PROBABILISTIC RESIDUAL TRANSPORT BETWEEN MULTI-FIDELITY MANIFOLDS

Sahil Bhola*
Department of Aerospace Engineering
University of Michigan
Ann Arbor, MI 48109, USA
sbhola@umich.edu

Karthik Duraisamy
Department of Aerospace Engineering
University of Michigan
Ann Arbor, MI 48109, USA
kdur@umich.edu

ABSTRACT

Learning probabilistic surrogates for partial differential equations (PDEs) remains challenging in data-scarce regimes, where high-fidelity simulations are expensive, and many generative models lack resolution invariance. We formulate multi-fidelity PDE learning as probabilistic transport between solution manifolds, where low-fidelity solvers define a reference measure, and the model learns residual flows toward high-fidelity solutions in function space. We propose an operator-valued flow matching framework that parameterizes conditional vector fields directly in infinite-dimensional spaces, enabling resolution-invariant inference without retraining. Unlike prior operator-valued flow matching methods that learn full solution distributions at a single fidelity, we explicitly parameterize transport in a residual-function space defined by a physics-based reference solver, enabling data-efficient probabilistic operator learning. The resulting model combines linear operator structure with a FiLM-conditioned Fourier neural operator to capture expressive input-dependent generative dynamics. Across advection, Burgers', and flow-through-porous-media problems, including cross-resolution tasks, our method improves data efficiency, discretization generalization, and uncertainty quantification compared with single-fidelity probabilistic neural operators.

1 INTRODUCTION

Solving PDEs remains computationally demanding, particularly when high-fidelity simulations are required for design optimization, uncertainty quantification, or real-time inference. While neural operators and physics-informed learning have enabled data-driven surrogates (Brunton et al., 2020), most approaches rely on large volumes of high-fidelity data and struggle to incorporate inexpensive low-fidelity models in a principled probabilistic framework.

We view multi-fidelity PDE learning as *probabilistic transport between solution manifolds*. Low-fidelity solvers define a reference measure, and the learning objective is to model residual flows that transport this measure toward high-fidelity solution distributions in function space. This perspective shifts the problem from learning full solution operators to learning structured residual corrections, improving data efficiency while preserving physical inductive bias.

To realize this view, we introduce a probabilistic operator-learning framework based on operator-valued flow matching. The proposed model learns conditional vector fields directly in function space and parameterizes generative dynamics in a residual space defined relative to a low-fidelity solver. This formulation enables resolution-invariant inference, cross-resolution generalization, and calibrated predictive uncertainty.

Our main contributions are summarized as follows:

1. **Residual Transport View of Multi-Fidelity PDE Learning:** We reinterpret probabilistic multi-fidelity operator learning as transport between solution manifolds, where low-fidelity

*Corresponding author. This work was supported in part by the OUSD(RE) grant *'Physics-Aware Reduced Order Modeling for Nonequilibrium Plasma Flows'* and in part by the Los Alamos National Laboratory grant *'Algorithm/Software/Hardware Co-design for High Energy Density applications'*.

solvers characterize a reference measure, and the model learns residual flows toward high-fidelity solutions in function space.

2. **Operator-Valued Flow Matching Dynamics:** We introduce a conditional neural operator that parameterizes flow matching vector fields directly in infinite-dimensional function space, enabling probabilistic inference across arbitrary spatial resolutions without retraining.

3. **Residual-Augmented Probabilistic Operators:** We develop a residual-based formulation in which generative dynamics learn probabilistic corrections from low-fidelity surrogates, improving data efficiency and reducing predictive variance.

4. **Empirical Evaluation Across Fidelities and Resolutions:** We demonstrate improved accuracy, cross-resolution generalization, and uncertainty calibration on canonical PDE benchmarks, including 1D advection, the Burgers' equation, and flow through porous media governed by Darcy's law.

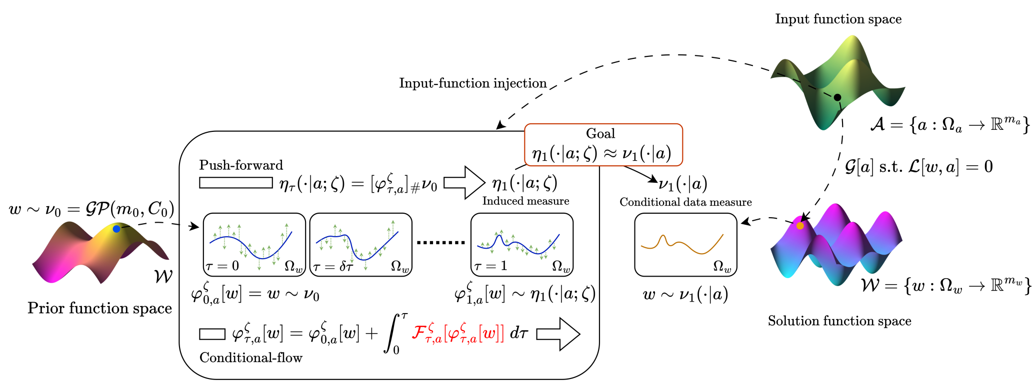

Figure 1: *Schematic of operator view of flow matching for PDE solution generation. The vector field operator $\mathcal{F}^\zeta_{\tau,\mathbf{a}}$ induces a probability measure $\eta_\tau(\cdot|a;\zeta)$ that seeks to approximate the conditional data measure $\nu_1(\cdot|a)$ at $\tau = 1$. Here, $\nu_1(\cdot|a)$ characterizes the PDE solution for a given $a \sim \nu_a$.*

2 RELATED WORK

We consider the *forward problem* of approximating the PDE solution operator $\mathcal{G} : \mathcal{A} \to \mathcal{W}$ that maps an input function $a \in \mathcal{A}$ to the corresponding solution field $w \in \mathcal{W}$. Here $\mathcal{A} \triangleq \{a : \Omega_a \subset \mathbb{R}^{\kappa_a} \to \mathbb{R}^{m_a}\}$ denotes the space of admissible input functions (e.g., boundary or initial conditions), and $\mathcal{W} \triangleq \{w : \Omega_w \subset \mathbb{R}^{\kappa_w} \to \mathbb{R}^{m_w}\}$ denotes the corresponding solution function space. Recent work approaches this problem using operator learning and conditional generative modeling.

Operator Learning for PDEs. Operator-learning methods such as DeepONet (Lu et al., 2019) and FNO (Li et al., 2020) learn mappings between function spaces and exhibit resolution-invariant generalization compared to CNN or physics-informed approaches (Guo et al., 2016; Karniadakis et al., 2021). While probabilistic extensions introduce uncertainty-aware formulations (Winovich et al., 2025; Bülte et al., 2025), most approaches perform direct operator regression and do not model relationships between fidelity levels as transport between solution measures.

Conditional Generative Modeling for PDEs. Generative approaches learn conditional distributions over solution fields, including conditional GANs (Goodfellow et al., 2014), diffusion models (Ho et al., 2020; Song et al., 2020), and flow matching methods (Lipman et al., 2022; Tong et al., 2023). Although these models capture uncertainty, most operate on finite-dimensional representations tied to a fixed discretization and do not explicitly leverage low-fidelity solvers as structural priors or transport references.

Functional Generative Models. Recent work extends generative modeling to function spaces, including GANO (Rahman et al., 2022), infinite-dimensional diffusion (Kerrigan et al., 2022; Yao et al., 2025), and operator-valued flow matching (Kerrigan et al., 2023; Shi et al., 2025). These approaches enable resolution-invariant probabilistic modeling, but typically assume data from a single fidelity. In contrast, our work introduces a residual transport formulation in which low-fidelity solvers define a reference measure, and generative dynamics learn probabilistic corrections toward high-fidelity solution manifolds. This perspective enables the transfer of contextual information between fidelity levels via conditional residual flows, thereby enabling multi-fidelity conditioning without retraining across resolutions.

