Ocean Engineering 343 (2026) 123314

Contents lists available at ScienceDirect

Ocean Engineering

journal homepage: www.elsevier.com/locate/oceaneng

Research paper

Ensemble adaptive gated multi-fidelity neural network for Bayesian optimization: Application to hydrofoil design

Passakorn Paladaechanan [a], Yao Zhong [b], Maokun Ye [a], Decheng Wan [a,*], Moustafa Abdel-Maksoud [c]

[a] Computational Marine Hydrodynamics Laboratory (CMHL), School of Ocean and Civil Engineering, Shanghai Jiao Tong University, Shanghai, China
[b] Zhongnan Engineering Corporation Limited, Changsha, China
[c] Hamburg University of Technology (TUHH), Institute for Fluid Dynamics and Ship Theory (M8), Hamburg, Germany

ARTICLE INFO

Keywords:
Multi-fidelity surrogate modeling
Bayesian optimization
Deep learning
Mixture of experts
Uncertainty quantification
Hydrofoil optimization
Computational fluid dynamics

ABSTRACT

High-fidelity computational fluid dynamics (CFD) simulations provide critical predictive accuracy in marine and ocean engineering design; however, their substantial computational expense often renders direct optimization infeasible. To alleviate this limitation, surrogate models approximate expensive objective functions from a finite set of observations, thereby enabling more tractable design exploration and optimization. Our objective is to build a novel, general-purpose multi-fidelity surrogate modeling approach that integrates seamlessly into a Bayesian optimization framework and remains robust under sparse high-fidelity data. We propose an adaptive gated multi-fidelity neural network (AGMF-Net), which incorporates three specialized expert subnetworks—linear, nonlinear, and residual—combined through a deep Mixture-of-Experts gating network that dynamically adjusts their contributions based on the input. To improve predictive uncertainty estimation, we ensemble multiple independently initialized AGMF-Net instances and use the resulting variance to guide sampling decisions. We embed this surrogate into a Bayesian optimization workflow driven by the logarithmic expected improvement acquisition function, which balances exploration and exploitation while maintaining numerical stability. We evaluated the proposed method against co-Kriging and the multi-fidelity neural network baseline on benchmark functions. AGMF-Net achieved higher initial predictive accuracy, rapidly converged to global optima, and maintained lower mean absolute relative error during optimization iterations. Finally, we applied the framework to a hydrofoil design optimization. The model successfully identified a subtle camber modification that improved the lift-to-drag ratio by 41.6 % compared to the baseline geometry, demonstrating that AGMF-Net can accelerate CFD-driven hydrodynamic design scenarios that combine sparse high-fidelity data with cheaper simulations. These results highlight the potential of adaptive gating and ensemble uncertainty quantification to accelerate design exploration and improve solution quality when only limited high-fidelity evaluations are feasible.

1. Introduction

High-fidelity CFD simulations have become indispensable in ocean and marine engineering, where designers must resolve turbulent and multi-scale flow phenomena around hulls, propellers, and hydrofoils. Although CFD delivers reliable hydrodynamic prediction, its high computational cost restricts direct use in iterative design optimization. In response, data-efficient surrogate models have emerged as a critical tool for accelerating high-cost simulation-based design loops. By providing fast approximations of expensive simulators, surrogates enable practitioners to explore design spaces and assess system performance with dramatically fewer expensive evaluations (Forrester et al., 2008; Peherstorfer et al., 2018a).

This efficiency is particularly important in Bayesian optimization (BO), which iteratively uses a probabilistic surrogate to balance exploration and exploitation of a black-box objective (Shahriari et al., 2015; Jones et al., 1998). BO relies on a probabilistic surrogate (commonly a Gaussian Process in classical approaches) to predict outcomes and quantify uncertainty, guiding the selection of new experiments via acquisition functions. Integrating cheaper, lower fidelity simulations or

* Corresponding author.
E-mail address: dcwan@sjtu.edu.cn (D. Wan).

https://doi.org/10.1016/j.oceaneng.2025.123314
Received 15 July 2025; Received in revised form 24 October 2025; Accepted 25 October 2025

models into this process— a concept known as multi-fidelity Bayesian optimization (MFBO)—can further reduce the overall optimization cost by allowing occasional queries of fast, approximate information sources in lieu of expensive high-fidelity ones (Peherstorfer et al., 2018a). In recent years, MFBO frameworks have demonstrated substantial speed-ups in applications ranging from materials discovery (Sabanza et al., 2024; Fare et al., 2022) to aerospace design (Mukhopadhaya et al., 2020) by optimally allocating resources across fidelity levels. These advances underscore the importance of developing surrogate modeling techniques that are not only accurate and data-efficient, but also capable of leveraging multiple fidelities and providing reliable uncertainty estimates to drive decision-making.

Multi-fidelity surrogate modeling has accordingly become a vibrant research area, aiming to fuse information from inexpensive low-fidelity (LF) sources and sparse high-fidelity (HF) data to build improved predictive models. Early approaches in this domain were largely based on Gaussian process regression; the seminal co-kriging framework of Kennedy and O'Hagan (2000) (Kennedy and O'Hagan, 2000) introduced a hierarchical GP model to correct LF predictions using HF data. Such GP-based multi-fidelity surrogates (and their extensions) can provide principled uncertainty quantification and have shown success in various engineering problems (Le et al., 2014; Ciarlatani and Gorlé, 2025; Novais et al., 2024). However, Gaussian process (GP) models face limitations when dealing with high-dimensional inputs or highly nonlinear system responses, often suffering the curse of dimensionality and loss of accuracy in complex scenarios (Williams and Rasmussen, 2006). For instance, studies have noted that classical multi-fidelity GP tends to degrade in performance on large-scale nonlinear problems (Perdikaris et al., 2017). This has motivated a shift toward deep learning methods, which can learn rich representations and handle large data with complex patterns (Raissi et al., 2019). Recent deep multi-fidelity surrogates exploit neural networks to capture correlations across fidelity levels and improve scalability (Perdikaris et al., 2017). Recurrent architectures such as Long Short-Term Memory (LSTM) networks have been used to automatically detect cross-fidelity features and achieve accurate multi-fidelity regression (Conti et al., 2023). Likewise, advanced convolutional neural nets (Halder et al., 2022) and neural operators (Li et al., 2020; Lu et al., 2021) have been explored to learn mappings between coarse and fine solution fields in physics problems.

Notably, Meng and Karniadakis (2020) (Meng and Karniadakis, 2020) proposed a composite neural network that couples subnetworks across fidelity levels to model both linear and nonlinear correlations. Their architecture integrates a low-fidelity encoder with two higher-fidelity branches—one linear, one nonlinear—capturing multi-scale relationships in a unified model. This concept of decomposing fidelity interactions has inspired further advances in deep multi-fidelity learning. For instance, Lu et al. (2020) (Lu et al., 2020) introduced a residual-learning framework for inferring material properties from indentation data, where the neural network explicitly learned the discrepancy between a low-fidelity analytical solution and high-fidelity finite element simulations. Building on such ideas, Zhan et al. (2024) (Zhan et al., 2024) proposed Ada2MF, a dual-adaptive multi-fidelity model for turbulent wake flow prediction. By integrating residual learning with learnable gating and adaptive loss weighting, Ada2MF improved robustness and mitigated negative transfer in regimes with limited high-fidelity data. In parallel, attention-based strategies have also emerged. Cheng et al. (2024) (Cheng et al., 2025) developed MF-Net, an architecture that uses self-attention to fuse multi-source low-fidelity features with sparse high-fidelity data in an end-to-end manner. Their model achieved state-of-the-art accuracy in a welding mechanics case study. Taken together, these developments highlight a clear trend: robust multi-fidelity models increasingly rely on mechanisms that adaptively gate, weight, or attend to each fidelity level—leveraging correlations while guarding against misleading signals.

Despite progress, two major challenges remain: (1) negative transfer due to naively mixing inconsistent fidelity sources, and (2) the lack of adaptive architectures that generalize across domains. Addressing these, Zhan et al. (2024) proposed Ada2MF, a dual-adaptive network combining an Adaptive Multi-fidelity (AMF) module and Adaptive Fast Weighting (AFW). AMF blends three experts—linear, nonlinear, and residual— via learnable weights, while AFW adjusts loss contributions from each fidelity dynamically.

This line of research connects with the Mixture of Experts (MoE) paradigm (Jacobs et al., 1991), where a gating network selects among expert sub-models. Modern MoE approaches, such as sparsely-gated models (Shazeer et al., 2017), scale this idea to billions of parameters with minimal computational cost. Applying this gating philosophy to surrogate modeling promises robustness by emphasizing trustworthy experts depending on the input.

Another key component in BO is uncertainty quantification (UQ). While Gaussian processes offer built-in UQ, they lack scalability. Bayesian neural networks (BNNs) (Neal, 2012) provide a principled alternative but are difficult to train. Lakshminarayanan et al. (2017) proposed deep ensembles as a practical and competitive method for UQ, outperforming many BNNs in calibration and robustness.

In multi-fidelity contexts, uncertainty-aware surrogates remain scarce. Some researchers extended BNNs to multi-fidelity modeling (Meng et al., 2021), but training complexity remains a bottleneck. Thus, integrating scalable UQ into expressive multi-fidelity architectures is a critical research frontier.

Surrogate modeling in hydrodynamics and fluid mechanics further motivates our work. Optimizing hydrofoils, propellers, and marine structures requires resolving turbulent, multi-scale physics—often needing unsteady Reynolds-averaged Navier–Stokes (URANS) simulations (Li et al., 2023; Bonfiglio et al., 2018). Recent works have combined XFOIL with CFD for multi-fidelity optimization of airfoils (Aye et al., 2023), but most rely on traditional co-kriging or basic neural nets, lacking adaptive fusion or robust UQ.

To address these gaps, we propose AGMF-Net, an Adaptive Gated Multi-Fidelity neural network. AGMF-Net builds on Ada2MF (Zhan et al., 2024) and enhances it in two ways: (1) using a deep Mixture-of-Experts (DMoE) gating sub-network to assign context-aware weights to three specialized experts (linear, nonlinear, residual), and (2) employing deep ensembles for uncertainty-aware predictions. This results in a surrogate that is accurate, general, and robust to fidelity inconsistency.

The objective of this study is to develop a novel, general-purpose multi-fidelity surrogate model integrated within a Bayesian optimization framework that can accurately predict expensive functions using limited high-fidelity observations.

This study introduces AGMF-Net, a novel surrogate architecture that integrates deep gated expert fusion with ensemble-based uncertainty quantification, enabling accurate and robust multi-fidelity learning under limited high-fidelity data. We incorporate AGMF-Net into a Bayesian optimization loop and evaluate its effectiveness on challenging mathematical benchmark functions. To demonstrate its practical utility, we then apply it to a realistic hydrofoil optimization problem that requires expensive URANS simulations.

2. Methods

This section centers on the proposed Adaptive Gating Multi-Fidelity Network (AGMF-Net) and its assessment. We direct readers to the Appendix for comprehensive methodological background: the Bayesian optimization procedure—including the numerically stable formulation of logarithmic Expected Improvement (logEI)—the benchmark surrogate models (Kriging, MLP, co-Kriging, MFNN), and the evaluation metrics (RMSE, r^2, MARE). In Section 2.1, we develop the AGMF-Net architecture and training objective. In Section 2.2, we evaluate AGMF-Net on benchmark functions, specifying design domains, initialization, acquisition optimization, and the reporting procedure. In Section 2.3,

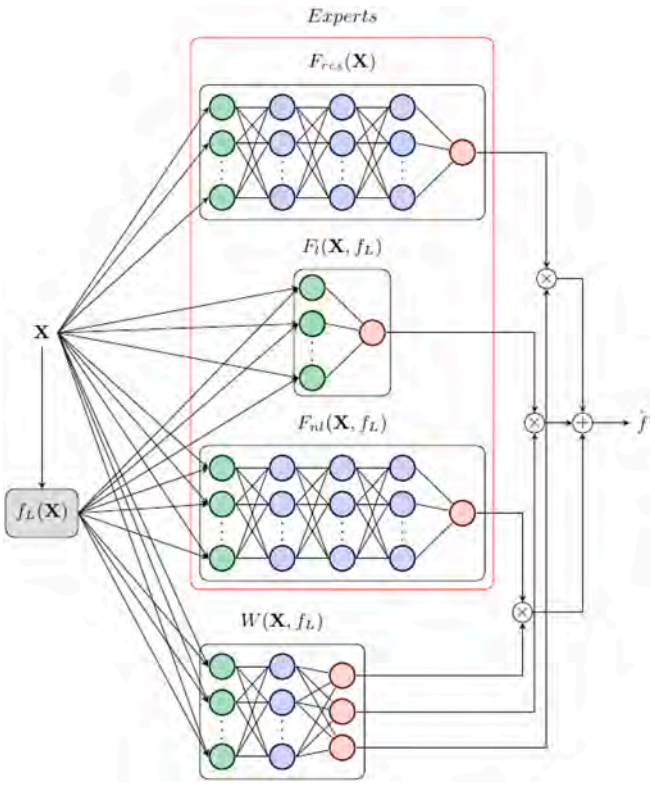

Fig. 1. Architecture of adaptive gated multi-fidelity neural network (AGMF-Net).

we perform a hydrofoil optimization with coupled URANS-XFOIL fidelities, and we describe the simulation configuration and assessment workflow.

2.1. Adaptive gating multi-fidelity neural network (AGMF-Net)

The proposed adaptive gating multi-fidelity neural network (AGMF-Net) aims to construct a generalized multi-fidelity surrogate model capable of accurately approximating complex functions from limited data. This framework builds upon the nonlinear correlation learning strategy of Meng and Karniadakis (2020) and the residual learning approach introduced by Lu et al. (2020) to improve predictive fidelity. Additionally, AGMF-Net addresses the need to effectively map design variables to response quantities in practical engineering settings.

