Accuracy, Performance, and Robustness of Physics-Informed Surrogate Models

Author: Filip Nedić Supervisors: Dr. Jing Sun & Dr. Tiexing Wang Examiner: Prof. Dr. Mathijs de Weerdt

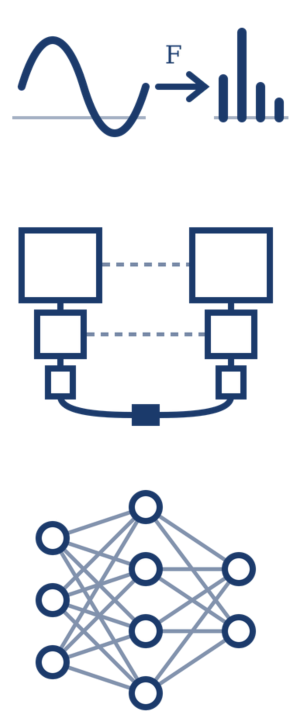
fnedic@tudelft.nl

1. Motivation

High-fidelity CFD simulations are essential for engineering design, but solving the underlying PDEs is computationally expensive, limiting rapid design iteration and large-scale parameter studies.

Neural surrogate models promise much faster predictions, but it's unclear how different architectures and training strategies hold up under noise or conditions outside their training data, so we benchmark MLP, U-Net, and Fourier Neural Operator (FNO) surrogates to find out.

3. Architectures

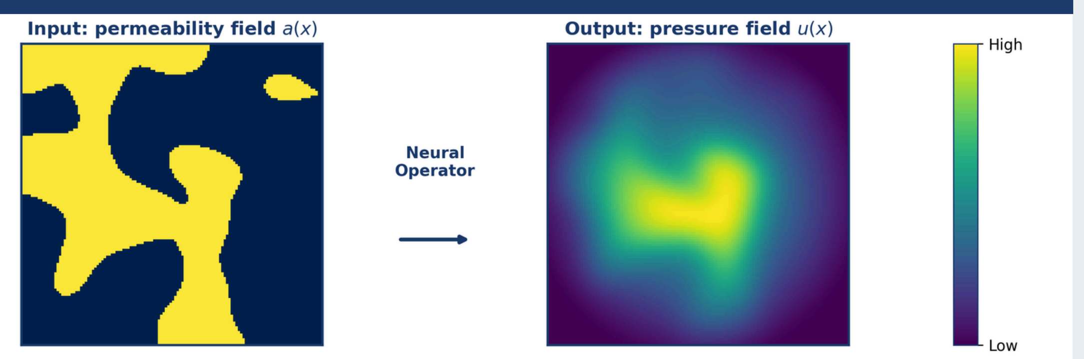


FNO: Operates in frequency space. captures global patterns, resolution-independent

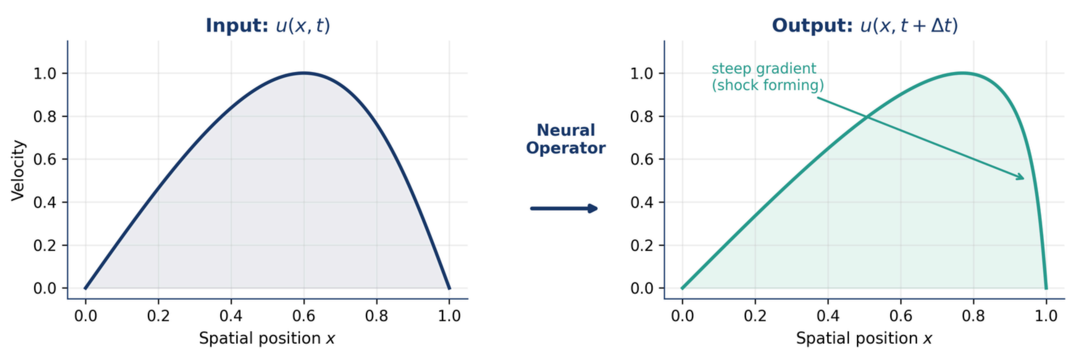
U-Net: Convolutional encoder-decoder. captures local detail via skip connections

MLP: Fully-connected baseline. no spatial awareness, treats the grid as a flat vector

4. PDEs



2D Darcy Flow: Models how fluid pushes through a porous material. Given the material's permeability, the operator predicts the resulting pressure field, a steady-state mapping with no time dependence.