3 METHODOLOGY

We interpret multi-fidelity operator learning as probabilistic transport between solution manifolds in function space. Given a low-fidelity solver that induces a reference measure and provides valuable inductive biases, we seek to learn an operator-valued transformation that transports this toward the high-fidelity solution measure through residual corrections. This transport perspective motivates the operator-valued flow matching formulation introduced next.

3.1 PROBABILISTIC OPERATOR FOR SOLVING THE FORWARD PROBLEM

To solve the forward problem in a probabilistic setting, we seek to approximate the conditional data measure $\nu_1(\cdot|a)$ defined on $\mathcal{B}(\mathcal{W})$ for the input function $a \sim \nu_a$. Here, $\mathcal{B}(\cdot)$ denotes the Borel σ-algebra and ν_a is a probability measure defined on $\mathcal{B}(\mathcal{A})$. Following Kerrigan et al. (2023), we define a conditional flow map $\varphi^{\zeta}_{\tau,a} : \mathcal{W} \to \mathcal{W}$ with parameters $\zeta \in \mathbb{R}^{n_\zeta}$ governed by

$$\begin{aligned} \frac{d}{d\tau}\varphi^{\zeta}_{\tau,a}[w] &= \mathcal{F}^{\zeta}_{\tau,a}\left[\varphi^{\zeta}_{\tau,a}[w]\right], \quad \tau \in [0,1], \\ \varphi^{\zeta}_{0,a}[w] &= w, \qquad w \sim \nu_0, \end{aligned}$$

where $\mathcal{F}^{\zeta}_{\tau,a} : \mathcal{W} \to \mathcal{W}$ is a vector field operator. The flow induces the push-forward measure $\eta_\tau(\cdot|a;\zeta) \triangleq [\varphi^{\zeta}_{\tau,a}]_{\#}\nu_0$ defined on $\mathcal{B}(\mathcal{W})$. The goal is to learn a vector field operator $\mathcal{F}^{\zeta}_{\tau,a}$ that induces a conditional probability path $\eta_\tau(\cdot|a;\zeta)$ such that $\eta_1(\cdot|a;\zeta) \approx \nu_1(\cdot|a)$ for all $a \sim \nu_a$, as illustrated in fig. 1.

Similar to Shi et al. (2025), we learn an operator $\mathcal{H}^{\xi}_{\tau,a}$ with parameters $\xi \in \mathbb{R}^{n_\xi}$ to approximate $\mathcal{F}^{\zeta}_{\tau,a}$ by introducing a latent conditional vector field $\tilde{\mathcal{F}}^{\zeta}_{\tau,a,z} : \mathcal{W} \to \mathcal{W}$ that induces $\eta_\tau(\cdot|a,z;\zeta)$ and minimizing the objective

$$J_{\text{PDE-COFM}}(\zeta,\xi) \triangleq \mathbb{E}_{\substack{\tau\sim\mathcal{U}[0,1],\ a\sim\nu_a,\\ z\sim\kappa(\cdot|a),\ w\sim\eta_\tau(\cdot|a,z;\zeta)}}\left[\left\|\gamma_\tau\left(\tilde{\mathcal{F}}^{\zeta}_{\tau,a,z}[w] - \mathcal{H}^{\xi}_{\tau,a}[w]\right)\right\|^2_{\mathcal{W}}\right], \tag{1}$$

such that $\mathcal{H}^{\xi}_{\tau,a}$ also induces the measure $\eta_\tau(\cdot|a;\zeta)$. Here, we introduce a conditional coupling $\kappa(\cdot|a)$ such that $z \sim \kappa(\cdot|a)$ with $z = \{w_0, w_1\}$, where $w_0 \sim \nu_0$ and $w_1 \sim \nu_1(\cdot|a)$. While independent couplings have been explored in finite-dimensional flow matching (Tong et al., 2023), here we extend this construction to the infinite-dimensional setting of the forward operator learning problem, enabling conditional probability paths directly in function space. Further, we introduce a positive time-dependent deterministic scaling γ_τ to regularize the vector field as $\tau \to 1$, encouraging accurate learning near the target data measure by scaling $\nabla_\xi J_{\text{PDE-COFM}}$. To evaluate the learning objective, we need to efficiently sample from the conditional measure $\eta_\tau(\cdot|a,z;\zeta)$. We construct $\eta_\tau(\cdot|a,z;\zeta)$ as a Gaussian measure $\mathcal{GP}(\mu_\tau(z), C_\tau(z;\zeta))$ that interpolates between the coupled samples w_0 and w_1. The mean follows a linear path,

$$\mu_\tau(z) \triangleq (1-\tau)w_0 + \tau w_1, \quad z = \{w_0, w_1\},$$

while the covariance is defined as $C_\tau(z;\zeta) \triangleq \sigma_\tau^2 C_0$, where C_0 denotes a trace-class base covariance. Assuming the reference measure $\nu_0 = \mathcal{GP}(m_0, C_0)$, the resulting path remains well-defined for finite time-dependent scaling $\sigma_\tau > 0$ (Kerrigan et al., 2023). In practice, we set $\sigma_\tau = \sigma_{\min}||w_1 - w_0||_{\mathcal{W}}$ to maintain a favorable signal-to-noise ratio when w_0 and w_1 are close. Under this Gaussian construction, the induced vector field reduces to $\tilde{\mathcal{F}}^{\zeta}_{\tau,a,z} = w_1 - w_0$ (Kerrigan et al., 2023).

3.2 RESIDUAL-AUGMENTED PROBABILISTIC OPERATOR LEARNING

Building on the probabilistic operator formulation introduced in section 3.1, we incorporate multi-fidelity structure by parameterizing transport in a residual function space defined relative to a low-fidelity solution operator. The resulting model learns probabilistic residual maps rather than full solution operators, which we refer to as FLORAL (Flow Matching Operators For Residual Augmented Learning).

Let $\mathcal{G}_{\mathrm{LF}} : \mathcal{A} \to \mathcal{W}$ denote a low-fidelity solution operator that approximates the true solution operator $\mathcal{G} : \mathcal{A} \to \mathcal{W}$. Such operators may arise from simplified physics models, coarse discretizations, or surrogate approximations and provide useful inductive bias while remaining computationally inexpensive. We express the solution operator through a residual decomposition

$$\mathcal{G}[a] = \mathcal{G}_{\mathrm{LF}}[a] + \mathcal{C}[a], \tag{2}$$

where the residual correction operator $\mathcal{C} : \mathcal{A} \to \mathcal{W}$ captures deviations between low- and high-fidelity solutions. This decomposition allows us to leverage the low-fidelity operator while focusing learning capacity on modeling the correction operator, which is often simpler than the full solution operator. Rather than learning $\mathcal{G}$ directly, FLORAL learns a probabilistic model for the residual operator $\mathcal{C}$ using the operator-valued flow matching framework.

