RealPDEBench: A Benchmark for Complex Physical Systems with Real-World Data

[ICLR Oral] The first benchmark for scientific ML that integrates real-world measurements with paired numerical simulations.

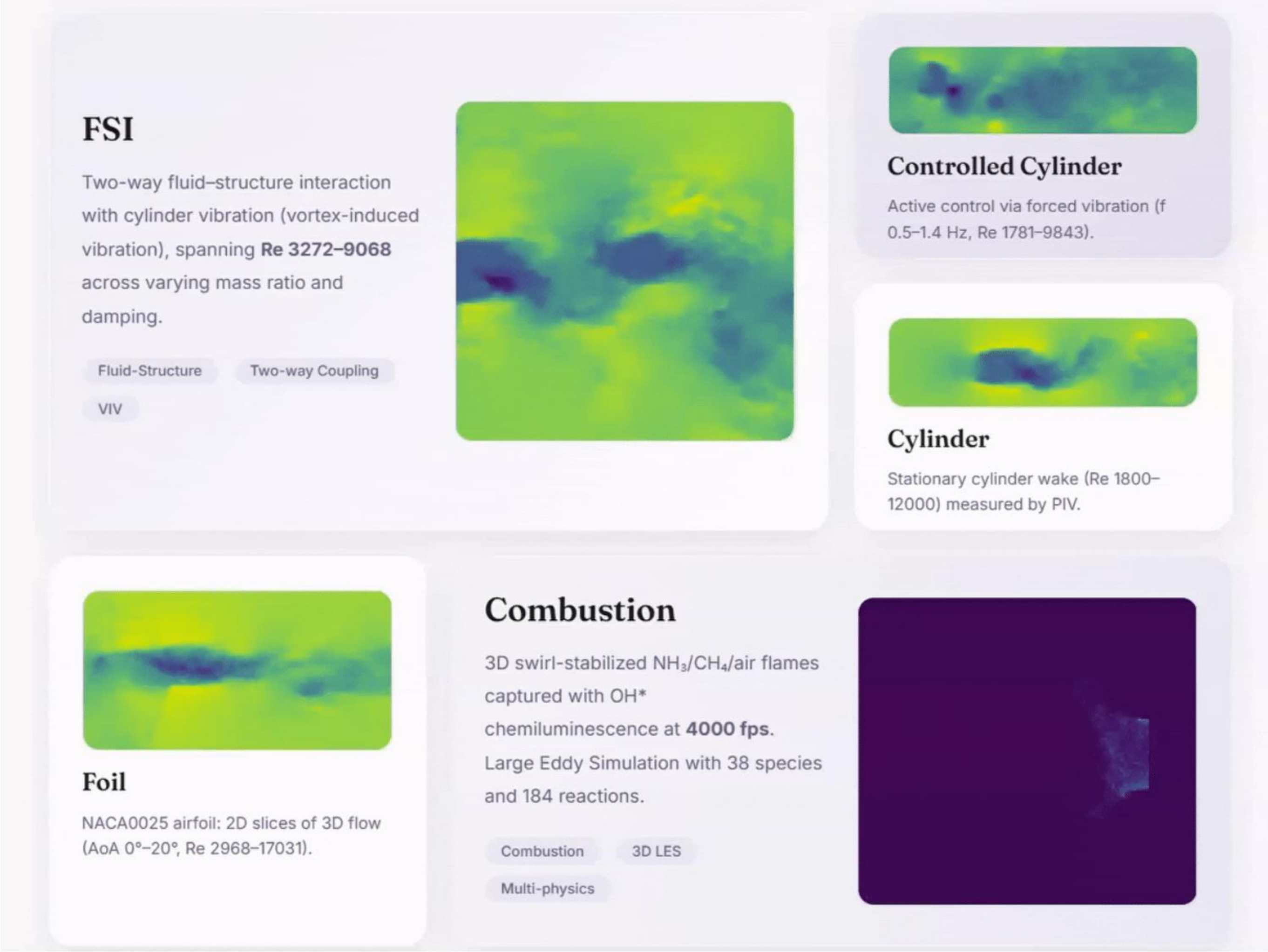
Peiyan Hu*, Haodong Feng*, Hongyuan Liu*, Tongtong Yan, Wenhao Deng, Tianrun Gao, Rong Zheng, Haoren Zheng, Chenglei Yu, Chuanrui Wang, Kaiwen Li, Zhi-Ming Ma, Dezhi Zhou, Xingcai Lu, Dixia Fan, Tailin Wu†

*Equal contribution. † Corresponding author.