To this end, the architecture adopts the additive formulation of Ada2MF (Zhan et al., 2024), expressed as:

$$\widehat{f}(\mathbf{X}, f_L(\mathbf{X})) = \tanh(\alpha_1)\, F_l(\mathbf{X}, f_L(\mathbf{X})) + \tanh(\alpha_2)\, F_{nl}(\mathbf{X}, f_L(\mathbf{X})) + \tanh(\alpha_3)\, F_{\text{res}}(\mathbf{X}), \tag{1}$$

where F_l, F_{nl}, and F_{res} represent the linear, nonlinear, and residual subnetworks, respectively. Scalar gating parameters $\alpha_i \in \mathrm{R}$ control the contribution of each component.

In Ada2MF, the adaptive multi-fidelity (AMF) module combines the linear, nonlinear, and residual subnetworks using $\tanh(\alpha_i)$ global coefficients that do not depend on $(\mathbf{X}, f_L(\mathbf{X}))$. Such input-invariant weights can be sub-optimal when correlation between the low-fidelity signal f_L and the high-fidelity target f varies across the domain. Moreover, $\tanh(\cdot)$ is unnormalized and permits negative coefficients, which can introduce subtractive combinations and sensitivity in poorly constrained regions; saturation near ± 1 can also slow learning. We therefore replace the global coefficients with input-dependent, normalized weights

$$w(\mathbf{X}) = \mathrm{softmax}(W(\mathbf{X}, f_L(\mathbf{X}))) \tag{2}$$

and define the predictor as

$$\widehat{f}(\mathbf{X}, f_L(\mathbf{X})) = w_1(\mathbf{X}) F_l(\mathbf{X}, f_L(\mathbf{X})) + w_2(\mathbf{X}) F_{nl}(\mathbf{X}, f_L(\mathbf{X})) + w_3(\mathbf{X}) F_{\text{res}}(\mathbf{X}). \tag{3}$$

Each summand corresponds to a distinct expert subnetwork. Fig. 1 depicts the architecture. We quantify predictive uncertainty by aggregating multiple independently initialized instances (deep ensembles), as in our MLP and MFNN baselines. The softmax gating yields non-negative weights that sum to one. This deep mixture-of-experts gate provides context-aware routing (up-weighting the linear/nonlinear experts where f_L aligns with f, and shifting mass to the residual expert where f_L is biased), and ensures each prediction is a convex combination of expert outputs. In practice this tends to improve robustness (no sign-flipped cancellations), interpretability (weights directly indicate each expert's local contribution), and smoother uncertainty aggregation when used with ensembles.

We trained AGMF-Net with paired low- and high-fidelity data. For each minibatch, we evaluated the three expert subnetworks on $(\mathbf{X}, f_L(\mathbf{X}))$ and obtained the prediction $\widehat{f}(\mathbf{X}, f_L(\mathbf{X}))$ and the input-dependent mixture weights $w(\mathbf{X})$ as defined in Equation (2). To guide learning, we decomposed the objective into three complementary Mean Squared Error (MSE) terms measured against the high-fidelity targets:

$$\mathscr{L}_H := \mathrm{MSE}(\widehat{f}(\mathbf{X}, f_L(\mathbf{X})), f(\mathbf{X})), \tag{4a}$$

$$\mathscr{L}_{LH} := \mathrm{MSE}(F_l(\mathbf{X}, f_L(\mathbf{X})) + F_{nl}(\mathbf{X}, f_L(\mathbf{X})), f(\mathbf{X})), \tag{4b}$$

$$\mathscr{L}_R := \mathrm{MSE}(F_{res}(\mathbf{X}), f(\mathbf{X}) - f_L(\mathbf{X})), \tag{4c}$$

$\mathscr{L}_H$ trained the full mixture to match f. $\mathscr{L}_{LH}$ aligned the experts that consume $(\mathbf{X}, f_L(\mathbf{X}))$ with the high-fidelity signal so the model could exploit informative low-fidelity structure. $\mathscr{L}_R$ taught the residual expert to correct the discrepancy $f(\mathbf{X}) - f_L(\mathbf{X})$, which supported robust bias correction when the low-fidelity model was inaccurate.

We then formed a single scalar objective via Adaptive Fast Weighting (AFW) (Zhan et al., 2024):

$$\mathscr{L} = \sum_{k \in \{H, LH, R\}} w_k^{(\text{task})} \mathscr{L}_k, \; w_k^{(\text{task})} = \mathrm{softmax}(u) \in \mathbb{R}^3 \tag{5}$$

Where $u \in \mathbb{R}^3$ are unconstrained logits. After each gradient step on $\mathscr{L}$, we updated u using the relative log-improvements of the component losses to reallocate emphasis toward slower-improving terms and away from faster-improving ones.

$$\ell^{(r)} = \left(\log \mathscr{L}_H^{(r)}, \log \mathscr{L}_{LH}^{(r)}, \log \mathscr{L}_R^{(r)}\right)^{\mathrm{T}}, \Delta\ell^{(r)} = \ell^{(r)} - \ell^{(r+1)}, u^{(r+1)} = u^{(r)} - \eta\left[\mathrm{diag}\left(w^{(\text{task})(r)}\right) - w^{(\text{task})(r)}\left(w^{(\text{task})(r)}\right)^{\mathrm{T}}\right]\Delta\ell^{(r)} \tag{6}$$

with step size $\eta > 0$. This schedule balanced the three learning signals automatically. When the low-fidelity model aligned with the high-fidelity target in a region, $\mathscr{L}_{LH}$ typically decreased rapidly and AFW reduced its weight, while $\mathscr{L}_R$ received less emphasis. When the low-fidelity model was biased, $\mathscr{L}_{LH}$ improved slowly and AFW reweighted toward $\mathscr{L}_R$, prompting the residual expert to explain the discrepancy. In this way, AFW mitigated negative transfer from low-fidelity to high-fidelity learning by down-weighting misleading supervision and up-weighting corrective signals.

The input-dependent softmax gate $w(\mathbf{X})$ at prediction time complemented AFW during training. The gate produced non-negative mixture weights that summed to one, which favored stable convex combinations of experts rather than subtractive cancellations and worked well with our deep-ensemble uncertainty estimates. We trained multiple independently initialized instances and aggregated them as a deep ensemble to quantify predictive uncertainty, following best practice for uncertainty-aware neural surrogates. We standardized inputs and targets before training and de-standardized predictions for report-

Table 1
Hyperparameters used in the Bayesian Optimization of the Forrester Function.

Surrogate Model	MLP	MFNN	Ada2MF	AGMF-Net
Activation Function	Mish	Mish (except F_l)	Mish (except F_l)	Mish (except F_l)
Optimizer	AdamW	AdamW	AdamW	AdamW
Learning Rate	1.00×10^{-3}	1.00×10^{-3}	1.00×10^{-3}	1.00×10^{-3}
Number of Epochs	4000	4000	4000	10000
Number of Ensemble Models	30	30	30	30
Number of Neurons in Hidden Layers	(5,5,5,5,5)	F_l (–) F_{nl} (5,5,5,5,5)	F_l (–) F_{nl} (5,5,5,5,5) F_{res} (5,5,5,5,5)	F_l (–) F_{nl} (5,5,5,5,5) F_{res} (5,5,5,5,5) W (1)

Table 2
Hyperparameters used in the Bayesian Optimization of the Branin Function.

Surrogate Model	MFNN	Ada2MF	AGMF-Net
Activation Function	Mish (except F_l)	Mish (except F_l)	Mish (except F_l)
Optimizer	AdamW	AdamW	AdamW
Learning Rate	1.00×10^{-3}	1.00×10^{-3}	1.00×10^{-3}
Number of Epochs	6000	6000	10000
Number of Ensemble Models	10	10	10
Number of Neurons in Hidden Layers	F_l (–) F_{nl} (10,10,10,10,10)	F_l (–) F_{nl} (10,10,10,10,10) F_{res} (10,10,10,10,10)	F_l (–) F_{nl} (10,10,10,10,10) F_{res} (10,10,10,10,10) W (2)

Table 3
Hyperparameters used in the Bayesian Optimization of the Hartmann-3D Function.

Surrogate Model	MFNN	Ada2MF	AGMF-Net
Activation Function	Mish (except F_l)	Mish (except F_l)	Mish (except F_l)
Optimizer	AdamW	AdamW	AdamW
Learning Rate	1.00×10^{-3}	1.00×10^{-3}	1.00×10^{-3}
Number of Epochs	10000	10000	12000
Number of Ensemble Models	10	10	10
Number of Neurons in Hidden Layers	F_l (–) F_{nl} (15,15,15,15,15)	F_l (–) F_{nl} (15,15,15,15,15) F_{res} (15,15,15,15,15)	F_l (–) F_{nl} (15,15,15,15,15) F_{res} (15,15,15,15,15) W (3)

ing.

In practice, we observed that Adaptive Fast Weighting (AFW) makes the optimization objective time-varying early in training because the task-weight logits u evolve based on relative log-loss improvements. This adaptive reweighting slowed—but stabilized—convergence compared with a single MSE objective. To reach a steady weighting regime and avoid premature bias toward any one component loss, we allocated a larger epoch budget to AGMF-Net than to the other surrogates in every test problem (see the hyperparameters in Tables 1–3). Empirically, the composite loss $\mathscr{L}$ stabilized later than a standard MSE loss and benefited from the extended training schedule.

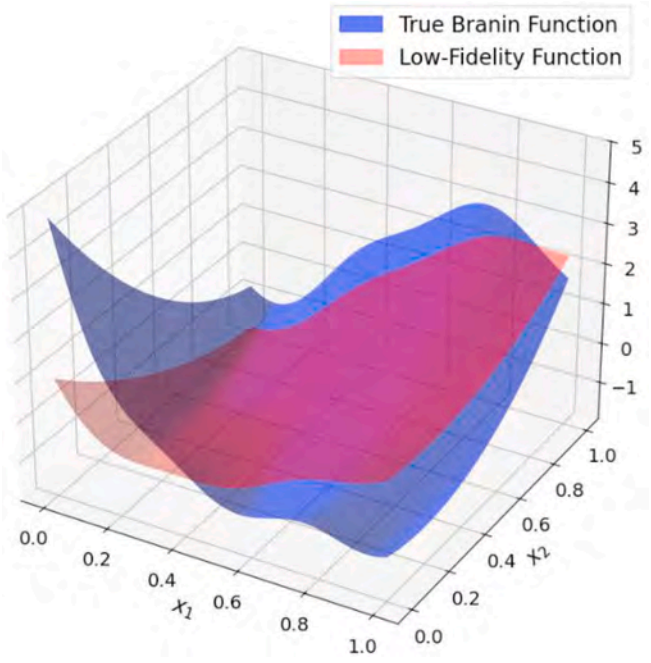

Fig. 2. Modified Branin Function via Dong et al. (Dong et al., 2015).

2.2. Benchmark function experiments

2.2.1. Forrester function

The Forrester function served as a single-variable benchmark to assess surrogate model performance. Forrester et al. (2008) originally proposed the function over the domain $x \in [0,1]$, defined as:

$$f(x) = (6x-2)^2 \sin(12x-4), \tag{7}$$

$$f_L(x) = \frac{1}{2} f(x) + 10(x-0.5) - 5. \tag{8}$$

Equation (7) represents the high-fidelity target, while Equation (8) introduces a biased low-fidelity approximation. The function contains a global minimum at $x \approx 0.75725$ and a local minimum at $x \approx 0.14259$, which challenge surrogate models to capture nonconvex behavior.

This study applied all six surrogate models described in Subsection 2.1 and Appendix to approximate the function. Table 1 reports the hyperparameters used for neural network-based models. The experiment initialized with three evenly spaced samples from the domain $[0,1]$. The Bayesian Optimization procedure, outlined in Appendix A, continued until one of the models predicted the global minimum with a MARE below 5 %.

We assessed predictive performance by computing RMSE and r^2 over 200 evenly distributed test points within the design space. To evaluate the surrogate's accuracy at the optimum, we determined MARE by comparing predictions with the known global minimum point.

2.2.2. Modified Branin Function

To extend the evaluation to two dimensions, the modified Branin function was selected. The design space $(x_1, x_2) \in [0,1]^2$ was transformed into the original Branin domain using:

$$u_1 = 15x_1 - 5, u_2 = 15x_2. \tag{9}$$

The high-fidelity Branin function was normalized and defined as:

$$f(u_1,u_2) = \frac{1}{51.95}\left[\left(u_2 - \frac{5.1u_1^2}{4\pi^2} + \frac{5u_1}{\pi} - 6\right)^2 + \left(10 - \frac{10}{8\pi}\right)\cos(u_1) - 44.81\right]. \tag{10}$$

Following the multi-fidelity structure proposed by Dong et al.