3.3 MODEL ARCHITECTURE

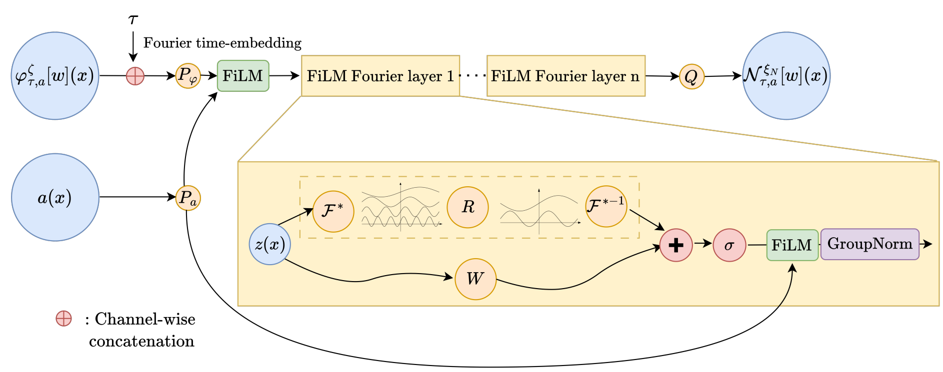

Figure 2: *Schematic of the proposed FiLMFNO architecture for modeling the nonlinear operator $\mathcal{N}^{\xi_N}_{\tau,a}$. The projection operators P_φ and P_a lift the inputs into a higher-dimensional channel space. A sequence of FiLM Fourier layers: each composed of an integral operator, nonlinear activation, and FiLM-based channel-wise modulation. The resulting representation is projected back to the target dimension by the operator Q. Within each FiLM Fourier layer, the input z is mapped to Fourier space via $\mathcal{F}^*$, transformed by a linear operator R acting on the retained low-frequency modes, and returned to physical space through the inverse transform $\mathcal{F}^{*-1}$. This output is combined with a skip connection obtained by a local linear map W, followed by nonlinear activation and FiLM conditioning.*

In this work, we decompose the model vector field operator into linear and nonlinear components (with respect to w) as $\mathcal{H}^{\xi}_{\tau,a}[w] = \mathcal{L}^{\xi_L}_{\tau,a}[w] + \mathcal{N}^{\xi_N}_{\tau,a}[w]$, where $\mathcal{L}^{\xi_L}_{\tau,a}$ and $\mathcal{N}^{\xi_N}_{\tau,a}$ are parameterized by ξ_L and ξ_N, respectively. This decomposition improves stability for long-horizon predictions, as observed in neural ODEs (Linot et al., 2023). To model the linear operator $\mathcal{L}^{\xi_L}_{\tau,a}$, we can use linear convolution layers with a stride and kernel size 1 and no activation. For the nonlinear operator $\mathcal{N}^{\xi_N}_{\tau,a}$, we propose FiLMFNO, a conditional Fourier neural operator (FNO) (Li et al., 2020) with feature-wise linear modulation (FiLM) (Perez et al., 2018) layers to effectively incorporate the input function $a \in \mathcal{A}$ as conditioning information. FiLM layers enable context-aware feature modulation by injecting auxiliary information directly into deeper network representations, providing a more effective alternative to input concatenation. The proposed architecture is illustrated in fig. 2.

4 EXPERIMENTS

Our goal is to evaluate whether probabilistic residual transport improves data efficiency, cross-resolution generalization, and uncertainty calibration compared to single-fidelity generative operators. We compare FLORAL, which combines low- and high-fidelity data, against two baselines: (i) a single-fidelity probabilistic neural operator (FLORA) trained only on high-fidelity data, as proposed by Kerrigan et al. (2023), and (ii) a deterministic low-fidelity physics-based solver. All methods are evaluated against the high-fidelity reference solution. Performance is measured using RMSE, NRMSE, and CRMSE (Takamoto et al., 2022), and probabilistic prediction performance is assessed using the mean L_2 error and the mean predictive standard deviation. All numerical experiments and datasets used are provided in `https://github.com/sahilbhola14/floral`.

4.1 PDE PROBLEM SETTING

We evaluate the proposed framework on canonical PDE operator-learning benchmarks, including 1D advection, the Burgers' equation, and Darcy flow. These problems span transport-dominated dynamics, nonlinear shock formation, and elliptic porous-media systems, enabling assessment of multi-fidelity learning. Low-fidelity models are physics-based numerical solvers with model-form uncertainty introduced through either (a) coarser spatial discretization or (b) spectral filtering of the initial condition or auxiliary model state. These settings reflect realistic scientific computing scenarios in which inexpensive simulations yield biased yet informative priors for high-fidelity solutions. Detailed PDE formulations, datasets, and numerical configurations are provided in appendix A.

4.2 MAIN EVALUATION RESULTS

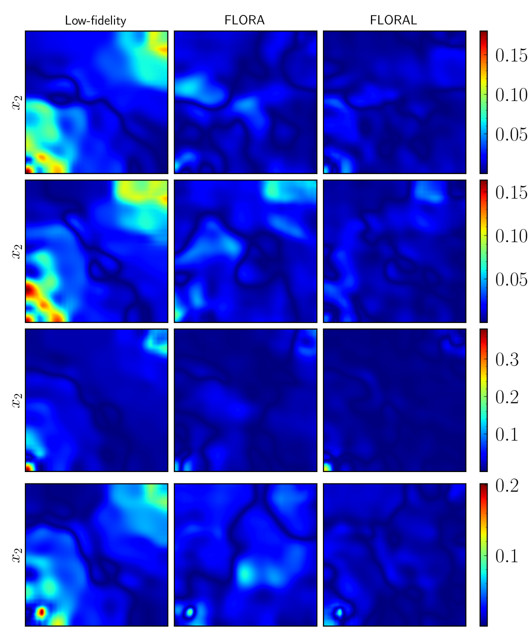

Figure 3: *Illustration of absolute error in the solution obtained using low-fidelity model (left),* FLORA *(center), and* FLORAL *(right) on unseen permeability fields for the Darcy flow. Model trained with* 5000 *high-fidelity training samples and corresponding low-fidelity training samples. To compute the mean prediction for probabilistic models,* 50 *ensembles are generated for each permeability field.*

Figure 3 illustrates the absolute error in the predicted pressure field for flow through porous media at unseen permeability fields using the low-fidelity model, FLORA, and FLORAL when trained with 5000 high-fidelity training samples. Qualitatively, we can reduce the absolute mean prediction error by leveraging low-fidelity information. Table 1 summarizes the quantitative comparison across PDE benchmarks for varying training sizes. All metrics are evaluated on 1000 unseen high-fidelity

Table 1: *Quantitative comparison across PDE benchmarks for varying training sizes. Lower is better for all metrics. For all PDEs, the metrics are computed on* 1000 *unseen samples, and for each sample,* 50 *ensemble predictions are generated to obtain the mean prediction.*

Method	**PDE**	**# Train samples**	RMSE↓	NRMSE↓	CRMSE↓
Low-fidelity (physics-based)	Advection	–	1.892×10^{-1}	3.742×10^{-1}	1.554×10^{1}
	Burgers'	–	1.115×10^{-1}	3.152×10^{-1}	1.625×10^{1}
	Darcy	–	4.595×10^{-2}	1.710×10^{-1}	4.674×10^{-1}
FLORA (single-fidelity)	Advection	500	4.964×10^{-3}	9.818×10^{-3}	8.651×10^{-1}
		1000	6.464×10^{-3}	1.278×10^{-2}	9.622×10^{-1}
		5000	2.943×10^{-3}	5.820×10^{-3}	6.770×10^{-1}
	Burgers'	500	9.369×10^{-3}	2.565×10^{-2}	1.133×10^{0}
		1000	8.172×10^{-3}	2.237×10^{-2}	2.676×10^{0}
		5000	5.351×10^{-3}	1.466×10^{-2}	1.722×10^{0}
	Darcy	500	2.777×10^{-2}	1.033×10^{-1}	1.482×10^{0}
		1000	2.394×10^{-2}	8.908×10^{-2}	1.206×10^{0}
		5000	2.512×10^{-2}	9.345×10^{-2}	1.071×10^{0}
FLORAL (ours)	Advection	500	$\mathbf{2.705 \times 10^{-3}}$	$\mathbf{5.351 \times 10^{-3}}$	$\mathbf{4.400 \times 10^{-1}}$
		1000	$\mathbf{2.261 \times 10^{-3}}$	$\mathbf{4.472 \times 10^{-3}}$	$\mathbf{2.141 \times 10^{-1}}$
		5000	$\mathbf{1.840 \times 10^{-3}}$	$\mathbf{3.639 \times 10^{-3}}$	$\mathbf{1.890 \times 10^{-1}}$
	Burgers'	500	$\mathbf{6.118 \times 10^{-3}}$	$\mathbf{1.675 \times 10^{-2}}$	$\mathbf{5.176 \times 10^{-1}}$
		1000	$\mathbf{3.803 \times 10^{-3}}$	$\mathbf{1.041 \times 10^{-2}}$	$\mathbf{4.163 \times 10^{-1}}$
		5000	$\mathbf{3.656 \times 10^{-3}}$	$\mathbf{1.001 \times 10^{-2}}$	$\mathbf{5.045 \times 10^{-1}}$
	Darcy	500	$\mathbf{1.341 \times 10^{-2}}$	$\mathbf{4.989 \times 10^{-2}}$	$\mathbf{4.151 \times 10^{-1}}$
		1000	$\mathbf{1.117 \times 10^{-2}}$	$\mathbf{4.158 \times 10^{-2}}$	$\mathbf{4.000 \times 10^{-1}}$
		5000	$\mathbf{1.538 \times 10^{-2}}$	$\mathbf{5.723 \times 10^{-2}}$	$\mathbf{4.243 \times 10^{-1}}$