Website: https://realpdebench.github.io/

Outline

- 1. Background and Motivation
- 2. Dataset:
  - from Fluid to Combustion
- 3. Tasks:
  - Simulated Training
  - Real-world Training
  - Real-world Finetuning
- 4. 9 Metrics
- 5. 10 Baselines
- 6. Experiment results



Background and Motivation

- **Scientific background:**
  - Core problem in science and engineering.
  - Dynamics are highly nonlinear, high-dimensional, and strongly coupled.
  - Fluid, combustion, nuclear fusion, aerospace…
- **Current limitation:**
  - Most scientific ML models are still trained and evaluated mainly on **simulated data**, rather than on **real-world measurements**.
  - → hinders the real-world application!
- **Key challenge:**

real-world data:
measurement noise,
missing modalities,
experimental imperfections...

GAP

simulated data:
numerical errors,
simplified physical assumptions...

Background and Motivation

- Existing benchmarks cannot fully answer a crucial question:

“How well do scientific ML models work in real-world settings?”

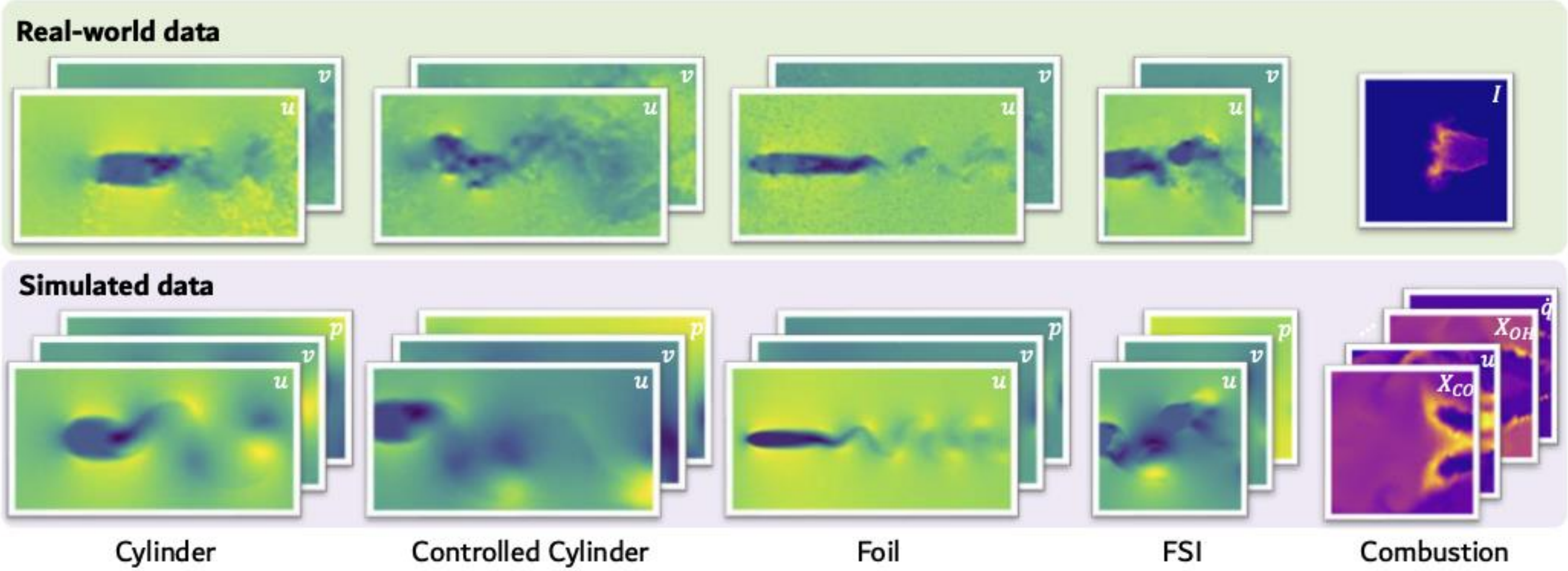
- Why this matters:
  - fair evaluation, practical deployment, sim-to-real transfer, learning from noisy observations…
- Gap in prior work:
  - Existing benchmarks mainly focus on **simulated PDE data**, while available real-world datasets are often **small-scale, sparse, and not designed for machine learning**.
- Motivation of us: **RealPDEBench** is proposed to bridge this gap.
  - Pairs **real-world measurements** with **matched simulations**,
  - enables **direct comparison** between the two,
  - supports studying new tasks such as **sim2real problems.**