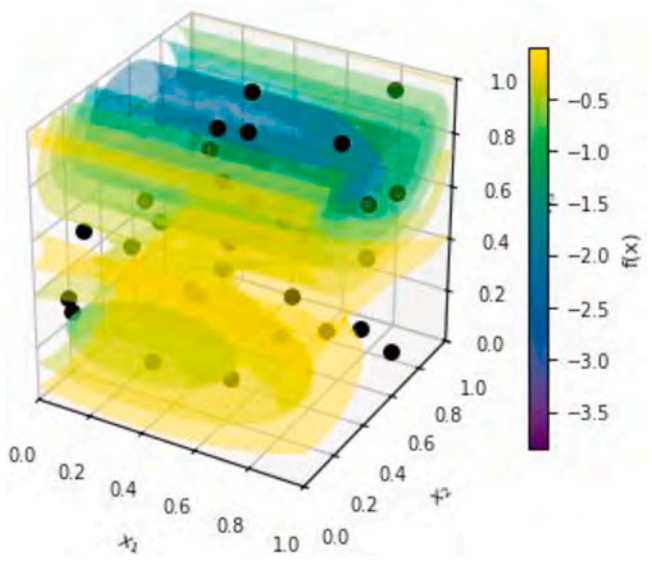

Fig. 3. Hartmann-3D high-fidelity objective visualized with semi-transparent isosurfaces (marching-cubes at 10/25/40/60/80th percentiles); black markers denote the 30 initial optimal Latin hypercube samples in $[0,1]^3$

Table 4
Design space and flow conditions for NACA $66_1 - 012$ hydrofoil optimization.

Parameter	Value	Unit
Design domain (c)	[0, 0.04]	% chord
Angle of attack	3	degrees
Flow speed	9.45	m/s
Reynolds number	8.93×10^5	
Kinematic viscosity	8.87×10^{-7}	m^2/s
Water density	997	kg/m^3
Mach number	6.33×10^{-3}	

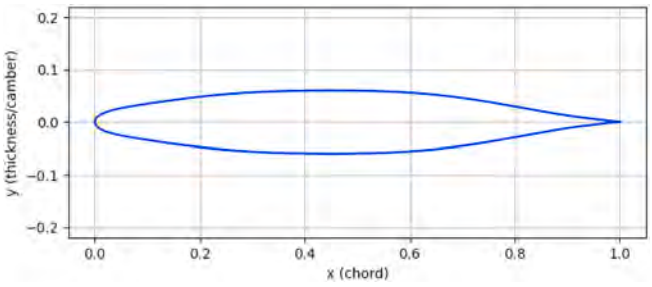

Fig. 4. NACA $66_1 - 012$ via Kermeen (Kermeen and Plesset, 1956) with CST resampling coordinates.

(2015), the low-fidelity function incorporated a quadratic bias:

$$f_L(u_1,u_2) = f(u_1,u_2) + 20(0.9+u_1)^2 - 50. \tag{11}$$

The Branin function contained three global minima in the (u_1,u_2) space, approximately located at:

$(-3.1950, 12.275)$, $(9.4248, 2.475)$, $(3\pi, 2.475)$.

Under the normalization in Equation (9), these minima corresponded to:

$$(x_1,x_2) \approx (0.1239, 0.8183), (0.5428, 0.1517), (0.9617, 0.1650).$$

Fig. 2 shows the resulting function landscape. This experiment applied four multi-fidelity surrogate models—co-Kriging, MFNN, Ada2MF and AGMF-Net—as described in Subsection 2.1 and Appendix Table 2 lists the corresponding neural network hyperparameters. The

Fig. 5. CFD simulation domain.

$[0,1]^2$ domain was initially sampled using 16 points generated via optimal Latin hypercube sampling (Franco, 2008). Bayesian Optimization followed the procedure in Appendix A and terminated once any model achieved a minimum with a MARE below 5 %.

We evaluated predictive performance by computing RMSE and r^2 over 40,000 evenly distributed grid sampling test points within the design domain. To assess local accuracy at the optima, we calculated MARE by comparing predictions with the known values at the three global minimum points.

2.2.3. Hartmann-3D

We used the conventional Hartmann-3D as the high-fidelity target on $[0,1]^3$:

$$f(x) = -\sum_{i=1}^{4} \alpha_i \exp\left(-\sum_{j=1}^{3} \beta_{ij}\left(x_j - P_{ij}\right)^2\right), \tag{12}$$

with

$$\alpha = \{1.0, 1.2, 3.0, 3.2\}, P = \begin{bmatrix} 0.3689 & 0.1170 & 0.2673 \\ 0.4699 & 0.4387 & 0.7470 \\ 0.1091 & 0.8732 & 0.5547 \\ 0.0381 & 0.5743 & 0.8828 \end{bmatrix}, \beta = \begin{bmatrix} 3 & 10 & 30 \\ 0.1 & 10 & 35 \\ 3 & 10 & 30 \\ 0.1 & 10 & 35 \end{bmatrix} \tag{13}$$

Following Toal's adjustable construction (Toal, 2015), we defined the low-fidelity function as

$$f_L(x) = -\sum_{i=1}^{4} \alpha_i \exp\left(-\sum_{j=1}^{3} \beta_{ij}\left(x_j - \frac{3}{4}P_{ij}(a+1)\right)^2\right), \tag{14}$$

And we fixed the correlation parameter at $a = 0.5$.

Fig. 3 renders the high-fidelity Hartmann-3D landscape as a set of percentile isosurfaces, revealing multiple separated basins and sharp ridges that indicate strong non convexity and anisotropy.

The optimization and evaluation procedure matched section 2.2.1 and 2.2.2. We initialized $[0,1]^3$ with 30 points from an optimal Latin hypercube. For evaluation, we computed RMSE and r^2 on a dense grid of $40^3 = 64,000$ test points. Neural-network hyperparameters for MFNN, Ada2MF, and AGMF-Net appear in Table 3.

2.3. Hydrofoil optimization

Following the benchmark evaluations of AGMF-Net, we applied the model to a real-world engineering problem: optimizing the lift-to-drag ratio (C_L/C_D) of a hydrofoil. The initial geometry was based on the NACA $66_1 - 012$ profile provided by Kermeen (Kermeen and Plesset,

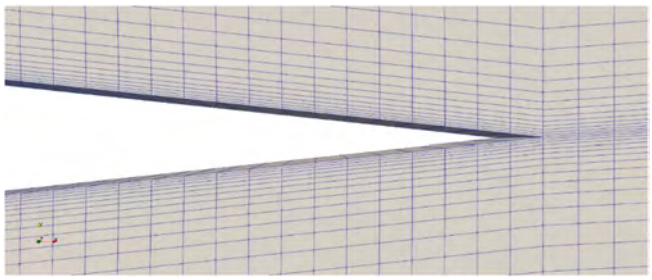

Fig. 6. Mesh boundary layer at trailing edge.

1956). The optimization involved a single design variable—the maximum camber coefficient located at 42 % of the chord length. Optimization was conducted within a non-cavitating regime, under the flow conditions and design domain summarized in Table 4.

The digitized coordinates of NACA $66_1 - 012$ provided by Kermeen (Kermeen and Plesset, 1956) lacked sufficient resolution for both CFD and XFOIL analysis. To address this limitation, we reconstructed a high-resolution geometry using the class-shape transformation (CST) method (Kulfan, 2008) with a sixth-order Bernstein polynomial. We scaled the airfoil to a unit chord length to simplify numerical treatment (Fig. 4).

We computed the high-fidelity objective function by solving the incompressible, unsteady Reynolds-averaged Navier–Stokes (URANS) equations using a finite volume method:

$$\text{Continuity equation}: \nabla \cdot \mathbf{u} = 0 \tag{16a}$$

$$\text{Momentum equation}: \frac{\partial \mathbf{u}}{\partial t} + \nabla \cdot (\mathbf{u} \otimes \mathbf{u}) = -\frac{1}{\rho}\nabla p + \nabla \cdot [\,(\nu + \nu_t)\nabla \mathbf{u}]. \tag{16b}$$

We modeled transition using the SST-$\gamma - Re_\theta$ transition model (Langtry and Menter, 2009), a correlation-based extension of the SST k-ω turbulence model. This framework introduces two additional transport equations for intermittency (γ) and the transition momentum-thickness Reynolds number (Re_θ), enabling accurate prediction of laminar-turbulent transition. The model is particularly suitable for simulating the hydrodynamic performance of airfoils and hydrofoils, where the location of the transition and separation points has a decisive influence on lift and drag. The governing equations are expressed as:

$$\frac{\partial(\rho\, k)}{\partial t} + \nabla \cdot (\rho\, \mathbf{u}\, k) = \nabla \cdot [(\mu + \sigma_k\, \mu_t)\nabla k] + P_k - \beta^*\, \rho\, k\, \omega \tag{17a}$$

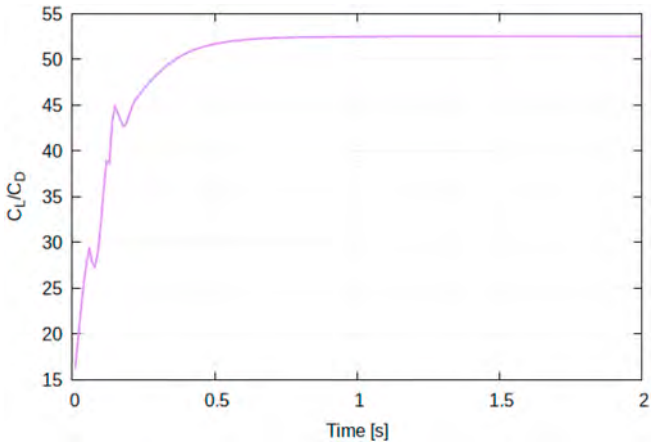

Fig. 7. The ratio of lift and drag coefficients time domain.

Table 6
Hyperparameters used in Bayesian Optimization of the Hydrofoil Problem.

Surrogate Model	XFOIL	Coupled XFOIL and CFD
Activation Function	Mish	Mish (except F_l)
Optimizer	AdamW	AdamW
Learning Rate	1.00×10^{-3}	1.00×10^{-3}
Number of Epochs	4000	4000
Number of Ensemble Models	10	10
Number of Neurons in Hidden Layers	(5,5,5,5,5)	F_l (−) F_{nl} (5,5,5,5,5) F_{res} (5,5,5,5,5) W (1)

$$\frac{\partial(\rho\, \omega)}{\partial t} + \nabla \cdot (\rho\, \mathbf{u}\, \omega) = \nabla \cdot [(\mu + \sigma_\omega\, \mu_t)\nabla \omega] + \alpha\, \frac{\rho\, P_k}{\mu_t} - \beta\, \rho\, \omega^2 + 2\, \rho\, (1 - F_1)\, \sigma_{\omega 2}\, \frac{\nabla k \cdot \nabla \omega}{\omega} \tag{17b}$$

$$\frac{\partial(\rho\, \gamma)}{\partial t} + \nabla \cdot (\rho\, \mathbf{u}\, \gamma) = P_\gamma - E_\gamma + \nabla \cdot \left[\left(\mu + \frac{\mu_t}{\sigma_\gamma}\right)\nabla \gamma\right] \tag{17c}$$

$$\frac{\partial(\rho\, Re_\theta)}{\partial t} + \nabla \cdot (\rho\, \mathbf{u}\, Re_\theta) = P_{Re_\theta} - E_{Re_\theta} + \nabla \cdot \left[\left(\mu + \frac{\mu_t}{\sigma_\theta}\right)\nabla Re_\theta\right] \tag{17d}$$

We used the pimpleFoam solver in OpenFOAM® v2412, which implements the PIMPLE algorithm (Weller et al., 1998). To accelerate convergence, we enabled local time stepping (LTS) (Jeanmasson et al., 2018). We selected pimpleFoam over simpleFoam due to stability advantages under the specified flow regime.

We generated a C-type mesh containing 4,662,000 hexahedral cells

Table 5
Boundary and initial conditions for each transported field in the OpenFOAM® simulation.

Field	Inlet	Outlet	Airfoil	Initial field
Pressure (m^2/s^2)	freestreamPressure = 0	freestreamPressure = 0	zeroGradient	uniform 0
Velocity U (m/s)	freestreamVelocity = (0.790913, 0.04145, 0)	freestreamVelocity = (0.790913, 0.04145, 0)	noSlip	uniform (0.790913, 0.04145, 0)
Turbulent viscosity ν_t (m^2/s^2)	calculated	calculated	fixedValue 1e-10 (for stability)	uniform 1e-10 (for stability)
Turbulent kinetic energy k (m^2/s^2)	turbulentIntensit yKineticEnergyInl et intensity = 0.01	zeroGradient	fixedValue 1e-10 (for stability)	uniform 9.40892e-5
Specific dissipation ω (s^{-1})	turbulentMixingLe ngthFrequencyInle t mixingLength = 0.0122474	zeroGradient	fixedValue 181,636	uniform 2.45404e-3
Intermittency γ	fixedValue 1	zeroGradient	zeroGradient	uniform 1
Momentum thickness Reynolds number $Re_{\theta t}$	fixedValue 1	zeroGradient	zeroGradient	uniform 1

