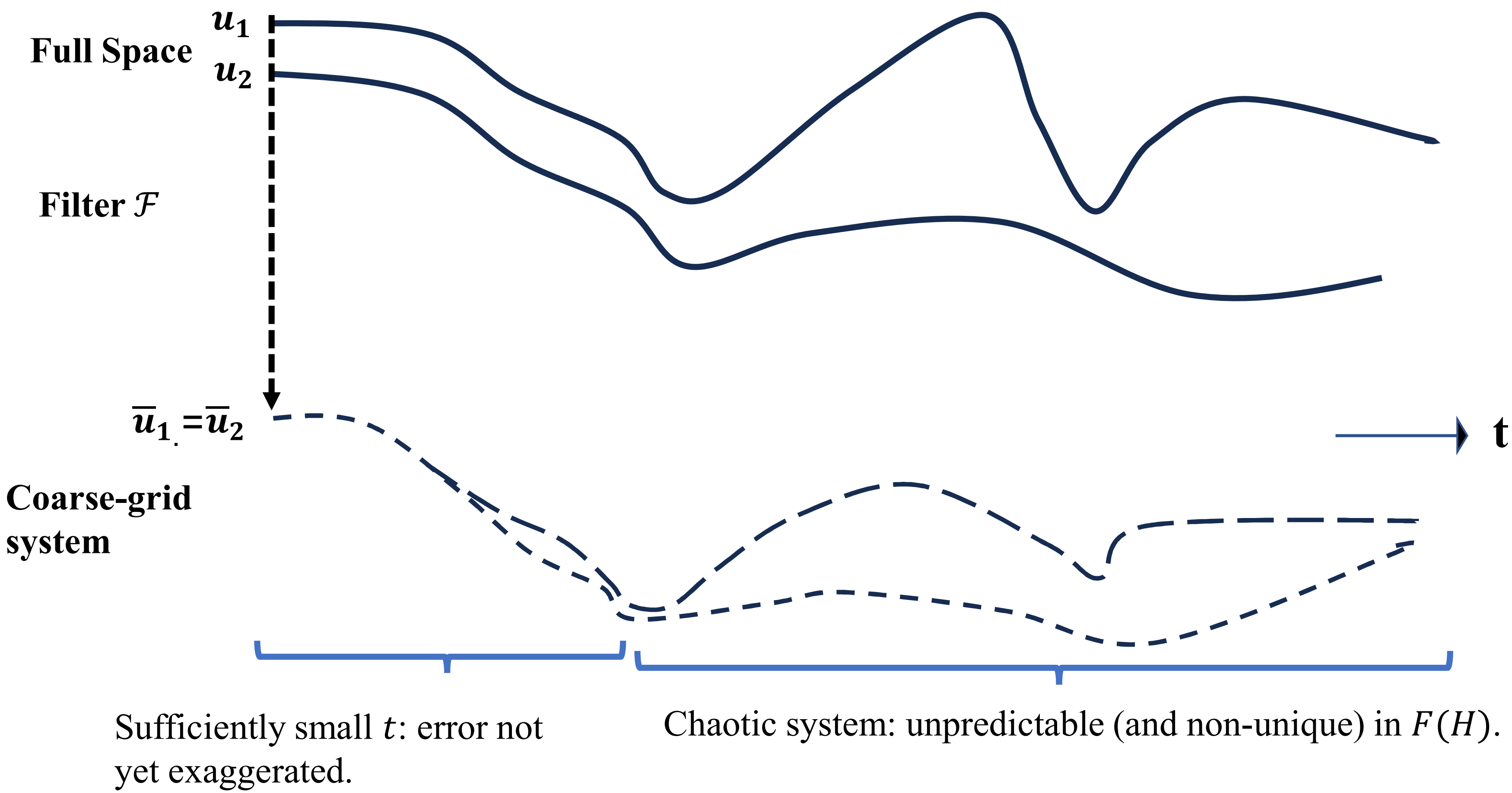
- Training with physics-informed loss.

  Random input functions. (Free!)

$$J_{pde}(\theta;\mathfrak{D}) = \frac{1}{|\mathfrak{D}|}\sum_{i\in\mathfrak{D}} \|(\partial_t - \mathcal{A})\mathcal{G}_\theta u_{0i}(x)\|_{L^2(\Omega\times[0,h])}$$

Evaluation: what is a good metric?

- RMSE (relative error) is meaningless!