Overall goal: Move scientific ML from **simulation-only success** toward **real-world applicability**.

Dataset: from Fluid to Combustion

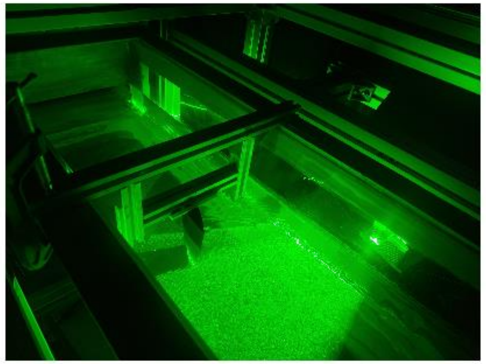
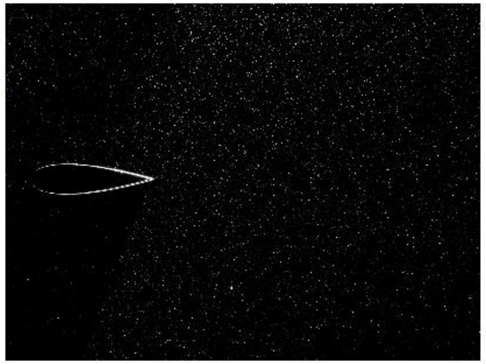
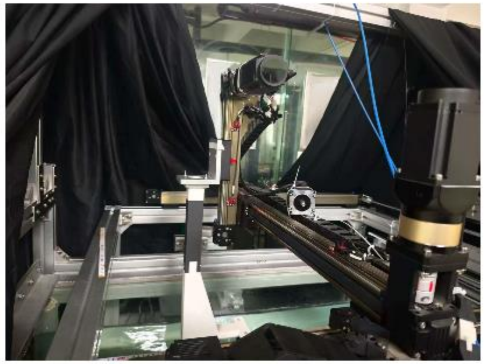
- **5 representative scenarios**:
  - Cylinder, Controlled Cylinder, Fluid-Structure Interaction, Foil, and Combustion.
  - Cover key physical challenges such as wake transition, control response, fluid–structure coupling, 3D flow effects, and reactive flows.

- **736 paired trajectories** of **real-world** and **simulated** data under multiple operating conditions.



Dataset: from Fluid to Combustion

- **Real-world data collection:**
  - Fluid datasets: water tunnels, Particle Image Velocimetry (PIV)
  - Combustion dataset: flame chemiluminescence imaging
- **Simulated data collection:**
  - CFD solvers
  - Fluid datasets: Lilypad and Waterlily
  - Combustion dataset: 3D unsteady LES and turbulence–chemistry modeling.
- **Data format:**
  - Each file contains one trajectory on a **uniform** spatial grid, together with the corresponding **physical parameters**.


3 Tasks

- **Purpose:**
  - reveal both the **limitations of simulated data** and the **benefits of combining simulation with real-world supervision**.

- **Testing**: In all settings, models are validated and tested on real-world data.

- **Simulated Training:** trained on **all simulated samples.**
- **Real-world Training:** trained directly on the **real-world samples in the training dataset.**
- **Real-world Finetuning:** first **pretrained on simulated data** and then **finetuned on real-world data.**