1D Burgers' Equation: Models a fluid velocity that steepens over time. Given the velocity field at one instant, the operator predicts it one step later, testing whether the model can capture sharp, evolving gradients.

Datasets have been taken from PDEBench.

2. Research Question

To what extent do different training strategies, including data-driven training, physics-informed regularisation, and noise augmentation, affect the accuracy, computational efficiency, and robustness of MLP, U-Net, and FNO surrogate models?

5. Methodology

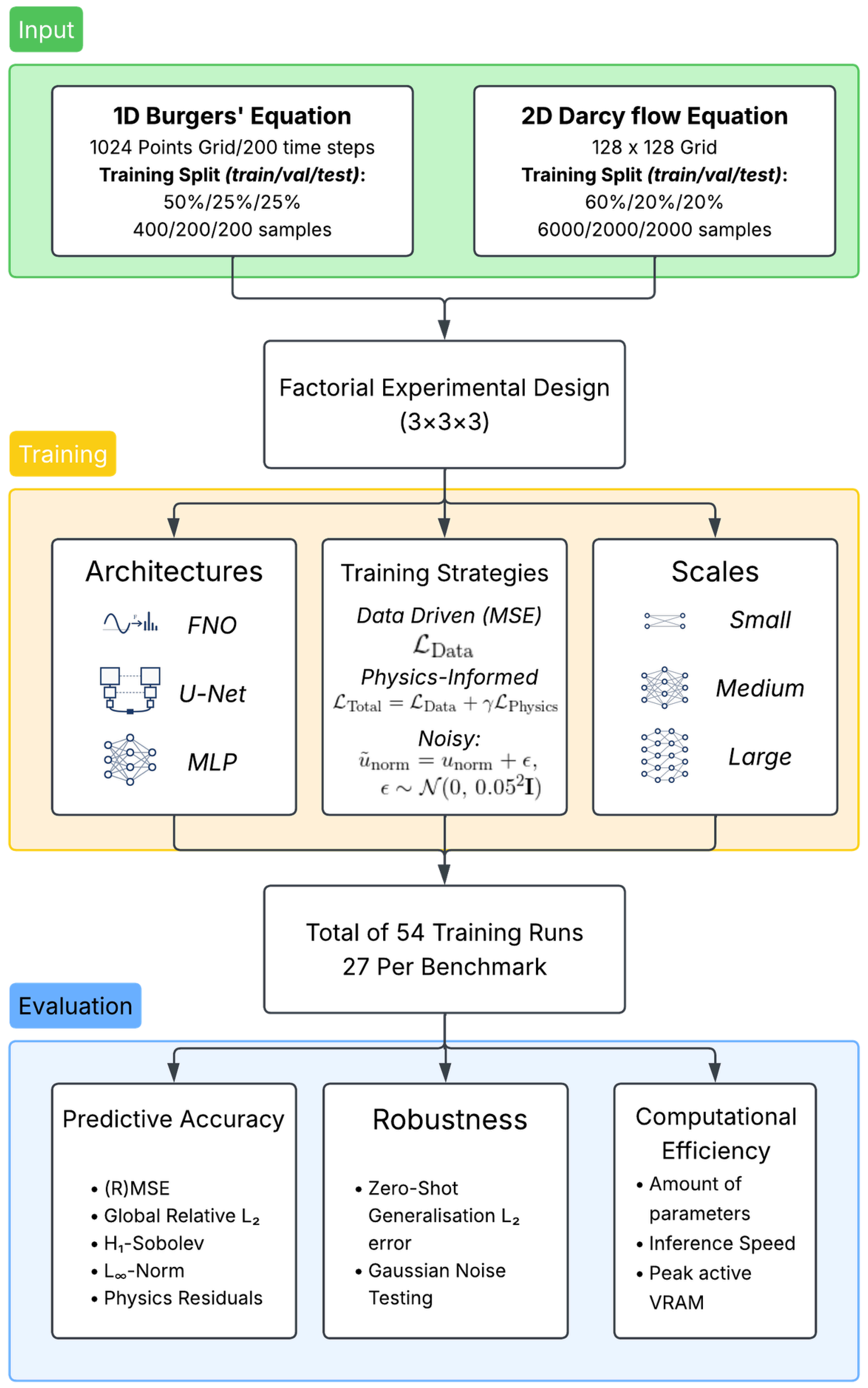


Table 1: Structural parameter scaling across architecture and scale.

Architecture	Small	Medium	Large
FNO	8 modes 16 latent channels 16 decoder channels	10 modes 24 latent channels 32 decoder channels	12 modes 32 latent channels 64 decoder channels