samples, and for probabilistic methods, we report the mean prediction over 50 generated ensembles per input. Leveraging low-fidelity information yields consistent gains in predictive accuracy, with the largest improvements observed in low-data regimes. Across all PDEs and training sizes, FLORAL achieves substantially lower CRMSE, resulting in nearly an order-of-magnitude improvement over the single-fidelity baseline. Notably, in the Darcy flow benchmark, the single-fidelity FLORA model fails to outperform the inexpensive low-fidelity solver, highlighting the difficulty of learning accurate probabilistic operators from limited high-fidelity data alone. In contrast, FLORAL attains nearly an order-of-magnitude reduction in CRMSE relative to FLORA using only 500 high-fidelity samples. These results indicate that learning probabilistic corrections in residual function space improves sample efficiency and stabilizes operator learning compared to modeling the full solution operator directly.

Table 2 reports the mean predictive L_2 error and predictive uncertainty for 500 training samples. Across all PDE benchmarks, FLORAL consistently reduces both predictive error and predictive variance relative to the single-fidelity FLORA baseline. This joint reduction suggests that multi-fidelity conditioning improves uncertainty calibration rather than merely producing overconfident predictions. Overall, the results demonstrate that residual probabilistic transport enables more reliable and data-efficient uncertainty-aware neural operators when high-fidelity data are scarce.

4.3 PROBABILISTIC SUPER-RESOLUTION

A key advantage of operator-based models is their ability to generalize across discretizations. To study this, we consider a 1D operator learning task where the correlation between low- and high-fidelity responses depends on the input function a, following Thakur et al. (2022). Let $\mathcal{A} = \{a : \Omega_a \subset \mathbb{R} \to \mathbb{R}\}$ and $\mathcal{W} = \{w : \Omega_w \subset \mathbb{R} \to \mathbb{R}\}$. The low- and high-fidelity operators $\mathcal{G}_{\text{LF}}, \mathcal{G}_{\text{HF}}$:

Table 2: *Mean predictive error and predictive uncertainty across PDE benchmarks for* 500 *training samples. Lower is better for both metrics. For all PDEs, the metrics are computed on* 1000 *unseen samples, and for each sample,* 50 *ensemble predictions are generated to obtain the mean prediction and the predictive uncertainty.*

Method	**PDE**	**Mean predictive L2 error** ↓	**Mean predictive uncertainty** ↓
FLORA (single-fidelity)	Advection	5.294×10^{-1}	5.851×10^{-3}
	Burgers'	8.797×10^{-1}	4.350×10^{-3}
	Darcy	3.385×10^{0}	1.277×10^{-2}
FLORAL (ours)	Advection	$\mathbf{2.815 \times 10^{-1}}$	$\mathbf{2.734 \times 10^{-3}}$
	Burgers'	$\mathbf{5.931 \times 10^{-1}}$	$\mathbf{2.677 \times 10^{-3}}$
	Darcy	$\mathbf{1.612 \times 10^{0}}$	$\mathbf{5.352 \times 10^{-3}}$

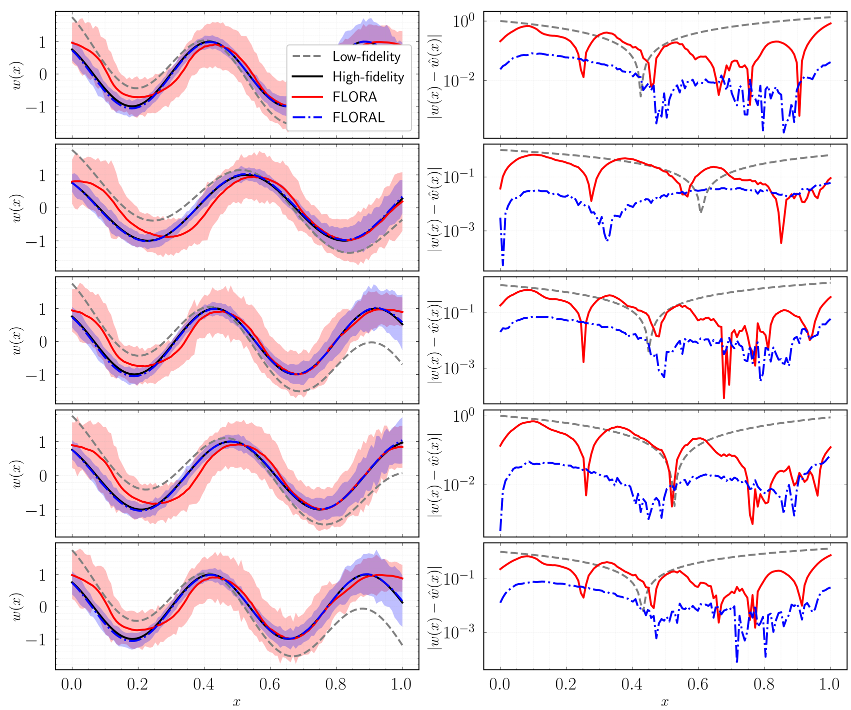

Figure 4: *Samples of the generated high-fidelity function (left) using* FLORA *(——) and* FLORAL *(-·-·-) and the absolute error (right) for the 1D problem with input function correlation. The shaded region indicates* ± 10 *standard deviations about the mean for both models, computed using* 100 *generated ensembles. Model trained with* 10 *high-fidelity training samples using* 8 *discretization points, and inference is performed on* 128 *points, that is, a* $16\times$ *resolution increase.*

$\mathcal{A} \to \mathcal{W}$ are defined as

$$\mathcal{G}_{\text{LF}}[a](x) = \sin(a(x)) + x - 0.25\, a(x),$$
$$\mathcal{G}_{\text{HF}}[a](x) = \sin(a(x)),$$

where $a(x) = kx - 4$, $k \sim \mathcal{U}(10, 14)$, and $\Omega_a = \Omega_w = [0, 1]$. Models are trained at coarse resolutions $N_x \in \{8, 16, 32, 64\}$ and evaluated on 128 points to assess super-resolution. Figure 4 illustrates samples of the generated high-fidelity function using FLORA and FLORAL along with the absolute error when trained with 8 spatial points. As shown, by leveraging the low-fidelity information, FLORAL can generate high-fidelity function samples that closely match the true high-fidelity function with high confidence, even when trained on a significantly coarser resolution. Quantitatively, as shown in fig. 5, FLORAL consistently outperforms FLORA for cross-resolution tasks, achieving nearly an order-of-magnitude accuracy gain for the $16\times$ resolution increase.