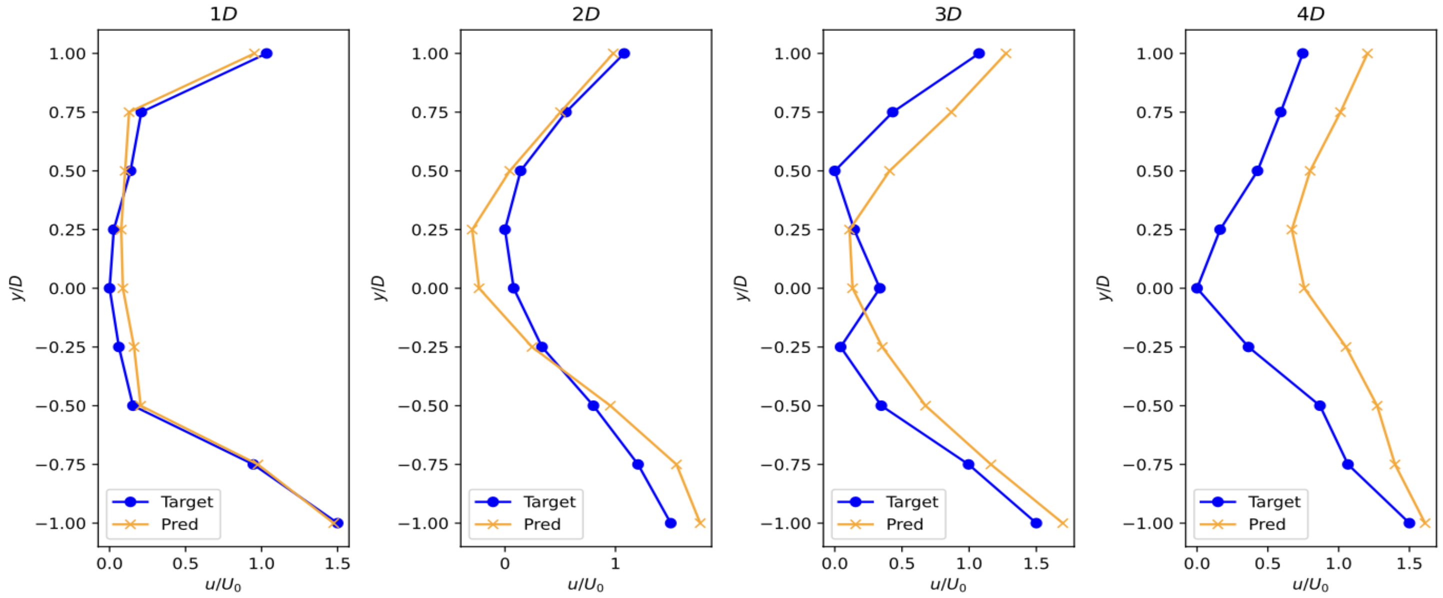
9 Metrics

- **Data-oriented metrics** & **physics-oriented metrics**:

  - **Data-oriented metrics:**
    - **RMSE**, **MAE**, **Relative L2 Error**
    - **R^2**: how much variance in the ground truth is explained by the model.
    - **Update Ratio:**
      - Evaluates **training efficiency**
      - Compares how many optimization updates needed to reach the target RMSE: **real-world finetuning** v.s **real-world training from scratch**.

9 metrics

- **Physics-oriented metrics:**
  - **Fourier Space Error (fRMSE):** a **frequency-domain metric**, and can also be analyzed across **low-, middle-, and high-frequency bands**.
  - **Frequency Error (FE):** focuses on **temporal periodicity**.
  - **Kinetic Energy Error (KE):** measures whether the predicted flow preserves the correct kinetic energy.
  - **Mean Velocity Profile Error (MVPE):** compares the **time-averaged velocity profile** at selected probe locations, evaluates **long-term behavior**.



⬅ MVP of U-Net's 10-round autoregressive prediction on Cylinder.

10 baselines

Class	Baseline
Foundation Model	DPOT-L (509M)
	DPOT-S (30M)
Neural Operator	Fourier Neural Operator (FNO)
	Convolutional Neural Operator (CNO)
	DeepONet
	Multiwavelet Transform Operator (MWT)
	Wavelet Diffusion Neural Operator (WDNO)
Transformer	Transolver
	Galerkin Transformer (GK-Transformer)
Traditional & CNN	Dynamic Mode Decomposition (DMD)
	U-Net