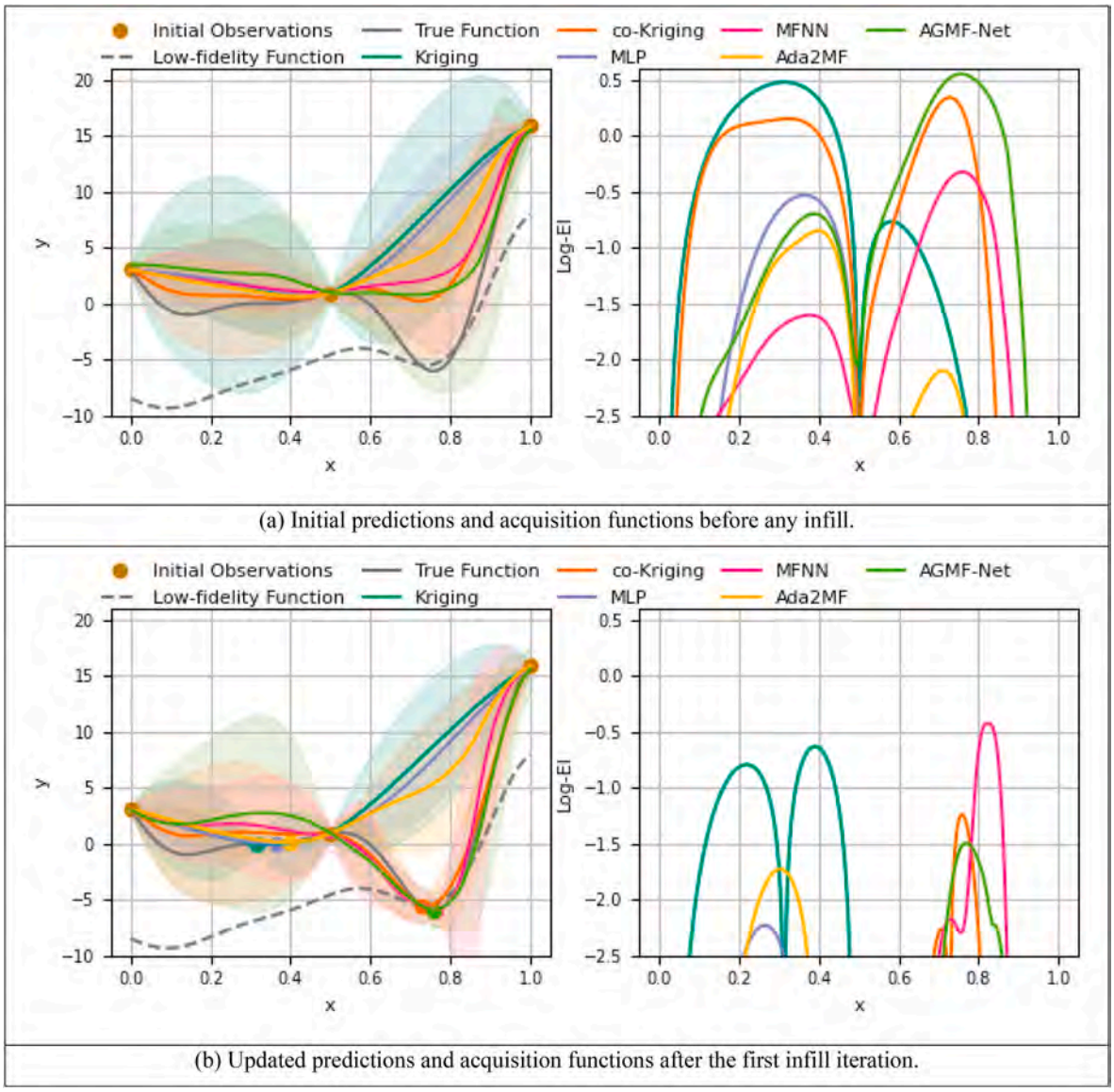

Fig. 8. Comparison of predictive performance and acquisition functions of six surrogate models for Bayesian optimization of the Forrester function.

Table 7
Comparison of surrogate model performance over infill iterations for the Forrester function optimization. The best performance in each row is highlighted in bold.

Metric	Infill Iteration	Kriging	MLP	co-Kriging	MFNN	Ada2MF	AGMF-Net
RMSE	0	44.7229	38.9339	**8.6760**	13.6389	24.8582	10.8987
	1	43.5161	36.5703	**0.8620**	2.7143	26.7417	2.9163
r^2	0	0.2069	0.2552	0.7822	0.7154	0.4586	**0.8393**
	1	0.1892	0.2433	**0.9711**	0.9326	0.3962	0.8919
MARE	0	2.4417	2.2814	**1.0777**	1.3832	1.7932	1.1554
	1	2.4463	2.2538	0.0974	0.0395	1.8896	**0.0308**

(Fig. 5) and partitioned it into 168 subdomains using the Scotch decomposition method (Pellegrini and Roman, 1996). The simulations were executed on the high-performance computing facilities at the Computational Marine Hydrodynamics Laboratory (CMHL). To fully resolve the boundary layer without resorting to wall functions, we set the first cell height to 1.25×10^{-4} m, yielding $y^+ \approx 1$ (Fig. 6). The boundary conditions for all transported fields are summarized in Table 5. We applied second-order upwind schemes (Van Leer, 1979) and continued simulations until the C_L/C_D ratio stabilized (Fig. 7).

To construct the low-fidelity model, we employed XFOIL 6.99, using viscous analysis for discrete design points. XFOIL is an interactive 2-D airfoil analysis/design program that solves an inviscid potential flow with a high-order panel method and couples it to an integral boundary-layer solver with transition modeling, enabling rapid prediction of lift, drag, and moment polars at subsonic Reynolds numbers (Drela, 1989) (Drela and Youngren, 2001). We validated both fidelity levels against experimental data for the initial design. Three high-fidelity design points were sampled uniformly across the design domain and evaluated using CFD.

To support the multi-fidelity model, we trained a single-fidelity

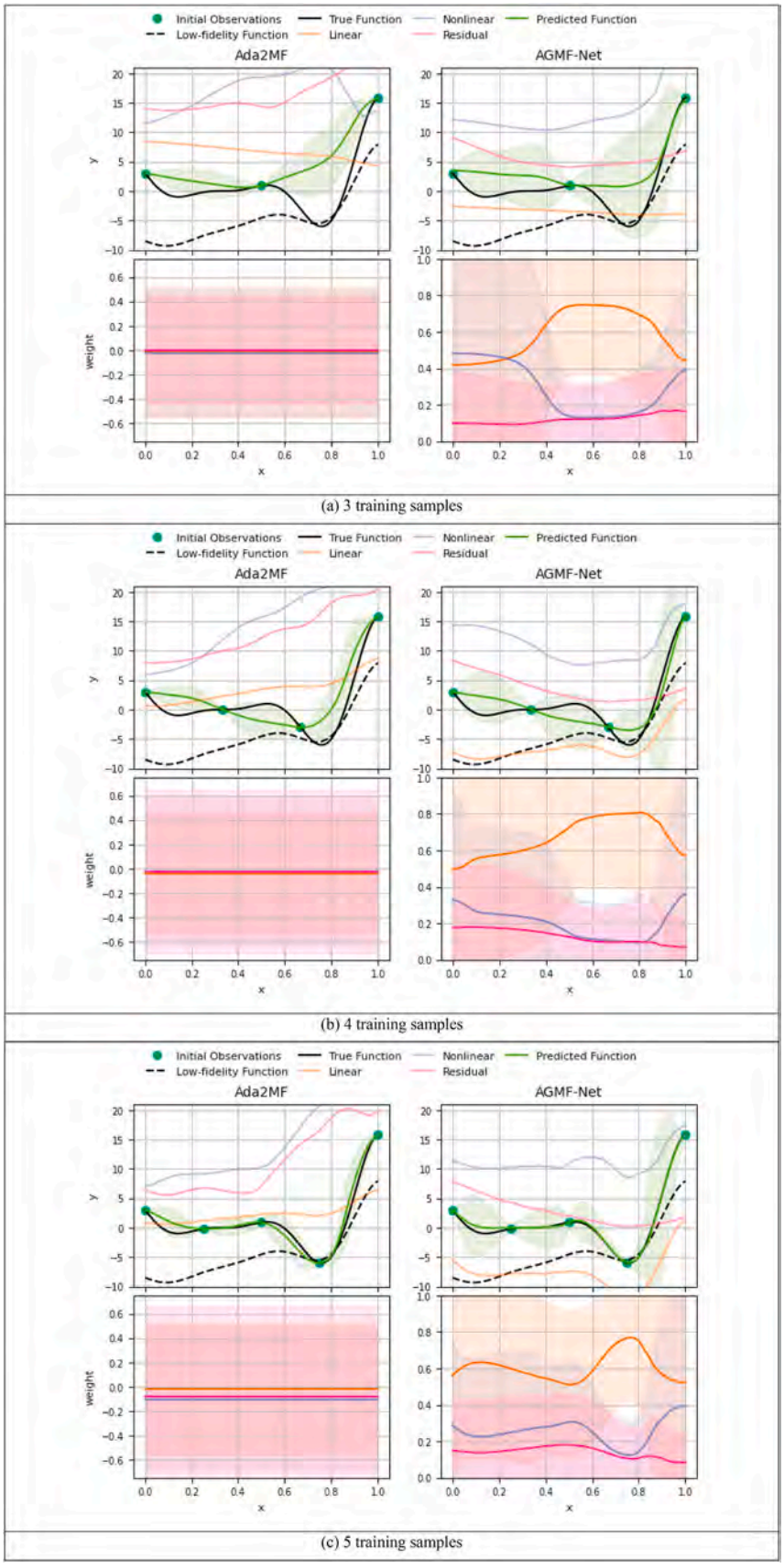

Fig. 9. Decomposition of Ada2MF and AGMF-Net on the Forrester function with increasing numbers of training samples. The upper panels illustrate the model predictions and the corresponding decomposed components, while the lower panels show the weight space that governs their combination across the domain.

Table 8
Comparison of Ada2MF and AGMF-Net on the Forrester function under different training sample sizes.

Metric	Training Samples	Ada2MF	AGMF-Net
RMSE	3	24.8582	**10.8987**
	4	7.7119	**3.8214**
	5	0.9724	**0.1455**
r^2	3	0.4586	**0.8393**
	4	0.7271	**0.8128**
	5	0.9668	**0.9943**
MARE	3	1.7932	**1.1554**
	4	0.6989	**0.4184**
	5	**0.0126**	0.0162

surrogate using a multilayer perceptron (MLP) on 100 low-fidelity samples. We then constructed the AGMF-Net multi-fidelity surrogate using the hyperparameters listed in Table 6. Two Bayesian optimization iterations explored the objective landscape, followed by surrogate optimization. The resulting optimal design was compared against the baseline.

3. Results and discussion

3.1. Evaluation of surrogate models using the forrester function

Fig. 8a (iteration 0) and Table 7 together show two clear patterns. First, in this one-dimensional setting, co-Kriging provides the strongest global fit in terms of RMSE at initialization, reflecting the well-known strength of Gaussian process co-Kriging for smooth, low-dimensional functions with an approximately stationary discrepancy. However, AGMF-Net achieves the highest r^2 among all models, indicating a stronger overall correlation with the true function even with sparse high-fidelity data. Second, AGMF-Net already ranks among the top methods at the optimum (low MARE) despite using the same sparse high-fidelity samples, indicating that its gated fusion and residual correction focus predictions near the true minimizer from the outset.

After the first infill (Fig. 8b), the divide between "global fit" and "optimization accuracy at the minimizer" becomes more pronounced. Co-Kriging continues to deliver the lowest RMSE, confirming that it still predicts the overall 1D function shape extremely well. In contrast, AGMF-Net and MFNN reduce the optimal-point error most aggressively, with AGMF-Net maintaining high correlation and the lowest MARE. Ada2MF improves more slowly under the same budget; and the single-fidelity baselines (Kriging, MLP) remain least competitive on all metrics.

The acquisition surfaces in Fig. 8a and b are consistent with these outcomes. Co-Kriging's calibrated uncertainty sharpens the full-domain reconstruction rapidly, while AGMF-Net concentrates Log-EI mass near the global minimizer early, which accelerates reduction of the error at the optimum. This contrast explains why co-Kriging leads the table on RMSE in this problem, yet AGMF-Net satisfies the MARE-based stopping criterion within a single iteration.

Fig. 9a–c compare Ada2MF and the proposed AGMF-Net on the Forrester benchmark using three, four, and five evenly spaced observations. Both models reconstruct the high-fidelity response through three experts—linear, nonlinear, and residual—combined by adaptive weights, where Ada2MF employs global coefficients and AGMF-Net introduces input-dependent gating over $(\mathbf{X}, f_L(\mathbf{X}))$. As shown in Table 8, AGMF-Net consistently outperforms Ada2MF across all metrics and sample sizes, achieving lower RMSE and higher r^2 for every case. The only exception is the MARE at five samples, where AGMF-Net is slightly higher but remains comparable. The improvement is most evident in the highly nonlinear region $(\mathbf{X} > 0.85)$ for the three- and four-sample cases (Fig. 9a and b), where Ada2MF fails to capture the strong curvature of the true function, whereas AGMF-Net closely tracks the target. With five samples (Fig. 9c), AGMF-Net reproduces the full function profile with near-perfect accuracy, demonstrating superior correlation adaptivity and representational robustness.

The upper panels show the decomposition of the ensemble-mean prediction into three ensemble-mean expert curves—linear, nonlinear, and residual—corresponding to the model's additive architecture. The lower panels illustrate the ensemble-mean weights. In Ada2MF, each ensemble member learns fixed global coefficients $\tanh(\alpha_i)$ that remain constant across the domain; thus, the averaged weights appear absolutely flat, directly reflecting their input-invariant nature. In contrast, AGMF-Net's softmax-based gating produces input-dependent, positive, and normalized weights that sum to one, enabling the model to form convex combinations of the experts and to adapt smoothly to local fidelity correlations.

It should also be noted that the linear expert is linear with respect to the joint input $(\mathbf{X}, f_L(\mathbf{X}))$, not $\mathbf{X}$ alone. Because $f_L(\mathbf{X})$ is nonlinear in $\mathbf{X}$, the corresponding $F_l(\mathbf{X}, f_L(\mathbf{X}))$ curve appears curved when plotted along $\mathbf{X}$, even though the mapping in the joint input space remains linear.