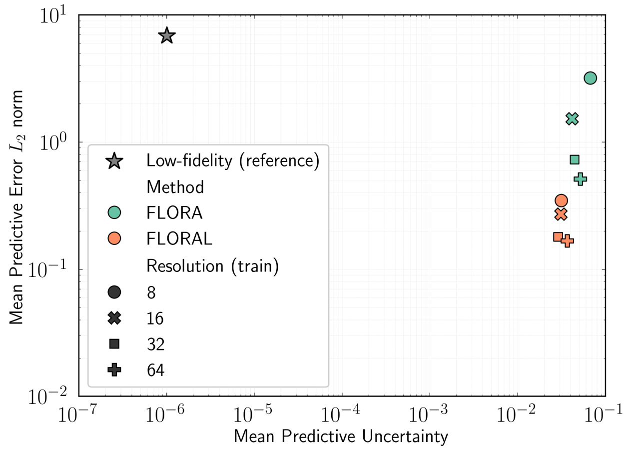

Figure 5: *Mean predictive error vs. predictive uncertainty for varying training resolutions for the 1D problem with input function correlation. Predictive uncertainty is the mean standard deviation of generated samples. The model is trained with* 10 *high-fidelity samples. To compute the statistics,* 1000 *unseen high-fidelity function samples are used, and for each sample,* 100 *ensembles are generated. Low-fidelity uncertainty is set to* 10^{-6} *for visualization since the model is deterministic.*

5 CONCLUSION

We introduced FLORAL, a probabilistic neural operator framework that formulates multi-fidelity learning as residual transport between solution manifolds. To enable probabilistic modeling in function space, we formulate the transport dynamics within a flow matching framework using conditional neural operators. To integrate multiple information sources, we introduce a low-fidelity reference measure that provides a structured prior and learn residual transport to the high-fidelity solution manifold. By learning conditional flow dynamics in residual function space, the proposed approach improves data efficiency, enables cross-resolution generalization, and produces calibrated uncertainty estimates while maintaining resolution-invariant inference. Across canonical PDE benchmarks, incorporating low-fidelity structure as a reference measure consistently improves predictive performance in data-limited regimes where single-fidelity generative operators struggle.

Beyond quantitative performance, the results suggest several broader perspectives for probabilistic operator learning. Modeling corrections in residual function space provides a structured alternative to learning full solution distributions, potentially simplifying the geometry of the learning problem. Similarly, conditioning the flow on the input function offers a flexible mechanism for contextual adaptation across discretizations without explicitly tying the model to a fixed resolution. Finally, structured vector-field parameterizations and time-dependent regularization appear to support stable training of operator-valued flows in practice.

While this work focuses on benchmark PDE problems, the residual transport formulation provides a general framework for probabilistic scientific modeling with heterogeneous fidelity sources. Future work includes systematic architectural studies, extending the framework to settings with mismatched input representations, and developing theoretical analyses of the observed gains in sample efficiency.

REFERENCES

Steven L Brunton, Bernd R Noack, and Petros Koumoutsakos. Machine learning for fluid mechanics. *Annual review of fluid mechanics*, 52(1):477–508, 2020.

Christopher Bülte, Philipp Scholl, and Gitta Kutyniok. Probabilistic neural operators for functional uncertainty quantification. *arXiv preprint arXiv:2502.12902*, 2025.

Ricky TQ Chen, Yulia Rubanova, Jesse Bettencourt, and David K Duvenaud. Neural ordinary differential equations. *Advances in neural information processing systems*, 31, 2018.

Jacob Gardner, Geoff Pleiss, Kilian Q Weinberger, David Bindel, and Andrew G Wilson. Gpytorch: Blackbox matrix-matrix gaussian process inference with gpu acceleration. *Advances in neural information processing systems*, 31, 2018.

Ian J Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. *Advances in neural information processing systems*, 27, 2014.

Xiaoxiao Guo, Wei Li, and Francesco Iorio. Convolutional neural networks for steady flow approximation. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, pp. 481–490, 2016.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020.

George Em Karniadakis, Ioannis G Kevrekidis, Lu Lu, Paris Perdikaris, Sifan Wang, and Liu Yang. Physics-informed machine learning. *Nature Reviews Physics*, 3(6):422–440, 2021.

Gavin Kerrigan, Justin Ley, and Padhraic Smyth. Diffusion generative models in infinite dimensions. *arXiv preprint arXiv:2212.00886*, 2022.

Gavin Kerrigan, Giosue Migliorini, and Padhraic Smyth. Functional flow matching. *arXiv preprint arXiv:2305.17209*, 2023.

Jean Kossaifi, Nikola Kovachki, Zongyi Li, David Pitt, Miguel Liu-Schiaffini, Valentin Duruisseaux, Robert Joseph George, Boris Bonev, Kamyar Azizzadenesheli, Julius Berner, and Anima Anandkumar. A library for learning neural operators. *arXiv preprint arXiv:2412.10354*, 2025.

Zongyi Li, Nikola Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Fourier neural operator for parametric partial differential equations. *arXiv preprint arXiv:2010.08895*, 2020.

Alec J Linot, Joshua W Burby, Qi Tang, Prasanna Balaprakash, Michael D Graham, and Romit Maulik. Stabilized neural ordinary differential equations for long-time forecasting of dynamical systems. *Journal of Computational Physics*, 474:111838, 2023.

Yaron Lipman, Ricky TQ Chen, Heli Ben-Hamu, Maximilian Nickel, and Matt Le. Flow matching for generative modeling. *arXiv preprint arXiv:2210.02747*, 2022.

Lu Lu, Pengzhan Jin, and George Em Karniadakis. Deeponet: Learning nonlinear operators for identifying differential equations based on the universal approximation theorem of operators. *arXiv preprint arXiv:1910.03193*, 2019.

Ethan Perez, Florian Strub, Harm De Vries, Vincent Dumoulin, and Aaron Courville. Film: Visual reasoning with a general conditioning layer. In *Proceedings of the AAAI conference on artificial intelligence*, volume 32, 2018.

Md Ashiqur Rahman, Manuel A Florez, Anima Anandkumar, Zachary E Ross, and Kamyar Azizzadenesheli. Generative adversarial neural operators. *arXiv preprint arXiv:2205.03017*, 2022.

Yaozhong Shi, Zachary E Ross, Domniki Asimaki, and Kamyar Azizzadenesheli. Stochastic process learning via operator flow matching. *arXiv preprint arXiv:2501.04126*, 2025.

Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. *arXiv preprint arXiv:2010.02502*, 2020.

Makoto Takamoto, Timothy Praditia, Raphael Leiteritz, Daniel MacKinlay, Francesco Alesiani, Dirk Pflüger, and Mathias Niepert. Pdebench: An extensive benchmark for scientific machine learning. *Advances in Neural Information Processing Systems*, 35:1596–1611, 2022.

Akshay Thakur, Tapas Tripura, and Souvik Chakraborty. Multi-fidelity wavelet neural operator with application to uncertainty quantification. *arXiv preprint arXiv:2208.05606*, 2022.

Alexander Tong, Kilian Fatras, Nikolay Malkin, Guillaume Huguet, Yanlei Zhang, Jarrid Rector-Brooks, Guy Wolf, and Yoshua Bengio. Improving and generalizing flow-based generative models with minibatch optimal transport. *arXiv preprint arXiv:2302.00482*, 2023.

Nick Winovich, Mitchell Daneker, Lu Lu, and Guang Lin. Active operator learning with predictive uncertainty quantification for partial differential equations. *arXiv preprint arXiv:2503.03178*, 2025.

Jiachen Yao, Abbas Mammadov, Julius Berner, Gavin Kerrigan, Jong Chul Ye, Kamyar Azizzadenesheli, and Anima Anandkumar. Guided diffusion sampling on function spaces with applications to pdes. *arXiv preprint arXiv:2505.17004*, 2025.