Experiment results

Dataset	Baseline	Params	Simulated Training RMSE	Simulated Training Rel L_2	Simulated Training fRMSE	Real-world Training RMSE	Real-world Training Rel L_2	Real-world Training fRMSE	Real-world Finetuning RMSE	Real-world Finetuning Rel L_2	Real-world Finetuning fRMSE	Real-world Finetuning Ratio
Cylinder	U-Net	23.0 M	0.0758	0.2165	0.0122	0.0700	**0.0701**	0.0103	0.0632	**0.0728**	0.0097	0.3636
	CNO	8.0 M	0.0729	0.1849	0.0113	0.0424	0.0897	0.0069	0.0403	0.1013	0.0066	1.0000
	DeepONet	3.6 M	0.0863	0.3592	0.0132	0.0713	0.1528	0.0108	0.0661	0.1503	0.0107	0.5758
	FNO	50.4 M	0.0739	0.2575	0.0115	0.0585	0.0855	0.0087	0.0545	0.0780	0.0087	**0.1111**
	WDNO	104.3 M	0.0699	0.2053	0.0105	0.0689	0.1666	0.0091	0.0513	0.0969	0.0073	0.1250
	MWT	2.9 M	0.0733	0.2240	0.0107	0.0540	0.0832	0.0081	0.0584	0.0850	0.0088	0.9253
	GK-Transformer	84.4 M	0.0876	0.2941	0.0139	0.1196	0.0898	0.0166	0.0995	0.0908	0.0146	0.4400
	Transolver	4.3 M	0.1121	0.3224	0.0174	0.1093	0.1887	0.0160	0.0965	0.1723	0.0144	1.0000
	DPOT-S-FT	30.8 M	0.0513	0.1474	0.0076	0.0586	0.0983	0.0086	0.0440	0.0766	0.0067	0.1250
	DPOT-L-FT	649.8 M	**0.0486**	**0.1446**	**0.0070**	**0.0390**	0.0812	**0.0056**	**0.0394**	0.0733	**0.0059**	1.0000
	ML Average	-	0.0752	0.2356	0.0115	0.0692	0.1106	0.0101	0.0613	0.0997	0.0093	0.5666
	DMD	-	-	-	-	-	-	-	0.0862	0.3590	0.0114	-
Controlled Cylinder	U-Net	23.0 M	0.0195	0.1360	0.0029	**0.0080**	**0.0555**	0.0010	**0.0079**	**0.0543**	**0.0009**	0.7632
	CNO	8.0 M	**0.0167**	**0.1209**	**0.0022**	0.0081	0.0583	**0.0009**	0.0080	0.0574	**0.0009**	0.8400
	DeepONet	3.6 M	0.0540	0.4256	0.0134	0.0309	0.2399	0.0058	0.0291	0.2284	0.0052	0.5200
	FNO	50.4 M	0.0285	0.2007	0.0059	0.0097	0.0723	0.0012	0.0094	0.0702	0.0011	0.5217
	WDNO	359.8 M	0.0240	0.1829	0.0037	0.0115	0.0927	0.0014	0.0101	0.0752	0.0013	**0.3617**
	MWT	2.9 M	0.0220	0.1594	0.0037	0.0102	0.0747	0.0011	0.0102	0.0746	0.0010	0.5487
	GK-Transformer	50.8 M	0.0261	0.1816	0.0050	0.0103	0.0767	0.0012	0.0101	0.0748	0.0012	0.5200
	Transolver	4.3 M	0.0294	0.1894	0.0069	0.0171	0.1200	0.0024	0.0168	0.1176	0.0022	0.4285
	DPOT-S-FT	30.8 M	0.0233	0.1722	0.0039	0.0084	0.0598	0.0010	0.0085	0.0615	0.0010	1.0000
	DPOT-L-FT	649.8 M	0.0248	0.1784	0.0043	0.0084	0.0603	0.0010	0.0085	0.0611	0.0010	1.0000
	ML Average	-	0.0268	0.1947	0.0052	0.0123	0.0910	0.0017	0.0119	0.0875	0.0016	0.6504
	DMD	-	-	-	-	-	-	-	0.0340	0.2233	0.0059	-