3.2. Evaluation of surrogate models using the Modified Branin Function

We restrict this comparison to the multi-fidelity surrogates—co-Kriging, MFNN, Ada2MF, and AGMF-Net—because multi-fidelity consistently dominated single-fidelity in our Forrester study and in prior reports (Kandasamy et al., 2017) (Peherstorfer et al., 2018b). Fig. 10a–e visualizes the predictive fields across four infill iterations, and Table 9 summarizes RMSE, r^2, and MARE. At initialization (Fig. 10a), AGMF-Net already achieved the lowest global error and highest explained variance while also yielding the smallest MARE at the minima, indicating that input-dependent gating with a residual expert aligned quickly with the biased low-fidelity signal under sparse high-fidelity data. As additional samples accumulated (Fig. 10b–e), MFNN drove the strongest reduction in RMSE and r^2, reflecting its emphasis on global fit, whereas AGMF-Net concentrated probability mass around the true basins and delivered the best optimum-focused accuracy by the final stage (lowest MARE in Table 9) while maintaining competitive r^2. Ada2MF started behind but improved substantially with more data, becoming globally competitive by the end; however, it did not match AGMF-Net near the optima. Co-Kriging remained the weakest of the four throughout this two-dimensional task—its conservative Gaussian-process posterior produced smooth, under-responsive fields and consistently inferior RMSE, r^2, and MARE compared with the neural network surrogates. Taken together, Figs. 10a–e and Table 9 show AGMF-Net's principal strength on this problem: rapid alignment with useful low-fidelity structure at the start and superior localization of the global minima as the BO loop progressed.

3.3. Evaluation of surrogate models using the Hartmann-3D function

Table 10 reports the evaluation of the four multi-fidelity surrogates on the Hartmann-3D benchmark under Bayesian optimization, listing performance at initialization and after nine infill iterations. Across the entire sequence, AGMF-Net consistently attains the strongest global fit and the most accurate estimates at the optimum. By the final iteration, it achieves the best entries in all three columns of Tables 10 and is the only model whose accuracy at the predicted optimum meets the 5 % threshold within the allotted iterations. Ada2MF ranks second overall: it narrows the gap as more data are acquired and maintains competitive global approximation quality, but it does not reach the accuracy achieved by AGMF-Net at the optimum. MFNN continues to improve with added information yet remains behind the adaptive-gated variants in both global fidelity and optimal-point accuracy. Co-Kriging performs worst throughout, indicating that on this three-dimensional landscape the neural multi-fidelity surrogates—particularly the proposed adaptive-gated design—offer clear advantages over a Gaussian-process baseline. These results substantiate the benefit of replacing fixed,

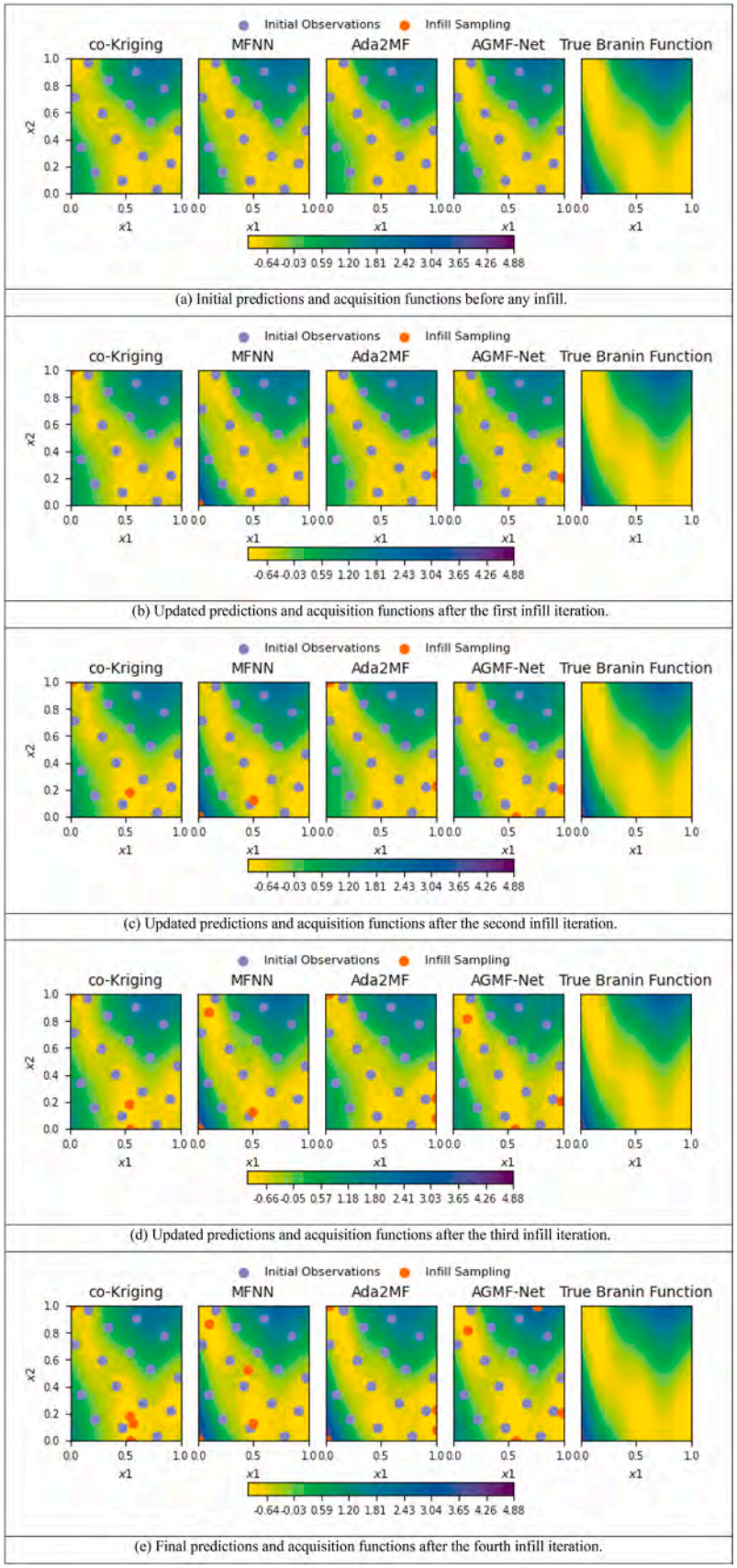

Fig. 10. Comparison of predictive performance of four multi-fidelity surrogate models for Bayesian optimization of the Branin function.

Table 9
Comparison of surrogate model performance over infill iterations for the modified Branin function optimization. The best performance in each row is highlighted in bold.

Metric	Infill Iteration	co-Kriging	MFNN	Ada2MF	AGMF-Net
RMSE	0	0.1273	0.1217	0.1705	**0.0945**
	1	0.1294	**0.0361**	0.1866	0.1489
	2	0.1218	**0.0159**	0.2038	0.1607
	3	0.1154	**0.0100**	0.1470	0.1469
	4	0.1137	**0.0091**	0.0221	0.0684
r^2	0	0.8902	0.8930	0.8408	**0.9195**
	1	0.8894	**0.9652**	0.8224	0.8683
	2	0.8961	**0.9867**	0.8047	0.8697
	3	0.9021	**0.9901**	0.8675	0.8937
	4	0.9036	**0.9910**	0.9786	0.9438
MARE	0	0.2185	0.1714	0.1298	**0.1206**
	1	0.2100	0.2152	**0.1366**	0.2364
	2	0.1736	**0.1161**	0.1280	0.1446
	3	0.1683	0.0685	0.1381	**0.0565**
	4	0.1619	0.0637	0.1206	**0.0425**

Table 10
Comparison of surrogate model performance over infill iterations for the Hartmann-3D function optimization. The best performance in each row is highlighted in bold.

Metric	Infill Iteration	co-Kriging	MFNN	Ada2MF	AGMF-Net
RMSE	0	0.2288	0.1915	**0.1383**	0.1797
	1	0.2353	0.2577	0.1405	**0.1264**
	2	0.2636	0.2919	0.1439	**0.0987**
	3	0.2476	0.2416	0.1249	**0.1023**
	4	0.2219	0.2215	0.1108	**0.0819**
	5	0.2275	0.2041	0.1156	**0.0727**
	6	0.2063	0.1890	0.0911	**0.0558**
	7	0.2121	0.1732	0.0964	**0.0617**
	8	0.1954	0.1438	0.0827	**0.0492**
	9	0.1886	0.1286	0.0789	**0.0461**
r^2	0	0.7506	0.7850	**0.8588**	0.8012
	1	0.7445	0.7124	0.8549	**0.8580**
	2	0.7228	0.6724	0.8499	**0.8925**
	3	0.7361	0.7165	0.9387	**0.9412**
	4	0.7749	0.7348	0.9552	**0.9719**
	5	0.7622	0.7648	0.9511	**0.9780**
	6	0.7988	0.8135	0.9672	**0.9857**
	7	0.7895	0.8345	0.9620	**0.9832**
	8	0.8123	0.8648	0.9724	**0.9876**
	9	0.8234	0.8850	0.9756	**0.9891**
MARE	0	0.5049	**0.3938**	0.5196	0.4198
	1	0.5070	0.4408	0.4311	**0.3606**
	2	0.5089	0.4838	0.3403	**0.2803**
	3	0.4971	0.3615	0.2988	**0.2011**
	4	0.4736	0.3052	0.1873	**0.1436**
	5	0.4819	0.3395	**0.1154**	0.1192
	6	0.4523	0.2761	0.0812	**0.0749**
	7	0.4597	0.2494	0.1052	**0.0816**
	8	0.4388	0.2379	0.0816	**0.0587**
	9	0.4215	0.2241	0.0690	**0.0419**

Table 11
Validation of lift and drag coefficients against experimental data.

Metric	Experiment	CFD	XFOIL	CFD Error (%)	XFOIL Error (%)
C_L	0.267	0.272257	0.295	**1.97**	10.5
C_D	0.0062	0.0073468	0.008	**18.5**	29.0
C_L/C_D	43.1	37.0580	36.875	**13.9**	14.4

domain-invariant mixing with the input-dependent softmax gate in AGMF-Net, which more effectively leverages the low-fidelity signal while routing corrections through the residual pathway, yielding superior performance under sparse high-fidelity sampling.

3.4. Optimization of hydrofoil

A comparative validation of CFD and viscous XFOIL predictions against Kermeen's experimental data (Pellegrini and Roman, 1996) is presented in Table 11. The CFD simulation yielded an accurate prediction of the lift coefficient, with an error of less than 2 %. However, it overpredicted the drag coefficient by 18.5 %. This deviation is likely due to the high sensitivity of drag to adverse pressure gradients and boundary layer separation. Even a minor shift in the separation point can significantly increase drag, especially since C_D is two orders of magnitude smaller than C_L, amplifying relative errors in the C_L/C_D ratio. As a result, the error in C_L/C_D reached 13.9 %.

Although XFOIL was executed in viscous mode, it exhibited higher discrepancies across all metrics compared to CFD. This outcome underscores CFD's superior capability to capture detailed flow physics, including transitional and separation effects, which remain challenging for integral boundary layer methods like those used in XFOIL.

The optimization progression driven by Bayesian inference is illustrated in Fig. 11. Initially, the surrogate model—trained on a small set of observations—captures a coarse landscape of the design space, with the acquisition function (Log-EI) indicating high potential near a camber coefficient of approximately 0.0248. This point is selected for the first infill iteration. After evaluating it, the model is updated, and the second acquisition maximum emerges near a slightly lower camber of 0.0227, which is then sampled next. The third iteration further concentrates around this optimal region, demonstrating the model's ability to refine its predictions and reduce uncertainty. Notably, the acquisition peaks become more localized across iterations, indicating increasing model confidence and convergence toward the optimal design region.

The hydrodynamic improvements are quantitatively summarized in Table 12 and physically illustrated through the pressure and velocity field distributions in Fig. 12. The optimized hydrofoil, corresponding to a maximum camber coefficient of 0.0239, yields a significant lift enhancement—from C_L to 0.2723—representing a 93.6 % increase. Although this comes with a moderate increase in drag (C_D rising from 0.0073 to 0.0100), the overall hydrodynamic efficiency (C_L/C_D) improves markedly by 41.6 %, from 37.06 to 52.49.

Fig. 12a and b reveal the flow mechanisms behind this performance improvement. The pressure field around the optimized hydrofoil exhibits a stronger low-pressure region on the suction side near the leading edge, creating a larger pressure differential and thus stronger lift. Simultaneously, the velocity fields in Fig. 12c and d shows greater flow acceleration over the upper surface of the optimized shape, confirming enhanced suction and a favorable pressure gradient. Importantly, the flow remains fully attached in both cases, indicating that the increased camber did not induce boundary layer separation, which is critical for maintaining drag control.

Fig. 13 highlights the geometric difference between the baseline and optimized foils. The optimized shape introduces a subtle positive camber—peaking at 0.0239—without modifying the original thickness distribution. This asymmetry results in a slightly forward-leaning mean camber line and an effectively increased angle of attack, which in turn promotes early suction buildup over the upper surface. Despite its small magnitude, this camber adjustment proves highly effective, confirming the hydrodynamic sensitivity of lift-to-drag ratio to fine geometric modifications under laminar conditions.