Yinhao Zhu and Nicholas Zabaras. Bayesian deep convolutional encoder–decoder networks for surrogate modeling and uncertainty quantification. *Journal of Computational Physics*, 366:415–447, 2018.

A DETAILED PROBLEM DESCRIPTION

Advection equation. We consider the 1D linear advection equation

$$\partial_t u(x,t) + \beta\, \partial_x u(x,t) = 0, \quad x \in [0,1],\ t \in [0,1],$$

which models the transport of a conserved scalar field with constant velocity β. Following Takamoto et al. (2022), periodic boundary conditions are imposed, and the initial condition is constructed as

$$u(x,0) = \sum_{i=1}^{N} A_i \sin(k_i x + \phi_i), \tag{3}$$

where $k_i = 2\pi n_i$, $n_i \sim \mathcal{U}(\{1,\ldots,n_{\max}\})$, $A_i \sim \mathcal{U}(0,1)$, and $\phi_i \sim \mathcal{U}(0,2\pi)$. We use $n_{\max} = 8$, $N = 2$, and $\beta = 0.05$. After evaluating eq. (3), we apply absolute value, random sign flip, and windowing with probabilities 0.1, 0.5, and 0.1, respectively. The PDE is solved using a first-order upwind finite-difference scheme with forward Euler time integration.

To generate high-fidelity (HF) data, we use $N_x = 128$ spatial grid points and $N_t = 128$ time steps. Low-fidelity (LF) solutions are defined on the same spatio-temporal grid but incorporate model-form error through perturbed initial conditions. Specifically, the LF initial condition is constructed by spectrally filtering the HF initial condition. Let $\hat{u}_{\mathrm{HF}}(k)$, $k = 0,\ldots,K$, denote the Fourier coefficients of $u(x,0)$. We define the LF spectrum as

$$\hat{u}_{\mathrm{LF}}(k) = s_A \hat{u}_{\mathrm{HF}}(k)\, m_k\, e^{-\gamma k/K},$$

where m_k retains $\max(2, \lfloor f_{\text{keep}} K \rfloor)$ low-frequency modes, $e^{-\gamma k/K}$ applies mode-dependent damping, and s_A scales the amplitude. We use $f_{\text{keep}} = 0.4$, $\gamma = 4$, and $s_A = 0.8$ across all experiments. The LF initial condition is obtained via inverse Fourier transform of $\hat{u}_{\mathrm{LF}}(k)$. An illustration of the high- and low-fidelity initial conditions is shown in fig. 6.

Burgers' equation. We consider the 1D viscous Burgers' equation

$$\partial_t u(x,t) + u(x,t)\partial_x u(x,t) = \nu\, \partial_{xx} u(x,t), \quad x \in [0,1],\ t \in [0,0.2],$$

which models nonlinear advection-diffusion dynamics. The viscosity is fixed to $\nu = 0.01$. Ground-truth solutions are generated using a second-order upwind scheme for advection, second-order central differences for diffusion, and a third-order SSP Runge-Kutta integrator.

High-fidelity (HF) data use $N_{x;\mathrm{HF}} = 128$ and $N_{t;\mathrm{HF}} = 128$. Initial conditions are obtained similar to the advection equation case, with $f_{\text{keep}} = 0.6$. Low-fidelity (LF) data introduce additional discretization error via a coarser grid with $N_{x;\mathrm{LF}} = 64$ and $N_{t;\mathrm{LF}} = 64$.

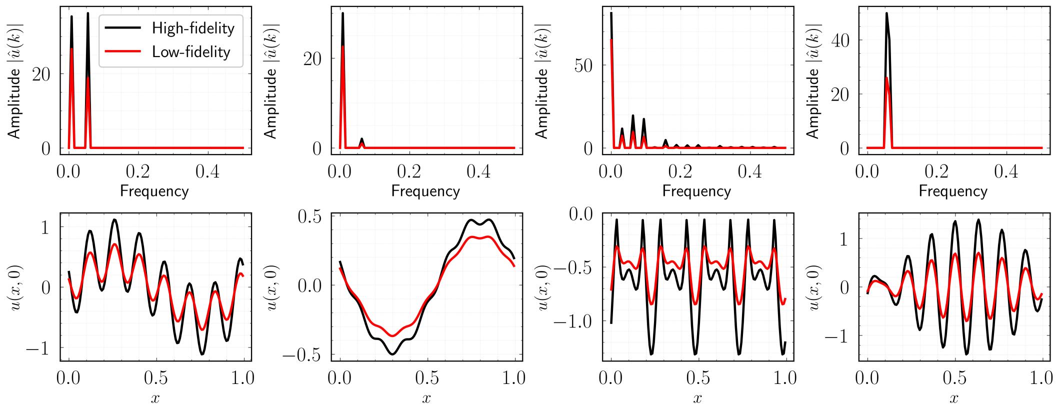

Figure 6: *Comparison of high- (——) and low-fidelity (——) initial conditions in the spectral space (top row) and the corresponding physical space state (bottom row) for the 1D advection equation. The low-fidelity initial condition is obtained by applying a spectral filter, followed by amplitude distortion, to the high-fidelity initial condition to remove high-frequency components and introduce an amplitude bias.*

Flow through porous media. We consider the flow through porous media characterized by Darcy's law that relates pressure $p(\mathbf{x})$ and permeability $K(\mathbf{x})$ on $\Omega = [0,1]^2$:

$$\mathbf{u}(\mathbf{x}) = -K(\mathbf{x})\nabla p(\mathbf{x}), \tag{4a}$$

$$\nabla \cdot \mathbf{u}(\mathbf{x}) = f_s(\mathbf{x}), \tag{4b}$$

$$\mathbf{u}(\mathbf{x}) \cdot \hat{n}(\mathbf{x}) = 0, \quad \mathbf{x} \in \partial\Omega, \tag{4c}$$

$$\int_\Omega p(\mathbf{x})\, d\mathbf{x} = 0, \tag{4d}$$

where $\hat{n}$ denotes the outward unit normal. Following Zhu & Zabaras (2018), the source term f_s models an injection well at the bottom-left and a production well at the top-right corner, with rate $r = 50$ and size $s = 0.125$.

The log-permeability is modeled as a Gaussian random field $K(\mathbf{x}) = \exp(G(\mathbf{x}))$, where G has zero mean and covariance $k(x, x') = \exp(-\|x - x'\|_2)/l$ with $l = 0.1$. Dimensionality reduction is performed via a truncated Karhunen–Loève expansion $G(\mathbf{x}) = \sum_{i=1}^{q} \sqrt{\lambda_i}\phi_i(\mathbf{x})z_i$, $z_i \sim \mathcal{N}(0,1)$.

The governing equations are solved using second-order central finite differences, enforcing boundary conditions via KKT conditions. High-fidelity (HF) data use a 128×128 grid and $q = 128$ modes. Low-fidelity (LF) data truncate the expansion to $q = 64$ and use a coarser 32×32 grid, introducing both model-form and discretization error.

B Training Details

For each experiment performed, we require samples from a Gaussian process measure. In our work, we consider a zero-mean (that is, $m_0 = 0$) Gaussian process parameterized by a Matern covariance kernel with smoothness factor $\nu = 0.5$ and length scale of 10^{-3}. We implement this using the `GPyTorch` library (Gardner et al., 2018). To model the vector fields using an operator, we propose a conditional Fourier neural operator (Li et al., 2020) leveraging feature-wise linear modulation (FiLM) (Perez et al., 2018), called FiLMFNO, using the `neuraloperator` library (Kossaifi et al., 2025). The hyperparameters for FiLMFNO have the same meaning as described in the `neuraloperator` library. For all experiments, we have considered 4 Fourier layers, 64 hidden channels, 4 lifting channel ratio, and 4 projection channel ratio. To compute the coefficients of the FiLM layers, we use fully connected neural networks with 3 hidden layers of 64 neurons each and SiLU activations. To train the model, we use Adam optimizer with a learning rate of 10^{-3} and a weight decay of 10^{-4}. We use an exponential learning rate scheduler with a decay rate of 0.99. To integrate the vector-fields, we use `dopri5` provided in `torchdiffeq` (Chen et al., 2018) with

$\texttt{atol} = \texttt{rtol} = 10^{-5}$. For all experiments, the training configurations are summarized in tables 3 to 6.