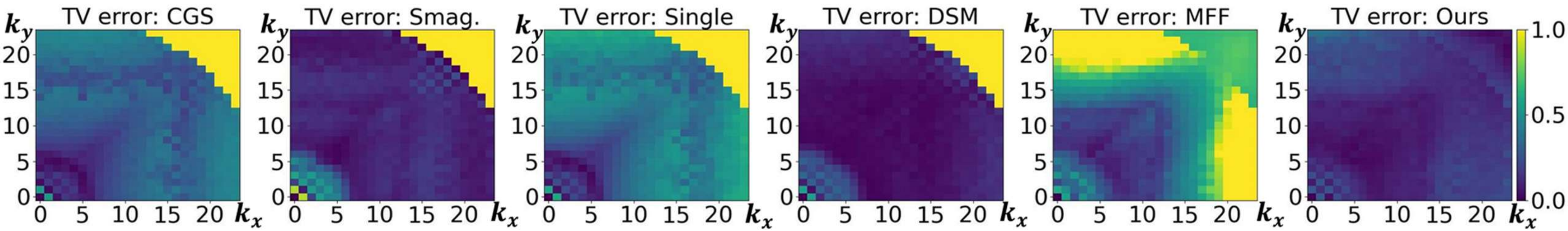


Evaluation: what is a good metric?

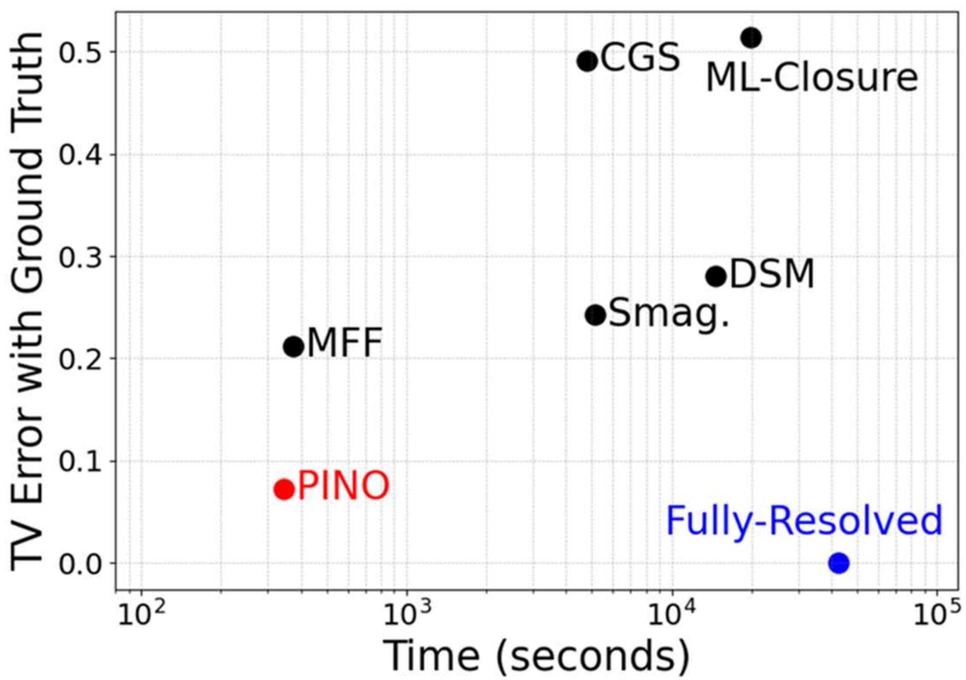
- RMSE (relative error) is meaningless!
- Statistics are not comprehensive.
- Directly compare the measure. Compare limit distribution from CGS and ρ_1^* .

Evaluation: what is a good metric?

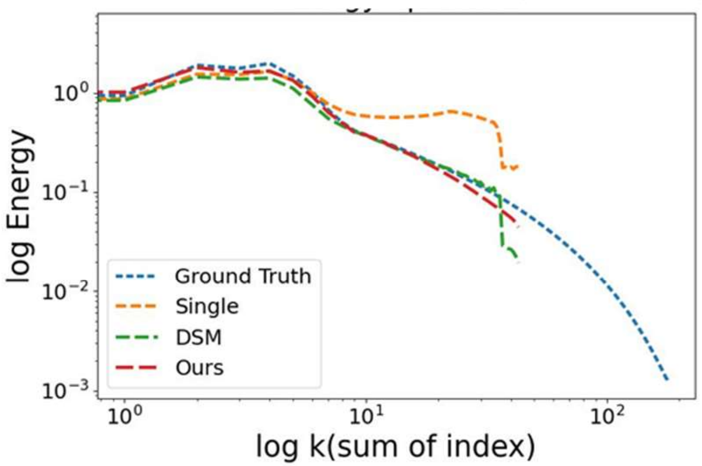
- RMSE (relative error) is meaningless!
- Statistics are not comprehensive.
- Directly compare the measure. Compare limit distribution from CGS and ρ_1^* .