4. Conclusions

This study introduced AGMF-Net, a novel adaptive gated multi-fidelity neural network that integrates deep Mixture-of-Experts gating

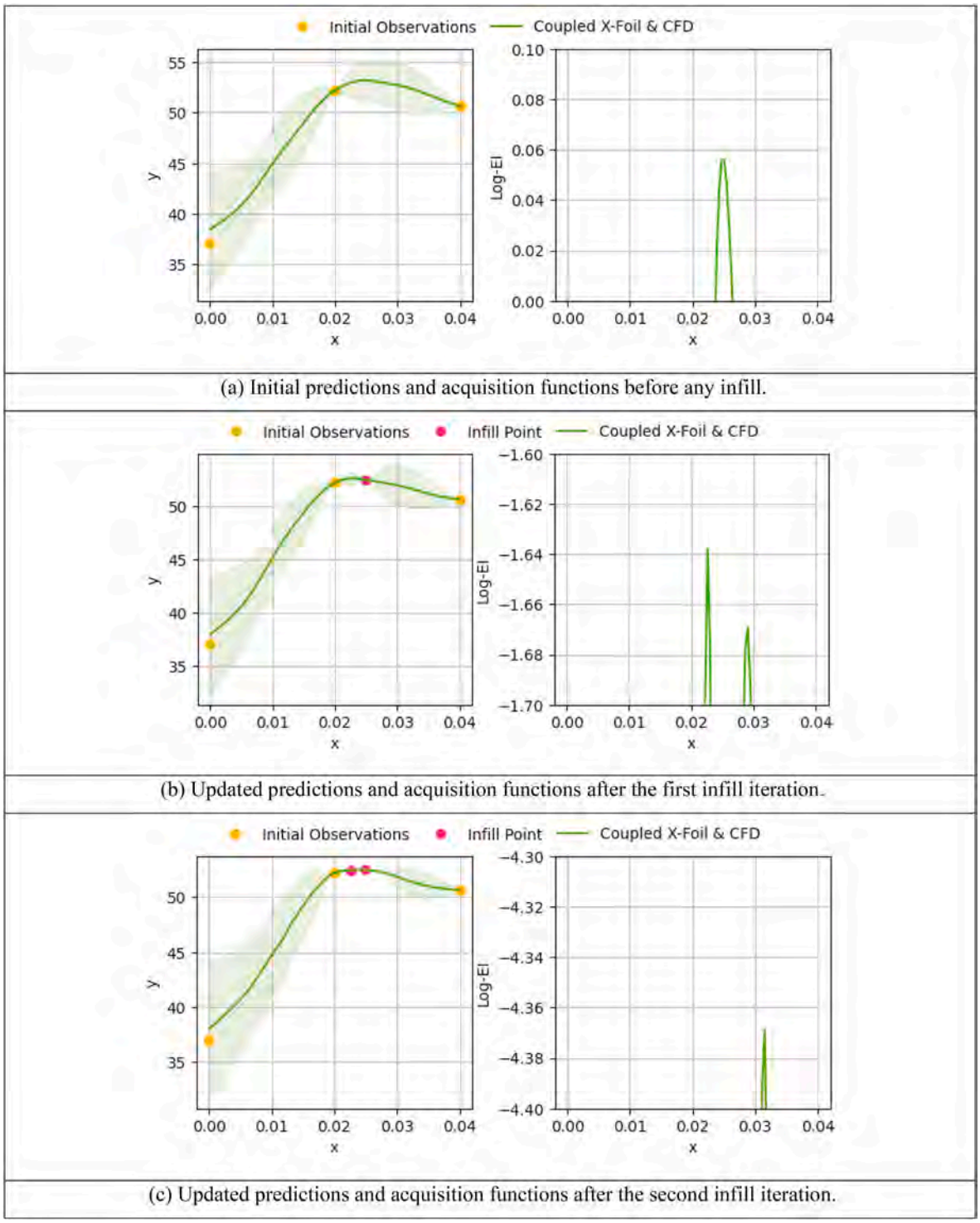

Fig. 11. Bayesian optimization of hydrofoil design using AGMF-Net with coupled XFOIL and CFD observations. The left subplots show surrogate predictions and uncertainty bands, while the right subplots display the Log-Expected Improvement (Log-EI) used for infill point selection.

Table 12
Comparison of hydrodynamic performance between the baseline and optimized hydrofoils.

Performance Metric	Baseline	Optimized	Relative Change (%)
C_L	0.2723	0.5270	+93.6
C_D	0.0073	0.0100	+36.7
$C_L/\ C_D$	37.06	52.49	+41.6

with ensemble-based uncertainty quantification within a Bayesian optimization framework. The development of AGMF-Net addressed the need for a surrogate model that is both accurate and robust under limited high-fidelity data, a scenario common in engineering design optimization. In our experiments, AGMF-Net achieved strong predictive accuracy even at the initial sampling stage, outperforming all other models in RMSE and r^2 metrics before any infill iterations. This early advantage was attributed to its ability to leverage both low-fidelity and sparse high-fidelity data through adaptive weighting of linear,

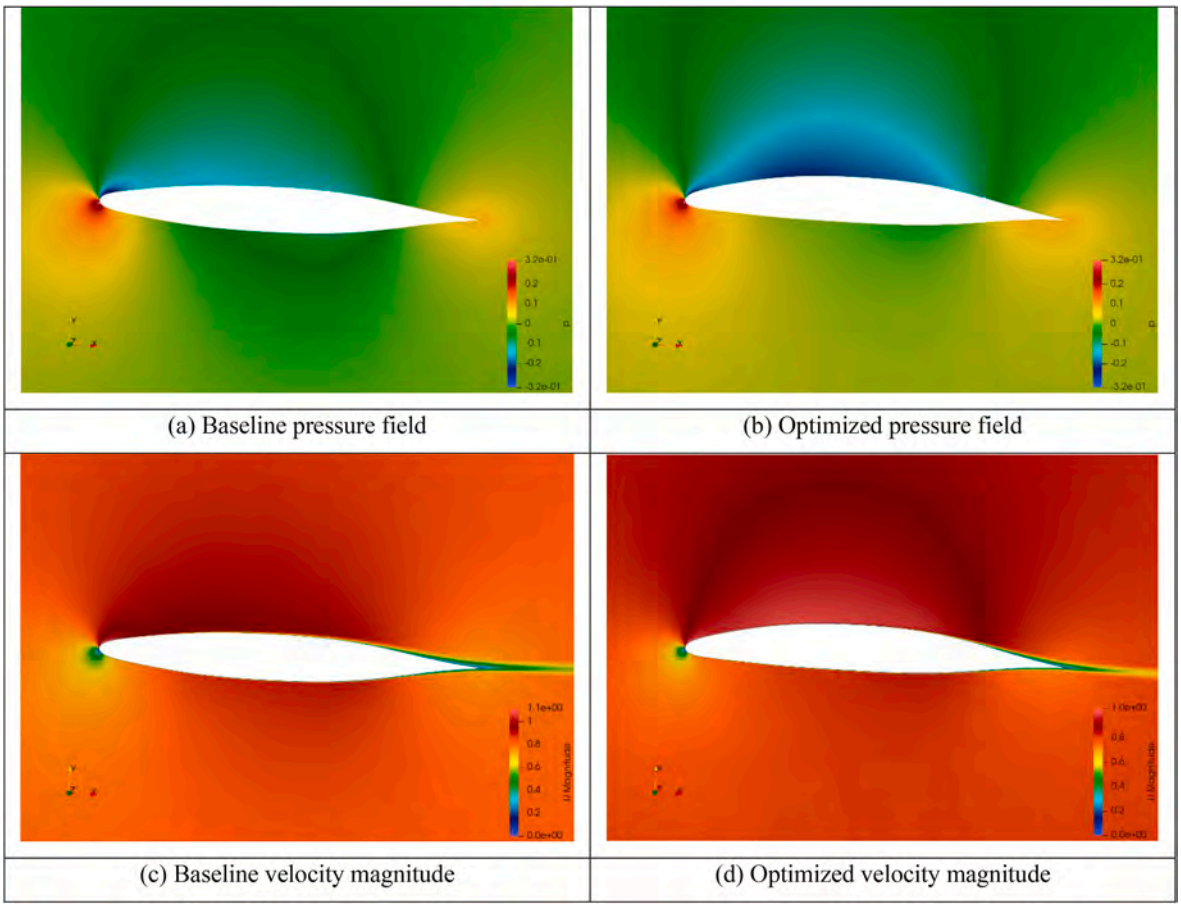

Fig. 12. Comparison of pressure and velocity fields between baseline and optimized hydrofoil.

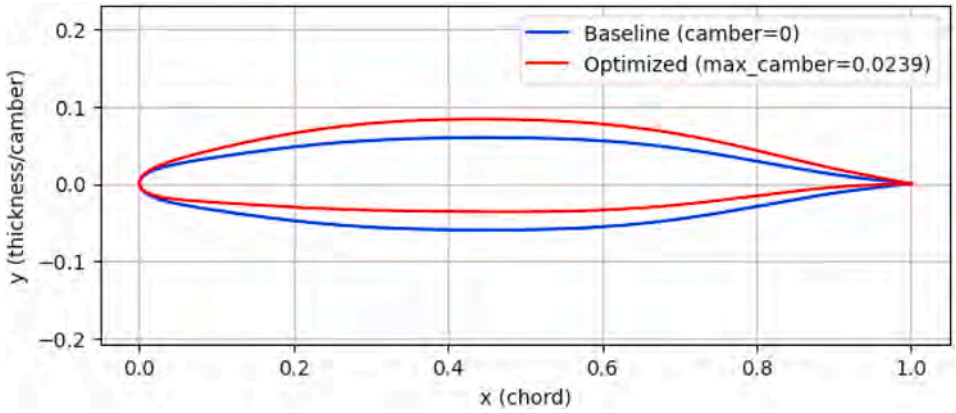

Fig. 13. Geometric comparison between baseline and optimized hydrofoils. The optimized shape introduces a maximum camber of 0.0239 while preserving the original thickness distribution, enabling higher lift with minimal drag penalty.

nonlinear, and residual subnetworks.

During Bayesian optimization of the mathematical benchmark functions, AGMF-Net rapidly converged toward the global optimum. The LogEI acquisition function effectively guided the model to focus exploitation in regions with the highest expected improvement, leading to the lowest mean absolute relative error (MARE) of all surrogates. However, this exploitation-oriented behavior also came with tradeoffs. Although AGMF-Net maintained excellent local accuracy near the optimum, its global predictive accuracy during infill was occasionally slightly lower than other test models, which achieved marginally better overall RMSE and r^2 in later iterations due to more balanced exploration across the design space. This pattern highlights that while AGMF-Net can quickly refine predictions around optimal regions, it may be less effective at globally capturing secondary features if exploration is not sufficiently promoted by the acquisition strategy.

In the hydrofoil optimization study, AGMF-Net demonstrated substantial practical value by efficiently identifying a subtle camber modification that improved the lift-to-drag ratio by 41.6 %. This outcome confirmed that the model's strengths—namely, its adaptive gating mechanism and ensemble-driven uncertainty quantification—translate effectively to real-world problems involving computationally expensive simulations. The ability to achieve significant design

improvements using only a limited number of high-fidelity CFD evaluations underscores AGMF-Nets promise for surrogate-assisted optimization in industrial settings.

Overall, these results validate the research objective of developing a general-purpose, data-efficient multi-fidelity surrogate modeling framework that can exploit prior knowledge and adaptively focus learning in critical regions of the design space. While this study has demonstrated AGMF-Net's effectiveness in ocean engineering optimization scenarios, its architecture is broadly applicable to other domains where high-fidelity data are limited, including materials discovery, structural analysis, and computational biology. Future work will extend AGMF-Net to higher-dimensional and multi-objective problems and explore advanced acquisition strategies, such as entropy-based and multi-task formulations, to improve the balance between exploration and exploitation. Additionally, we plan to benchmark the framework on diverse datasets beyond fluid dynamics to further demonstrate its robustness, scalability, and versatility in real-world surrogate modeling and optimization tasks.

CRediT authorship contribution statement

Passakorn Paladaechanan: Writing – original draft, Visualization, Validation, Investigation, Formal analysis, Data curation. **Yao Zhong:** Writing – review & editing, Validation, Investigation, Formal analysis, Data curation. **Maokun Ye:** Writing – review & editing, Visualization, Validation, Investigation, Formal analysis, Data curation. **Decheng Wan:** Writing – review & editing, Supervision, Software, Project administration, Investigation, Funding acquisition, Conceptualization. **Moustafa Abdel-Maksoud:** Writing – original draft, Visualization, Validation, Investigation, Data curation.

Ethics approval and consent to participate

Not applicable.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (52131102), to which the authors are most grateful.

Appendix A. Bayesian Optimization (BO)

Let $\mathscr{X}$ denote the domain of the k design variables, where $\mathscr{X}\subset \mathrm{R}^{\mathrm{k}}$, and let $f:\mathscr{X}\to \mathrm{R}$ be the objective function. The goal of optimization is to systematically search $\mathscr{X}$ for a solution vector $\mathbf{x}^{*}\in\mathscr{X}$ that achieves the global optimum $f^{*}=f(\mathbf{x}^{*})$. Depending on the problem formulation, this corresponds to either minimizing or maximizing f. In this study, we formulate all derivations within the minimization framework. Readers should appropriately adapt the expressions when applying them to maximization problems.

For a minimization task, the global optimum is defined as

$$\mathbf{x}^{*}\in \underset{\mathbf{x}\in\mathscr{X}}{\operatorname{argmin}} f(\mathbf{x});\quad f^{*}=\min_{\mathbf{x}\in\mathscr{X}} f(\mathbf{x})=f(\mathbf{x}^{*}) \tag{A.1}$$

Since f is treated as a black-box function that is expensive to evaluate, we approximate it using a surrogate model $\widehat{f}$ constructed from discrete observations of f. This surrogate provides inexpensive predictions of the objective value over the design space $\mathscr{D}$. We begin by uniformly sampling $\mathscr{D}$ to evaluate f at selected points, forming an initial dataset. We then train a probabilistic surrogate model $\widehat{f}$ on this data.