U-Net	16 channels	32 channels	64 channels
MLP	3 layers 512 hidden dim	4 layers 1024 hidden dim	5 layers 2048 hidden dim

- *400 epochs with early stopping (Patience 60)*
- *Batch size = 20*
- *Adam Optimiser*
- *Adaptive physics-loss weighting (γ): gradient-based EMA, 10-epoch ramp-up*

6a. Results (Predictive Accuracy) 2D Darcy flow

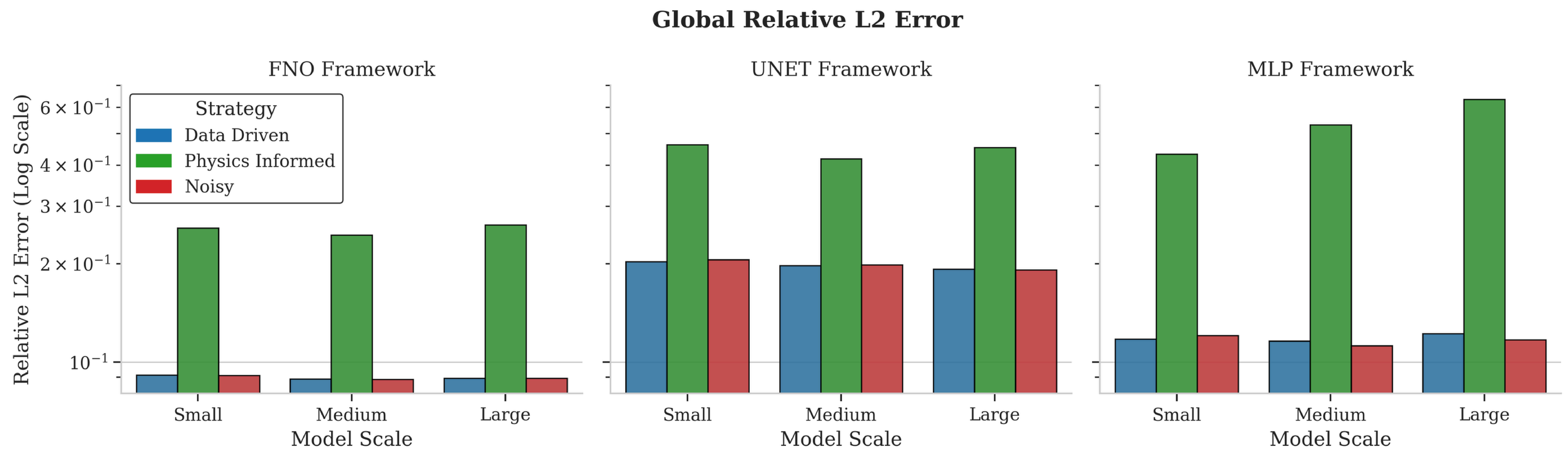


Figure 2: Global Relative L2 Error for 2D Darcy flow Equation

- FNO has the lowest, most stable baseline error. MLP beats U-Net's baseline on 2D Darcy Flow
- Physics-informed error rises with scale only for MLP. FNO and U-Net stay flat across scales
- On 1D Burgers', FNO and U-Net look strong on average, but huge per-sample variance means frequent, severe localised failures
- MLP repeats its negative scaling trend on Burgers' too, bigger models perform worse