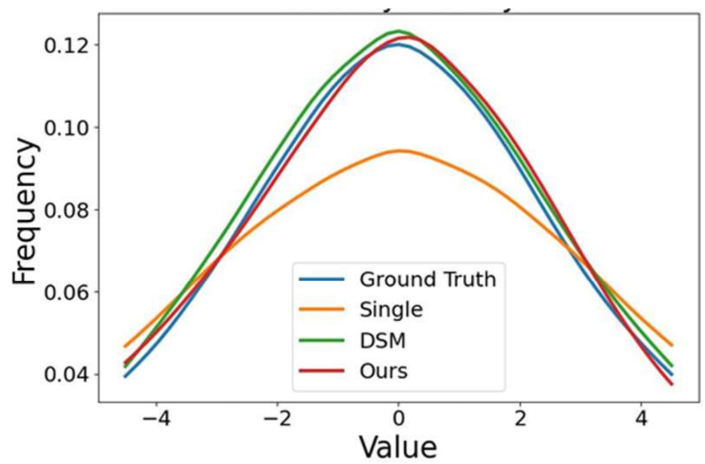
- 2D Kolmogorov Flow (Reynolds number 1.6×10^4)
- Compute marginal distribution over each basis function $\exp\big(i(jx + ky)\big), j, k \in \mathbb{Z}$.
- Total variation error of each marginal distribution.



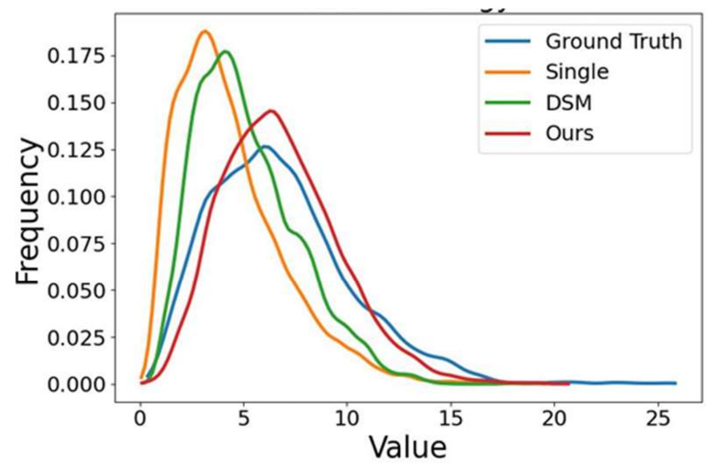
➢All ML method are trained with same fully-resolved dataset. ($O(10^2)$ snapshots.)



(a) Energy Spectrum



(b) Vorticity Distribution



(c) Kinetic Energy Distribution

Summary

Method	Optimal statistics	High-res. training data DNS Snapshots	Trajs.	Complexity
Fully-resolved Simulation, e.g., DNS [33, 34]	✓	-		$Re^{3.52}$
Coarse-grid Simulation, e.g., LES [33, 34]	✗	-		$Re^{2.48}$
Single-state model [28]	✗	24000	8	$Re^{2.48}$
History-aware model[35]	✗	250000	50	$Re^{2.48}$
Latent Neural SDE[32]	✗	179200	28	$\frac{1}{\delta t} Re^{1.86}$
Physics-Informed Operator Learning (Ours)	✓	110	1	$Re^{1.86}$

Re: Reynolds number

1. Require large number of FRS data (which are not available usually).
2. Require a coarse-grid solver (can be even faster).
3. Cannot give the optimal estimations of statistics in ideal case, i.e. perfect training.

Outline

- (1) Problem formulation: long-term statistics & coarse-grid simulations.
- (2) Limitation of Closure Models: Non-uniqueness issue
- (3) Theoretical Perspective via Measure Flow:
  - Learning-based closures: impractical reliance on hi-fidelity data.
- (4) Coarse-graining with Neural Operator
- (5) Conclusion & Future Direction & Discussion

Takeaway Messages

- Always remember the goal: long-term statistics instead of transient trajectory.
- Closure Modeling approach: fundamental shortcoming unless special structure of the system.
  - Non-uniqueness issue.
  - Impractical reliance on high-fidelity data.
  - A systematic way to verify convergence through functional measure flow.
- More promising to have implicit nonlinear interactions between different scales.
- Neural Operator as a CG approach.

Future Directions

- How much data do we need from fully-resolved simul?
  - Tradeoff: Time cost for generating data vs. time cost for tuning parameters.
  - Guide for practitioners: the more, the better.
  - Advanced optimization techniques for minimizing physics-informed loss.
- A unified model that generalizes.
  - Input=[initial condition] ➡ input= concatenate[initial, boundary, force, coeff, geometry]
  - Advanced model architectures.

Thanks!

Check more details at https://arxiv.org/abs/2408.05177

Beyond Closure Models: Learning Chaotic-Systems via Physics-Informed Neural Operators