FSI	U-Net	23.0 M	**0.0223**	**0.1589**	**0.0025**	**0.0085**	**0.0583**	**0.0007**	**0.0084**	**0.0579**	**0.0007**	0.6667
	CNO	8.0 M	0.0241	0.1724	0.0029	0.0105	0.0741	0.0009	0.0096	0.0679	0.0008	0.5600
	DeepONet	3.4 M	0.0637	0.4606	0.0097	0.0350	0.2502	0.0051	0.0332	0.2368	0.0048	0.3125
	FNO	268.5 M	0.0426	0.3095	0.0059	0.0129	0.0892	0.0012	0.0127	0.0881	0.0012	0.5714
	WDNO	91.7 M	0.0369	0.2700	0.0050	0.0117	0.0817	0.0011	0.0116	0.0824	0.0011	0.5116
	MWT	2.9 M	0.0339	0.2487	0.0046	0.0128	0.0910	0.0010	0.0128	0.0912	0.0010	0.9992
	GK-Transformer	67.6 M	0.0307	0.2241	0.0039	0.0127	0.0916	0.0011	0.0124	0.0898	0.0010	0.6200
	Transolver	4.3 M	0.0305	0.2119	0.0043	0.0236	0.1567	0.0027	0.0223	0.1480	0.0025	**0.1600**
	DPOT-S-FT	41.3 M	0.0260	0.1886	0.0031	0.0105	0.0746	0.0008	0.0099	0.0701	**0.0007**	0.2500
	DPOT-L-FT	673.5 M	0.0262	0.1898	0.0030	0.0099	0.0687	0.0008	0.0097	0.0668	0.0008	0.3125
	ML Average	-	0.0337	0.2434	0.0045	0.0148	0.1036	0.0015	0.0143	0.0999	0.0015	0.4964
	DMD	-	-	-	-	-	-	-	0.0450	0.2955	0.0060	-
Foil	U-Net	23.0 M	0.0272	0.0745	0.0039	**0.0100**	**0.0159**	0.0011	**0.0094**	**0.0145**	**0.0008**	0.5116
	CNO	8.0 M	0.0227	0.0438	0.0028	0.0136	0.0253	0.0018	0.0114	0.0206	0.0013	0.4400
	DeepONet	3.6 M	0.0339	0.0565	0.0051	0.0222	0.0363	0.0028	0.0226	0.0375	0.0029	0.5185
	FNO	50.4 M	0.0274	0.0540	0.0037	0.0130	0.0228	0.0015	0.0120	0.0206	0.0012	**0.3192**
	WDNO	358.4 M	0.0242	0.0490	**0.0026**	0.0162	0.0444	0.0018	0.0106	0.0181	0.0010	0.1035
	MWT	2.9 M	0.0264	0.0492	0.0036	0.0133	0.0227	0.0015	0.0125	0.0210	0.0012	0.4894
	GK-Transformer	50.8 M	0.0277	0.0512	0.0038	0.0142	0.0245	0.0018	0.0138	0.0237	0.0017	0.5918
	Transolver	4.3 M	0.0345	0.0465	0.0050	0.0220	0.0370	0.0027	0.0138	0.0237	0.0017	0.5918
	DPOT-S-FT	41.3 M	**0.0221**	**0.0397**	0.0027	0.0106	0.0166	**0.0010**	0.0109	0.0174	0.0012	1.0000
	DPOT-L-FT	673.5 M	0.0229	0.0402	0.0029	0.0105	**0.0159**	0.0011	0.0109	0.0161	0.0012	1.0000
	ML Average	-	0.0269	0.0505	0.0036	0.0146	0.0261	0.0017	0.0128	0.0213	0.0014	0.5566
	DMD	-	-	-	-	-	-	-	0.0322	0.0520	0.0041	-
Combustion	U-Net	23.3 M	**0.0358**	**0.7290**	**0.0051**	0.0216	0.5487	0.0026	0.0213	0.5403	0.0025	0.5682
	CNO	8.0 M	0.0401	0.8274	0.0057	0.0248	0.6030	0.0032	0.0238	0.5877	0.0030	0.7083
	DeepONet	3.5 M	0.0403	0.8565	0.0056	0.0229	0.5751	0.0028	0.0227	0.5723	0.0028	0.7800
	FNO	67.1 M	0.0363	0.7606	0.0052	0.0226	0.5664	0.0027	0.0225	0.5680	0.0027	0.4286
	WDNO	122.7 M	0.0439	1.1016	0.0061	0.0380	0.8605	0.0055	0.0314	0.7441	0.0044	**0.2632**
	MWT	2.9 M	0.0367	0.7536	0.0052	0.0221	0.5560	0.0027	0.0220	0.5549	0.0026	0.9982
	GK-Transformer	67.6 M	0.0400	0.9143	0.0056	0.0247	0.6083	0.0031	0.0258	0.6421	0.0034	1.0000
	Transolver	4.3 M	0.0442	0.9056	0.0061	0.0375	0.7825	0.0050	0.0375	0.7836	0.0050	1.0000
	DPOT-S-FT	41.5 M	0.0393	0.7842	0.0054	0.0209	0.5349	**0.0024**	0.0211	0.5378	0.0024	1.0000
	DPOT-L-FT	674.2 M	0.0378	0.7750	0.0053	**0.0208**	**0.5331**	**0.0024**	**0.0206**	**0.5318**	**0.0023**	0.8125
	ML Average	-	0.0394	0.8408	0.0055	0.0256	0.6169	0.0032	0.0249	0.6063	0.0031	0.7559
	DMD	-	-	-	-	-	-	-	0.0914	1.3360	0.0110	-