The surrogate yields a predictive posterior distribution, whose mean approximates f and whose variance informs the acquisition function used to guide infill sampling. In this work, we employ the logarithmic Expected Improvement (logEI) acquisition function due to its numerical stability (Ament et al., 2023). Given the posterior mean $\mu(\mathscr{X})$, standard deviation $\sigma(\mathscr{X})$, and the current best observed value y^{*}, the logEI is defined as

$$\log \mathrm{EI}_{y^{*}}(\mathscr{X})=\log_{h}\left(\frac{\mu(\mathscr{X})-y^{*}}{\sigma(\mathscr{X})}\right)+\log(\sigma(\mathscr{X})), \tag{A.2}$$

(following (Ament et al., 2023)), where

$$h(z)=\phi(z)+z\,\boldsymbol{\Phi}(z), \tag{A.3}$$

and ϕ, $\boldsymbol{\Phi}$ denote the standard Normal density and distribution functions, respectively, and $h(z)=\phi(z)+z\,\boldsymbol{\Phi}(z)$ is the classical EI factor (Jones et al., 1998). The function $\log_h$ is mathematically equivalent to $\log\circ h$ and is evaluated using the following numerically stable approximation:

$$\log_{h}(z)=\begin{cases}\log\left(\phi(z)+z\,\boldsymbol{\Phi}(z)\right) & z>-1,\\ -\dfrac{z^{2}}{2}-c_{1}+\log 1\mathrm{mexp}\Big(\log\Big(\mathrm{erfcx}\Big(-z\Big/\sqrt{2}\Big)\cdot|z|\Big)+c_{2}\Big)-1\Big/\sqrt{\epsilon} & <z\le -1,\\ -\dfrac{z^{2}}{2}-c_{1}-2\,\log(|z|) & z\le -1\Big/\sqrt{\epsilon},\end{cases} \tag{A.4}$$

with constants $c_1=\log(2\pi)/2$, $c_2=\log(\pi/2)/2$, and ϵ denoting numerical precision. Functions log1mexp and erfcx represent numerically stable implementations of $\log(1-\exp(z))$ and $\exp(z^{2})\mathrm{erfc}(z)$, respectively.

Directly evaluating $\log_h(z)=\log\left(\phi(z)+z\,\boldsymbol{\Phi}(z)\right)$ can be numerically fragile. In the extreme left tail ($z\ll 0$), both $\phi(z)$ and $\boldsymbol{\Phi}(z)$ become vanishingly small (e.g., $\phi(-30)\approx e^{-450}/\sqrt{2\pi}$), so a naive sum underflows to zero and $\log_h(z)$ collapses to -∞, destroying gradients. In this regime we rely on the Mills' ratio asymptotic $\boldsymbol{\Phi}(z)\sim\phi(z)/(-z)$, which yields the stable approximation $\log_h(z)\approx -\frac{z^{2}}{2}-c_1-2\,\log(|z|)$ with $c_1=\log(2\pi)/2$. Around the transition region ($-1/\sqrt{\epsilon}<z\le -1$) we have $h(z)=\phi(z)-|z|\,\boldsymbol{\Phi}(z)$, i.e., the difference of two small terms, so direct subtraction causes catastrophic cancellation. We therefore rewrite $\boldsymbol{\Phi}(z)$ using the scaled complementary error function, $\boldsymbol{\Phi}(z)=\frac{1}{2}e^{-z^{2}/2}\mathrm{erfcx}\left(-z/\sqrt{2}\right)$, and evaluate in log-space as

$\log_h(z) = \log \phi(z) + \log 1\text{mexp}(A)$ with $A = \log\left(\text{erfcx}\left(-z/\sqrt{2}\right)\cdot|z|\right) + c_2 \leq 0$ and $c_2 = \log(\pi/2)/2$. For $z > -1$ we evaluate $\log(\phi(z) + z\,\Phi(z))$ directly, which is stable in double precision. This piecewise evaluation, summarized in Eq. (A.4), follows Ament et al. (2023) and prevents underflow/overflow and cancellation across all z, preserving well-behaved values and gradients for acquisition optimization.

To identify the next candidate point $\mathbf{x}_{cand} \in \mathscr{D}$, we maximize $\log\text{EI}(\mathscr{X})$ using the L-BFGS-B algorithm (Zhu et al., 1997) with multiple random restarts to avoid local optima. After evaluating $f(\mathbf{x}_{cand})$, the dataset is augmented with this new observation, and the surrogate model is retrained. This process repeats iteratively until a stopping criterion is met. In this study, the stopping rule depends on the task: for benchmark functions, we stop when the mean absolute relative error at the current predicted optimum point(s) falls below a preset threshold; for the hydrofoil (real engineering) optimization, we stop after a fixed number of expensive evaluations (evaluation budget). Upon completion, we minimize the final surrogate $\widehat{f}$ using L-BFGS-B with multiple starting points to estimate the predicted optimum $\widehat{\mathbf{x}^*} = \underset{\mathbf{x}\in\mathscr{X}}{\text{argmin}}\widehat{f}(\mathbf{x})$. Finally, the true objective value at $\widehat{\mathbf{x}^*}$ is evaluated and compared with the best initial design. The complete Bayesian optimization workflow is illustrated in Figure A.1.

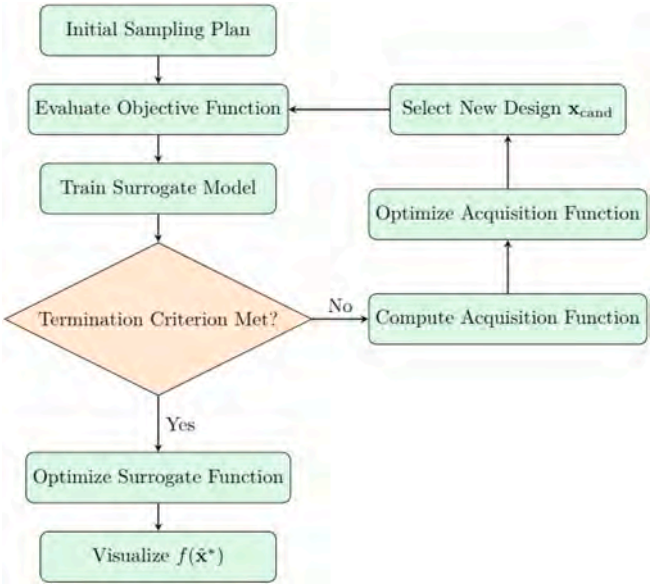

Fig. A.1. Bayesian Optimization Process.

Appendix B. Kriging

Kriging (Krige, 1951) implements a specific instance of Gaussian process regression with a characteristic kernel function:

$$k(\mathbf{x},\mathbf{x}') = \exp\left(-\sum_{j=1}^{k}\theta_j\left\|\mathbf{x}_j - \mathbf{x}_j'\right\|^{p_j}\right) \tag{A.5}$$

We define a Gaussian process by its mean and covariance functions as follows:

$$m(\mathbf{x}) = \text{E}[f(\mathbf{x})], \tag{A.6a}$$

$$k(\mathbf{x},\mathbf{x}') = \text{E}[(f(\mathbf{x}) - m(\mathbf{x}))(f(\mathbf{x}') - m(\mathbf{x}'))]. \tag{A.6b}$$

This specification leads to the GP prior over functions:

$$f(\mathbf{x}) \sim \mathscr{GP}(m(\mathbf{x}), k(\mathbf{x},\mathbf{x}')) \tag{A.7}$$

Appendix C. Multilayer Perceptron (MLP)

The multilayer perceptron (MLP) represents a standard feedforward neural network, structured with an input layer, one or more hidden layers, and an output layer. Each layer transforms its input through weighted summation followed by a non-linear activation function. Figure A.2 displays an MLP with three hidden layers.

The network computes the forward pass from layer n to $n+1$ using:

$$\mathbf{a}^{n+1} = \sigma(\mathbf{W}^n\mathbf{a}^n + \mathbf{b}^n) \tag{A.8}$$

Here, $\mathbf{a}^n$ denotes the activation vector at layer n, $\mathbf{W}^n$ is the weight matrix, $\mathbf{b}^n$ is the bias vector, and $\sigma(\cdot)$ is the activation function applied element-wise.

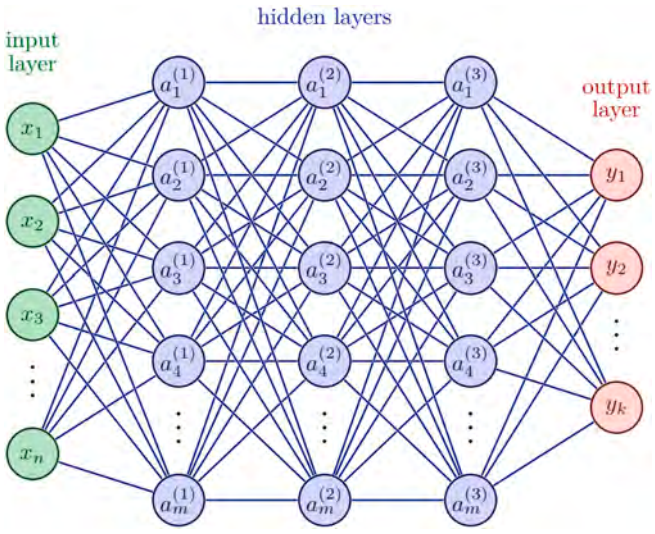

Fig. A.2. Topology of MLP with 3 Hidden Layers.

This study employs the Mish activation function (Misra, 2019), given by:

$$\sigma(\mathbf{x}) = x \cdot \tanh(\text{softplus}(x)) = x \cdot \tanh(\ln(1+e^x)) \tag{A.9}$$

An ensemble of MLPs is constructed using distinct random initializations of weights drawn from the Kaiming normal distribution (He et al., 2015):

$$\mathbf{W}_{ij} \sim \mathcal{N}\left(0, \frac{2}{\text{fan\_in}}\right) \tag{A.10}$$

All bias vectors are initialized to zero. This ensemble formulation (Lakshminarayanan et al., 2017) enables the modeling of predictive uncertainty and captures a distribution over function outputs.

Appendix D. Co-Kriging

Co-Kriging constructs a surrogate model by fusing information from both a low-fidelity function f_c and a high-fidelity function f_e, under the assumption that $f_e(\mathbf{x}) \approx \rho f_c(\mathbf{x}) + \delta(\mathbf{x})$ (Kennedy and O'Hagan, 2001), where $\rho \in \mathrm{R}$ is a scaling parameter and $\delta(\mathbf{x})$ is an independent Gaussian process capturing the discrepancy.

Let the combined training dataset be

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_c \\ \mathbf{X}_e \end{bmatrix}, \mathbf{y} = \begin{bmatrix} \mathbf{y}_c \\ \mathbf{y}_e \end{bmatrix}, \tag{A.11}$$

where $\mathbf{X}_c \in \mathrm{R}^{n_c \times k}$ and $\mathbf{X}_e \in \mathrm{R}^{n_e \times k}$ are the inputs and $\mathbf{y}_c, \mathbf{y}_e$ are the corresponding outputs.

Define the kernel matrices:

$$K_{cc} = k_c(\mathbf{X}_c, \mathbf{X}_c), \tag{A.12}$$

$$K_{ce} = k_c(\mathbf{X}_c, \mathbf{X}_e), \tag{A.13}$$

$$K_{ec} = k_c(\mathbf{X}_e, \mathbf{X}_c), \tag{A.14}$$

$$K_{ee} = \rho^2 k_c(\mathbf{X}_e, \mathbf{X}_e) + k_d(\mathbf{X}_e, \mathbf{X}_e), \tag{A.15}$$

where k_c and k_d are the ARD generalized exponential kernels defined in Eq. (A.5), used for the low- and discrepancy processes, respectively.

The full covariance matrix becomes:

$$K = \begin{bmatrix} k_c(\mathbf{X}_c, \mathbf{X}_c) & \rho k_c(\mathbf{X}_c, \mathbf{X}_e) \\ \rho k_c(\mathbf{X}_e, \mathbf{X}_c) & \rho^2 k_c(\mathbf{X}_e, \mathbf{X}_e) + k_d(\mathbf{X}_e, \mathbf{X}_e) \end{bmatrix}. \tag{A.16}$$

Let $\mathbf{c}$ denote the cross-covariance vector between the test point x and all training points:

$$\mathbf{c} = \begin{bmatrix} k_c(\mathbf{x}, \mathbf{X}_c) \\ \rho k_c(\mathbf{x}, \mathbf{X}_e) + k_d(\mathbf{x}, \mathbf{X}_e) \end{bmatrix}. \tag{A.17}$$

Then the predictive mean and variance are given by:

$$\widehat{f}(\mathbf{x}) = \mu + \mathbf{c}^\top K^{-1}(\mathbf{y} - \mu \mathbf{1}), \tag{A.18a}$$

$$\widehat{\sigma}^2(\mathbf{x}) = \rho^2\sigma_c^2 + \sigma_d^2 - \mathbf{c}^\top K^{-1}\mathbf{c} + \frac{\left(1 - \mathbf{1}^\top K^{-1}\mathbf{c}\right)^2}{\mathbf{1}^\top K^{-1}\mathbf{1}}, \tag{A.18b}$$

where μ is the common mean, and σ_c^2, σ_d^2 are the marginal variances of the low-fidelity and discrepancy processes, respectively.

Appendix E. Multi-Fidelity Neural Network (MFNN)

The MFNN proposed by Meng and Karniadakis (2020) captures nonlinear correlations between low- and high-fidelity data using a composite neural architecture.