Table 3: *Training configurations for the 1D advection equation. Training and inference share the same spatial–temporal resolution, and the number of Fourier modes is identical across spatial and temporal dimensions.*

Train→Infer Resolution	**Epochs**	**Train Size**	**Val Size**	**Batch Size**	**Fourier Modes**
128 → 128	500	500 1,000 5,000	1,000	16 64 128	64

Table 4: *Training configurations for the 1D Burgers' equation. Training and inference share the same spatial–temporal resolution, and the number of Fourier modes is identical across dimensions.*

Train→Infer Resolution	**Epochs**	**Train Size**	**Val Size**	**Batch Size**	**Fourier Modes**
128 → 128	500	500 1,000 5,000	1,000	16 64 128	64

Table 5: *Training configurations for the Darcy flow problem. Training and inference share the same resolution, and the number of Fourier modes is identical across spatial dimensions.*

Train→Infer Resolution	**Epochs**	**Train Size**	**Val Size**	**Batch Size**	**Fourier Modes**
128 → 128	500	500 1,000 5,000	1,000	16 64 128	64

Table 6: *Training configurations for the 1D input-correlation experiment. Both equal-resolution and multi-resolution settings are reported.*

Train→Infer Resolution	**Epochs**	**Train Size**	**Val Size**	**Batch Size**	**Fourier Modes**
128 → 128	300	10 50 500 1,000	1,000	2 4 16 64	64
8 → 128 16 → 128 32 → 128 64 → 128	300	10	1,000	2	4 8 16 32

C ADDITIONAL RESULTS

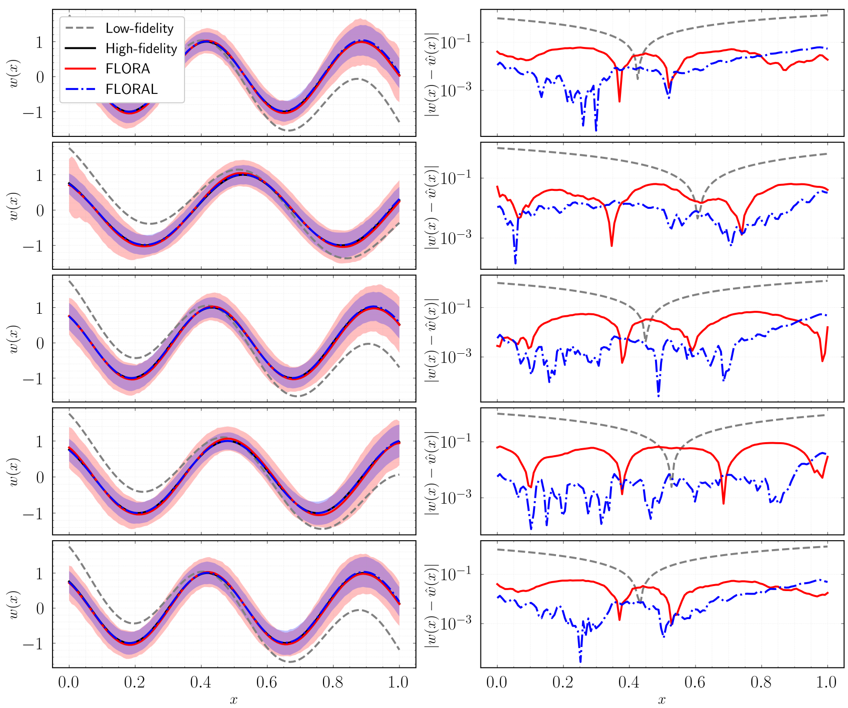

Figure 7: *Samples of the generated high-fidelity function (left) using* FLORA *(—) and* FLORAL *(-·-·) and the absolute error (right) for the 1D problem with input function correlation. The shaded region indicates* ± 10 *standard deviations about the mean for both models, computed using* 100 *generated ensembles. Model trained with* 10 *high-fidelity and corresponding low-fidelity training samples for a training and inference resolution of* 128 *discretization points. Samples are generated on the same input function as in fig. 4.*

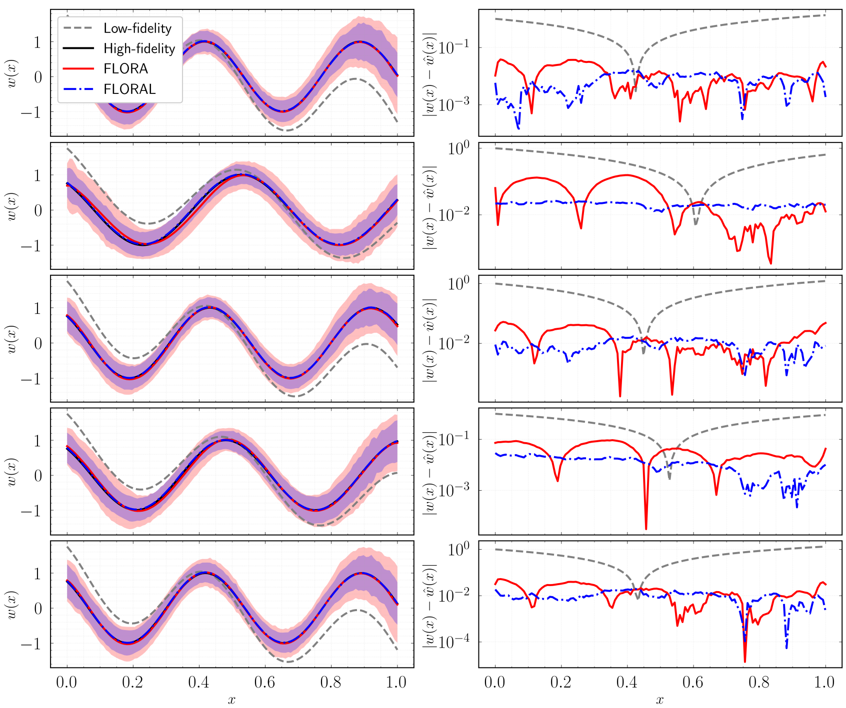

Figure 8: *Samples of the generated high-fidelity function (left) using* FLORA *(—) and* FLORAL *(-·-·) and the absolute error (right) for the 1D problem with input function correlation. The shaded region indicates* ± 10 *standard deviations about the mean for both models, computed using* 100 *generated ensembles. Model trained with* 10 *high-fidelity training samples using* 64 *discretization points, and inference is performed on* 128 *points, that is, a* $2\times$ *resolution increase. Samples are generated on the same input function as in fig. 4.*