6b. Results (OOD Robustness) 2D Darcy

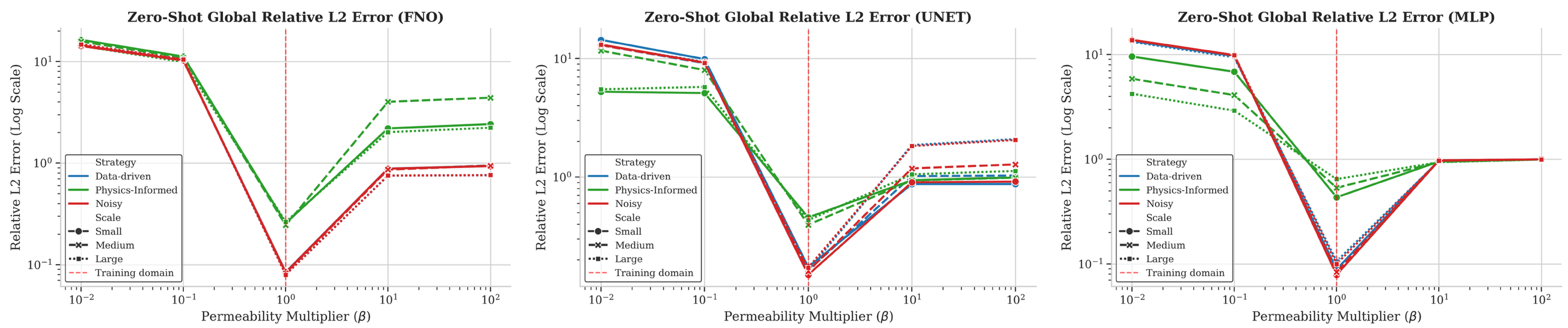


Figure 1: Out-of-Distribution Global Relative L2 error for 2D Darcy flow Equation

- Every architecture bottoms out at the training distribution (β=1) and rises with distance from it
- MLP and U-Net's physics-informed error stays lower than FNO's at extreme shifts
- FNO's physics-informed variant is the outlier, it never recovers, even at the most extreme β
- 1D Burgers' stays more stable than Darcy across the same shift in viscosity (ν), for every architecture

6c. Results (Computational Efficiency) 2D Darcy

Table 2: Computational profiling

Architecture	Scale	Parameters	Latency (s)	Peak VRAM (MB)
FNO	S	263,670	32.97	294.5
	M	925,160	52.14	434.8
	L	2,366,300	68.88	581.6
U-Net	S	483,150	22.25	351.0
	M	1,927,800	42.75	631.3
	L	7,701,800	92.28	1,266.3
MLP	S	17,582,000	0.41	80.3
	M	37,770,000	0.81	157.3
	L	88,109,000	1.87	349.5

- **U-Net:** highest latency and memory cost at every scale
- **MLP:** fastest and lightest, despite far more parameters
- **FNO:** sits in between on both cost and accuracy
- 1D Burgers' statistics are roughly the same

7. Conclusion and Future Work

- Architecture fit depends on spatial complexity: MLP beats U-Net on Darcy despite losing on Burgers'.
- Physics constraints don't guarantee robustness under domain shift, benefit depends entirely on architecture.
- Compute trade-offs are architecture-specific. U-Net gives the best detail, FNO scales predictably, MLP trades accuracy for speed.

Future Work: Burgers' dataset is small. 400 samples can't rule out data scarcity as the cause. Only single-step prediction tested. Rollout would expose compounding errors single-step misses.