The surrogate function $\widehat{f}$ is formulated as:

$$\widehat{f}(\mathbf{X}, f_L(\mathbf{X})) = \alpha F_l(\mathbf{X}, f_L(\mathbf{X})) + (1-\alpha)F_{nl}(\mathbf{X}, f_L(\mathbf{X})), \alpha \in [0,1], \tag{A.19}$$

where f_L denotes the low-fidelity function, and F_l, F_{nl} represent the linear and nonlinear subnetworks, respectively. The scalar α serves as an adaptive gating coefficient inferred from training data.

The linear subnetwork F_l consists of a single linear layer without hidden layers or activation functions, enabling direct affine transformation. In contrast, the nonlinear subnetwork F_{nl} adopts a multilayer perceptron (MLP) structure with hidden layers and non-linear activation functions to model complex relationships. Figure A.3 illustrates the MFNN architecture. The model ensembles multiple independently initialized instances to quantify predictive uncertainty, following the same strategy as the MLP ensemble.

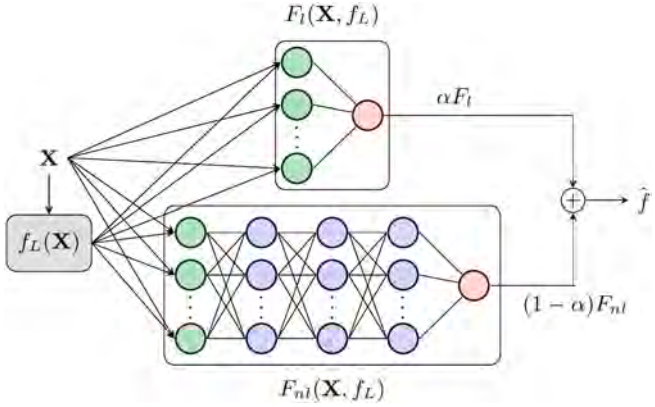

Fig. A.3. Architecture of the Multi-Fidelity Neural Network (MFNN), adapted from (Meng and Karniadakis, 2020).

Appendix F. Root Mean Square Error (RMSE)

RMSE quantified the global predictive accuracy of each surrogate. It is defined as the square root of the mean of the squared differences between predicted and true values across all test points:

$$\text{RMSE} = \sqrt{\frac{1}{n_t}\sum_{i=1}^{n_t}\left(f(x_t^{(i)}) - \widehat{f}(x_t^{(i)})\right)^2}, \tag{A.20}$$

where n_t denotes the number of test points. Lower RMSE values indicated better overall prediction accuracy.

Appendix H. Squared Pearson Correlation (r^2)

r^2 assessed the degree to which surrogate predictions correlated with the true function values. It is defined as:

$$r^2 = \left(\frac{\text{cov}(f(\mathbf{x_t}), \widehat{f}(\mathbf{x_t}))}{\sqrt{\text{var}(f(\mathbf{x_t}))\,\text{var}(\widehat{f}(\mathbf{x_t}))}}\right)^2 \tag{A.21}$$

$$= \left(\frac{n_t\sum_{i=1}^{n_t} f(x_t^{(i)})\,\widehat{f}(x_t^{(i)}) - \left(\sum_{i=1}^{n_t} f(x_t^{(i)})\right)\left(\sum_{i=1}^{n_t}\widehat{f}(x_t^{(i)})\right)}{\sqrt{\left[n_t\sum_{i=1}^{n_t}\left(f(x_t^{(i)})\right)^2 - \left(\sum_{i=1}^{n_t} f(x_t^{(i)})\right)^2\right]\left[n_t\sum_{i=1}^{n_t}\left(\widehat{f}(x_t^{(i)})\right)^2 - \left(\sum_{i=1}^{n_t}\widehat{f}(x_t^{(i)})\right)^2\right]}}\right)^2. \tag{A.22}$$

This metric focused on the similarity of the functional shape rather than absolute value agreement. Values of r^2 closer to 1 implied strong shape correspondence.

Appendix I. Mean Absolute Relative Error (MARE)

MARE specifically evaluated the prediction quality at the function's optimal point. It was computed as:

$$\mathrm{MARE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{f(x_i^*) - \widehat{f}(x_i^*)}{f(x_i^*)}\right|. \tag{A.23}$$

This metric emphasized local accuracy around the predicted optimum, complementing the global metrics RMSE and r^2.

Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

Ament, S., Daulton, S., Eriksson, D., et al., 2023. Unexpected improvements to expected improvement for bayesian optimization. Adv. Neural Inf. Process. Syst. 36, 20577–20612.

Aye, C.M., Wansaseub, K., Kumar, S., et al., 2023. Airfoil shape optimisation using a multi-fidelity surrogate-assisted metaheuristic with a new multi-objective infill sampling technique. CMES-Computer Model. Eng. Sci. 137 (3).

Bonfiglio, L., Perdikaris, P., Brizzolara, S., et al., 2018. Multi-fidelity optimization of super-cavitating hydrofoils. Comput. Methods Appl. Mech. Eng. 332, 63–85.

Cheng, J., Wang, L., Jin, H., et al., 2025. Attention-based multi-fidelity deep neural network for efficient estimation of welding residual stresses in V-shaped butt-welded high strength steel plate. Expert Syst. Appl. 266, 126137.

Ciarlatani, M.F., Gorlé, C., 2025. A (co-) kriging multi-fidelity framework for wind loading predictions. J. Build. Eng., 112940

Conti, P., Guo, M., Manzoni, A., et al., 2023. Multi-fidelity surrogate modeling using long short-term memory networks. Comput. Methods Appl. Mech. Eng. 404, 115811.

Dong, H., Song, B., Wang, P., et al., 2015. Multi-fidelity information fusion based on prediction of kriging. Struct. Multidiscip. Optim. 51 (6), 1267–1280.

Drela, M., 1989. XFOIL: an analysis and design system for low reynolds number airfoils [C]. In: Low Reynolds Number Aerodynamics: Proceedings of the Conference Notre Dame, Indiana, USA, 5–7 June 1989. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 1–12.

Drela, M., Youngren, H., 2001. XFOIL 6.9 User Primer[EB/OL], 11-30[2025-08-23]. https://web.mit.edu/drela/Public/web/xfoil/xfoil_doc.txt.

Fare, C., Fenner, P., Benatan, M., et al., 2022. A multi-fidelity machine learning approach to high throughput materials screening. npj Comput. Mater. 8 (1), 257.

Forrester, A., Sobester, A., Keane, A., 2008. Engineering Design via Surrogate Modelling: a Practical Guide[M]. John Wiley & Sons.

Franco, J., 2008. Exploratory Designs for Computer Experiments of Complex Physical Systems Simulation. HAL, p. 2008.

Halder, R., Fidkowski, K.J., Maki, K.J., 2022. Non-intrusive reduced-order modeling using convolutional autoencoders. Int. J. Numer. Methods Eng. 123 (21), 5369–5390.

He, K., Zhang, X., Ren, S., et al., 2015. Delving deep into rectifiers: surpassing human-level performance on imagenet classification[C]. Proceedings of the IEEE Int. conf. Comput. vision 1026–1034.

Jacobs, R.A., Jordan, M.I., Nowlan, S.J., et al., 1991. Adaptive mixtures of local experts. Neural Comput. 3 (1), 79–87.

Jeanmasson, G., Mary, I., Mieussens, L., 2018. Explicit local time stepping scheme for the unsteady simulation of turbulent flows[C]. In: ICCFD10-Tenth International Conference on Computational Fluid dynamics-Barcelona (Spain).

Jones, D.R., Schonlau, M., Welch, W.J., 1998. Efficient global optimization of expensive black-box functions. J. Global Optim. 13, 455–492.

Kandasamy, K., Dasarathy, G., Schneider, J., et al., 2017. Multi-fidelity bayesian optimisation with continuous approximations. In: International Conference on Machine Learning. PMLR, pp. 1799–1808.

Kennedy, M.C., O'Hagan, A., 2000. Predicting the output from a complex computer code when fast approximations are available. Biometrika 87 (1), 1–13.

Kennedy, M.C., O'Hagan, A., 2001. Bayesian calibration of computer models. J. Roy. Stat. Soc. B 63 (3), 425–464.

Kermeen, R.W., Plesset, M.S., 1956. THE NACA 661-012 HYDROFOIL IN.

Krige, D.G., 1951. A statistical approach to some basic mine valuation problems on the witwatersrand. J. S. Afr. Inst. Min. Metall 52 (6), 119–139.

Kulfan, B.M., 2008. Universal parametric geometry representation method. J. Aircraft 45 (1), 142–158.

Lakshminarayanan, B., Pritzel, A., Blundell, C., 2017. Simple and scalable predictive uncertainty estimation using deep ensembles. Adv. Neural Inf. Process. Syst. 30.

Langtry, R.B., Menter, F.R., 2009. Correlation-based transition modeling for unstructured parallelized computational fluid dynamics codes. AIAA J. 47 (12), 2894–2906.

Le Gratiet, L., Garnier, J., 2014. Recursive co-kriging model for design of computer experiments with multiple levels of fidelity. Int. J. Uncertain. Quantification 4 (5).

Li, Z., Kovachki, N., Azizzadenesheli, K., et al., 2020. Fourier neural operator for parametric partial differential equations. arXiv preprint arXiv:2010.08895.

Li, J., Li, Y., Liu, T., et al., 2023. Multi-fidelity graph neural network for flow field data fusion of turbomachinery. Energy 285, 129405.

Lu, L., Dao, M., Kumar, P., et al., 2020. Extraction of mechanical properties of materials through deep learning from instrumented indentation. Proc. Natl. Acad. Sci. 117 (13), 7052–7062.

Lu, L., Jin, P., Pang, G., et al., 2021. Learning nonlinear operators via DeepONet based on the universal approximation theorem of operators. Nat. Mach. Intell. 3 (3), 218–229.

Meng, X., Karniadakis, G.E., 2020. A composite neural network that learns from multi-fidelity data: application to function approximation and inverse PDE problems. J. Comput. Phys. 401, 109020.

Meng, X., Babaee, H., Karniadakis, G.E., 2021. Multi-fidelity Bayesian neural networks: Algorithms and applications. J. Comput. Phys. 438, 110361.

Misra, D., 2019. Mish: a self regularized non-monotonic activation function. arXiv preprint arXiv:1908.08681.

Mukhopadhaya, J., Whitehead, B.T., Quindlen, J.F., et al., 2020. Multi-fidelity modeling of probabilistic aerodynamic databases for use in aerospace engineering. Int. J. Uncertain. Quantification 10 (5).

Neal, R.M., 2012. Bayesian Learning for Neural Networks[M]. Springer Science & Business Media.

Novais, H.C., da Silva, S., Figueiredo, E., 2024. Co-Kriging strategy for structural health monitoring of bridges. Struct. Health Monit., 14759217241265375

Peherstorfer, B., Willcox, K., Gunzburger, M., 2018a. Survey of multifidelity methods in uncertainty propagation, inference, and optimization. SIAM Rev. 60 (3), 550–591.

Peherstorfer, B., Willcox, K., Gunzburger, M., 2018b. Survey of multifidelity methods in uncertainty propagation, inference, and optimization. SIAM Rev. 60 (3), 550–591.

Pellegrini, F., Roman, J., 1996. Scotch: a software package for static mapping by dual recursive bipartitioning of process and architecture graphs[C]. In: High-Performance Computing and Networking: International Conference and Exhibition HPCN EUROPE 1996 Brussels, Belgium, April 15–19, 1996 Proceedings 4. Springer Berlin Heidelberg, pp. 493–498.

Perdikaris, P., Raissi, M., Damianou, A., et al., 2017. Nonlinear information fusion algorithms for data-efficient multi-fidelity modelling. Proc. R. Soc. A 473 (2198), 20160751.

Raissi, M., Perdikaris, P., Karniadakis, G.E., 2019. Physics-informed neural networks: a deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. J. Comput. Phys. 378, 686–707.

Sabanza-Gil, V., Barbano, R., Gutiérrez, D.P., et al., 2024. Best practices for multi-fidelity bayesian optimization in materials and molecular research. arXiv preprint arXiv:2410.00544.

Shahriari, B., Swersky, K., Wang, Z., et al., 2015. Taking the human out of the loop: a review of Bayesian optimization. Proc. IEEE 104 (1), 148–175.

Shazeer, N., Mirhoseini, A., Maziarz, K., et al., 2017. Outrageously large neural networks: the sparsely-gated mixture-of-experts layer. arXiv preprint arXiv:1701.06538.

Toal, D.J.J., 2015. Some considerations regarding the use of multi-fidelity kriging in the construction of surrogate models. Struct. Multidiscip. Optim. 51 (6), 1223–1245.

Van Leer, B., 1979. Towards the ultimate conservative difference scheme. V. A second-order sequel to Godunov's method. J. Comput. Phys. 32 (1), 101–136.

Weller, H.G., Tabor, G., Jasak, H., et al., 1998. A tensorial approach to computational continuum mechanics using object-oriented techniques. Comput. Phys. 12 (6), 620–631.

Williams, C.K.I., Rasmussen, C.E., 2006. Gaussian Processes for Machine Learning[M]. MIT press, Cambridge, MA.

Zhan, L., Wang, Z., Chen, Y., et al., 2024. Ada2MF: dual-adaptive multi-fidelity neural network approach and its application in wind turbine wake prediction. Eng. Appl. Artif. Intell. 137, 109061. Engineering Applications of Artificial Intelligence, 2024, 137: 109061. Zhan L, Wang Z, Chen Y, et al. Ada2MF: Dual-adaptive multi-fidelity neural network approach and its application in wind turbine wake prediction[J].

Zhu, C., Byrd, R.H., Lu, P., et al., 1997. Algorithm 778: L-BFGS-B: fortran subroutines for large-scale bound-constrained optimization. ACM Trans. Math Software 23 (4), 550–560.