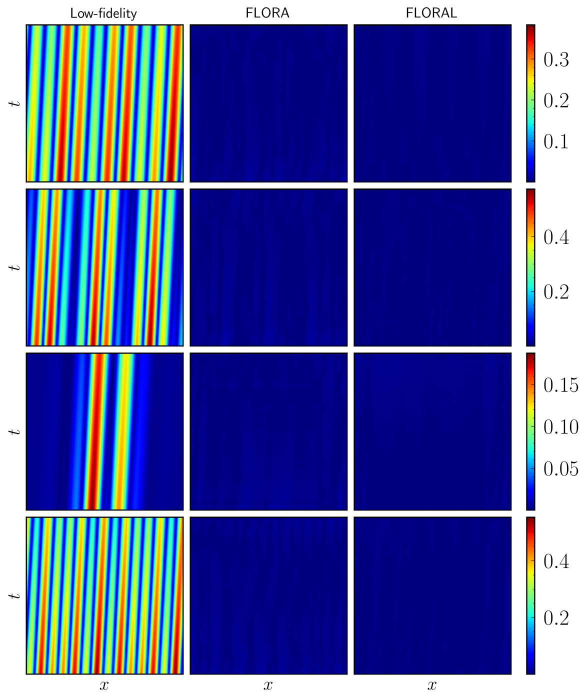

Figure 9: *Illustration of absolute error in the solution obtained using low-fidelity model (left),* FLORA *(center), and* FLORAL *(right) on unseen initial conditions for the 1D advection equation. Model trained with* 500 *high-fidelity and corresponding low-fidelity training samples. To compute the mean prediction for probabilistic models,* 50 *ensembles are generated for each unseen initial condition.*

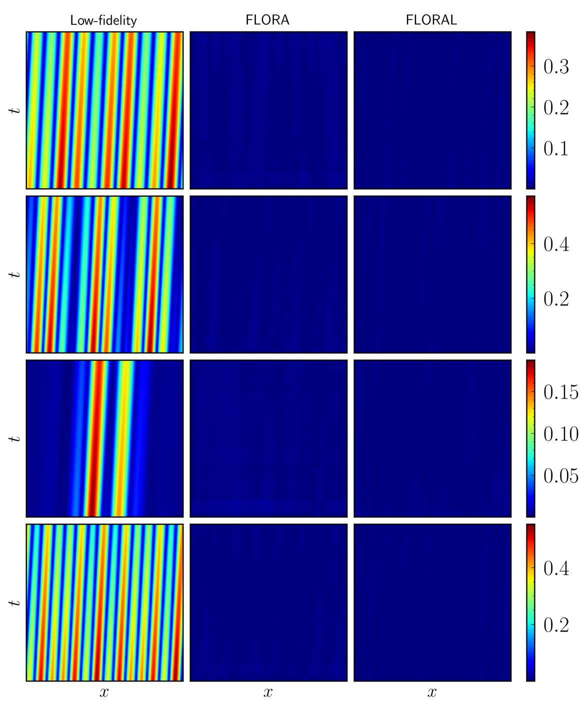

Figure 10: *Illustration of absolute error in the solution obtained using low-fidelity model (left),* FLORA *(center), and* FLORAL *(right) on unseen initial conditions for the 1D advection equation. Model trained with* 5000 *high-fidelity and corresponding low-fidelity training samples. To compute the mean prediction for probabilistic models,* 50 *ensembles are generated for each unseen initial condition. Samples are generated on the same initial conditions as in fig. 9.*

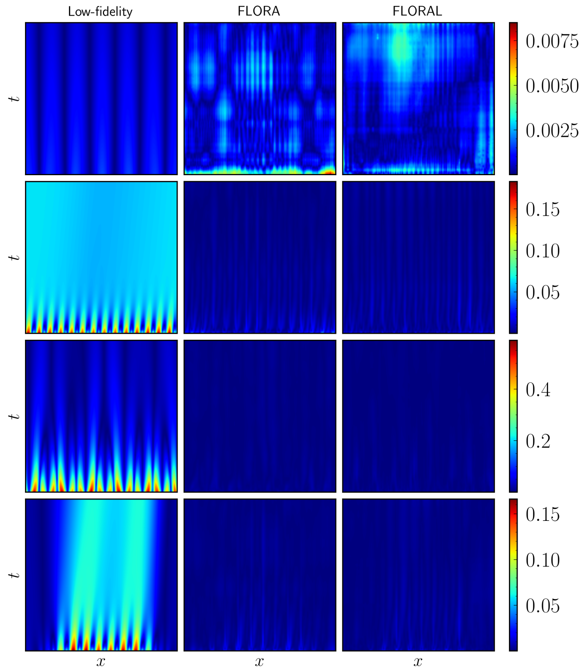

Figure 11: *Illustration of absolute error in the solution obtained using low-fidelity model (left),* FLORA *(center), and* FLORAL *(right) on unseen initial conditions for the 1D Burgers' equation. Model trained with* 500 *high-fidelity training samples and corresponding low-fidelity training samples. To compute the mean prediction for probabilistic models,* 50 *ensembles are generated for each unseen initial condition.*

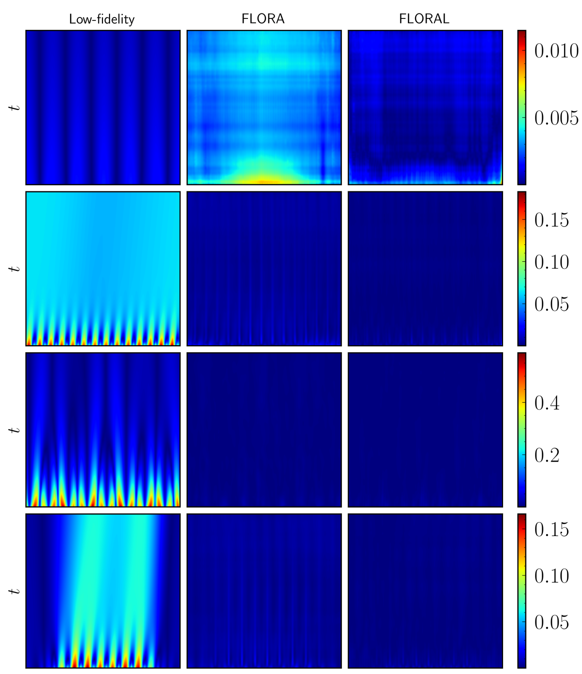

Figure 12: *Illustration of absolute error in the solution obtained using low-fidelity model (left),* FLORA *(center), and* FLORAL *(right) on unseen initial conditions for the 1D Burgers' equation. Model trained with* 5000 *high-fidelity training samples and corresponding low-fidelity training samples. To compute the mean prediction for probabilistic models,* 50 *ensembles are generated for each unseen initial condition. Samples are generated on the same initial conditions as in fig. 11.*

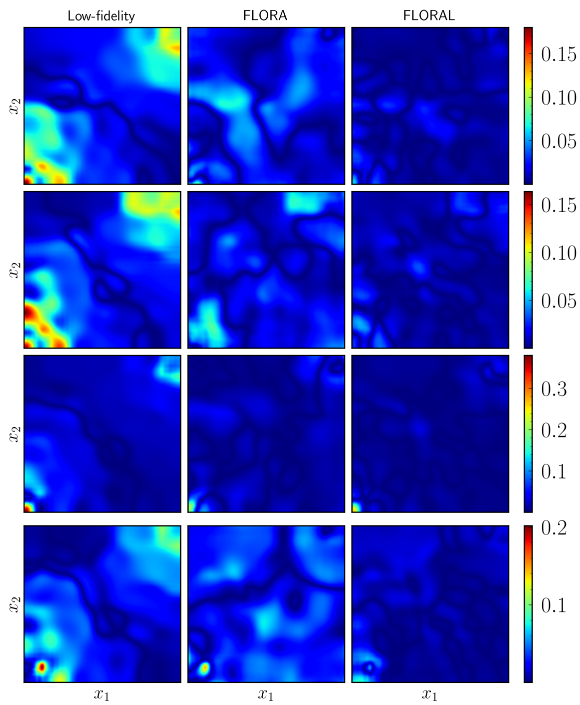

Figure 13: *Illustration of absolute error in the solution obtained using low-fidelity model (left),* FLORA *(center), and* FLORAL *(right) on unseen permeability fields for the Darcy flow. Model trained with* 500 *high-fidelity training samples and corresponding low-fidelity training samples. To compute the mean prediction for probabilistic models,* 50 *ensembles are generated for each unseen permeability field. Samples are generated on the same permeability fields conditions as in fig. 3.*