Experiment results

Findings:

1. Error characteristics:

Real-world datasets are mainly affected by **sensor noise**, whereas simulated datasets are mainly limited by **numerical and modeling errors**.

2. Advantages of simulated data:

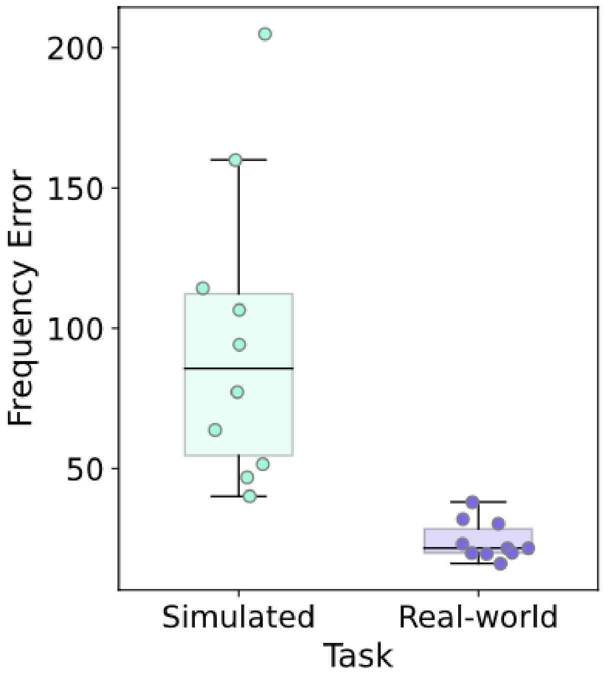
Have lower cost, more accessible modalities, and no measurement-induced noise.

3. Sim-to-real performance gap:

Models trained on simulated data perform worse than those trained on real-world data when evaluated on real-world benchmarks.

4. Superiority of real-world training:

Real-data-trained models outperform simulated-data-trained models by a large margin, with **relative L2 improvements of 9.39%–78.91%**.



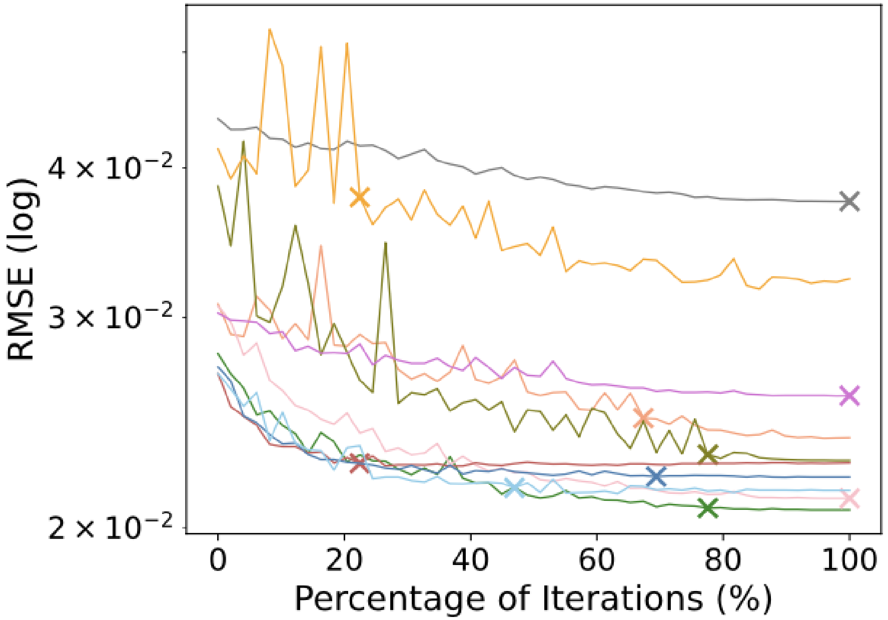
Findings:

5. Benefit of pretraining:

Simulated-data pretraining followed by real-world finetuning improves performance.

6. Faster convergence:

Simulated pretraining also accelerates optimization, reflected by **Update Ratios below 1** and faster RMSE reduction.



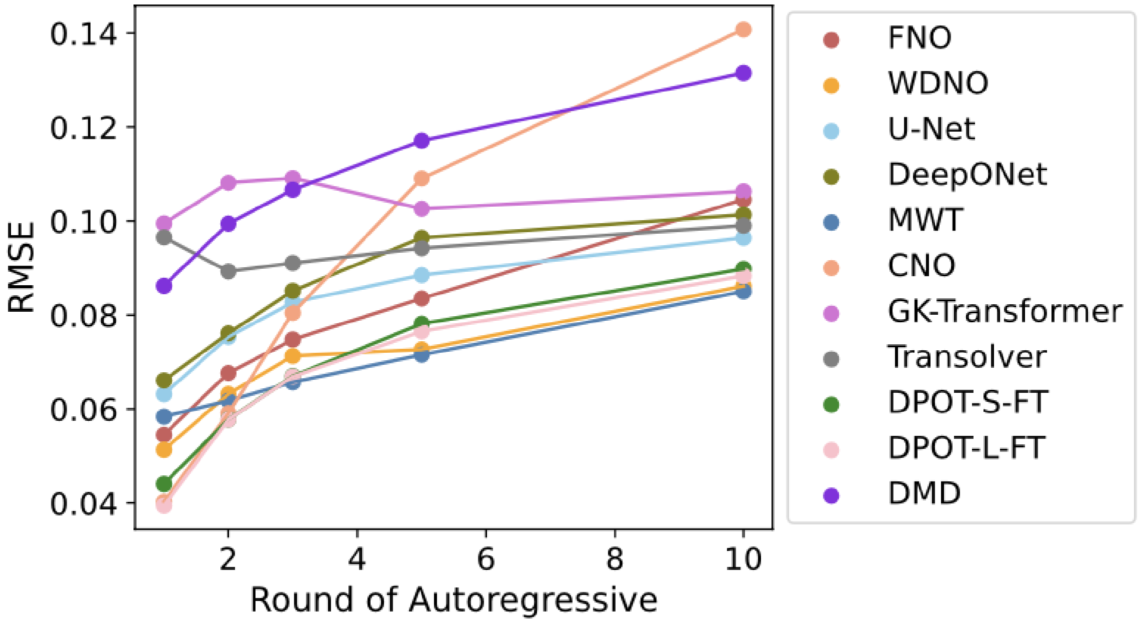
⬆️ Validation RMSE curves of real-world finetuning on Combustion.
×: the best RMSE of real-world training.

Experiment results

Findings:

7. Architecture trade-off:

Convolution-based models perform well on **pointwise RMSE**, while **operator-based and spectral models** better preserve **periodicity and global physical dynamics**.



8. Long-horizon stability:

Large pretrained operator models such as **DPOT** remain more stable over time.

Takeaway: The **first** benchmark for scientific ML that integrates real-world measurements with paired numerical simulations.

Peiyan Hu

Haodong Feng

Hongyuan Liu

Tailin Wu

Website: https://realpdebench.github.io/

Contact email:

hupeiyan18@mails.ucas.ac.cn

{fenghaodong, liuhongyuan, wutailin}@westlake.edu.cn