



# International Operational Modal Analysis Conference

20 - 23 May 2025 | Rennes, France

## Probabilistic Super-Resolution for High-Fidelity Physical System Simulations with Uncertainty Quantification

Pengyu Zhang<sup>1</sup>, Connor Duffin<sup>2</sup>, Alex Glyn-Davies<sup>3</sup>, Arnaud Vadeboncoeur<sup>4</sup> and Mark Girolami<sup>5</sup>

<sup>1</sup> University of Cambridge, UK, pz281@cam.ac.uk

<sup>2</sup> University of Cambridge, UK, connor.duffin@nhm.ac.uk

<sup>3</sup> University of Cambridge, UK, ag933@cam.ac.uk

<sup>4</sup> University of Cambridge, UK, av537@cam.ac.uk

<sup>5</sup> University of Cambridge, The Alan Turing Institute, UK, mag92@cam.ac.uk

### ABSTRACT

Super-resolution (SR) is a promising tool for generating high-fidelity simulations of physical systems from low-resolution data, enabling fast and accurate predictions in engineering applications. However, existing deep-learning based SR methods, require large labeled datasets and lack reliable uncertainty quantification (UQ), limiting their applicability in real-world scenarios. To overcome these challenges, we propose a probabilistic SR framework that leverages the Statistical Finite Element Method and energy-based generative modeling. Our method enables efficient high-resolution predictions with inherent UQ, while eliminating the need for extensive labeled datasets. The method is validated on a 2D Poisson example and compared with bicubic interpolation upscaling. Results demonstrate a computational speed-up over high-resolution numerical solvers while providing reliable uncertainty estimates.

*Keywords: Super Resolution; Uncertainty Quantification; Bayesian Methods*

### 1. INTRODUCTION

A long standing challenge in the engineering sciences is accurately modelling physical systems, most notably when these are described by partial differential equations (PDEs). High-resolution simulations are critical in fields such as automotive and structural engineering [1, 2], where precise modelling of subtle physical behaviours is essential to inform engineering decisions. However, repeated evaluations of high-fidelity simulations using traditional numerical solvers, such as the Finite Element Method (FEM), has high computational costs and significant time requirements. This limitation poses challenges in applications like optimal design, where iterative simulations across varied parameter sets are necessary

to achieve optimal configurations, making the process both slow and resource-intensive [3]. With the growing reliance on simulation-based predictions, ensuring computational efficiency alongside accuracy in high-fidelity simulations is paramount.

To address some of these challenges, researchers have proposed using super-resolution (SR) techniques, from the field of computer vision [4, 5], to learn a mapping from low-resolution (LR) images to high-resolution (HR) images. The aim is to obtain high-fidelity solutions of physical fields defined by PDEs by first performing simulations on a coarse grid and then upscaling these coarse predictions to a finer grid. Most existing SR techniques for physical simulations focus on learning the mapping using upscaling deep learning models, such as Convolutional Neural Networks (CNNs) [6, 7] and Generative Adversarial Networks (GANs) [8], augmented with physics-based constraints. However, current methodologies exhibit several critical limitations. Firstly, supervised deep learning models such as CNNs and GANs require extensive labelled HR datasets for effective training, the generation of which is computationally expensive. Secondly, existing methods also overlook *predictive uncertainty*, an essential aspect in engineering decision-making. Without accurate uncertainty quantification (UQ) in the HR predictions, the reliability of data-driven decision-making is compromised, especially in engineering systems where uncertainty directly impacts model accuracy and trustworthiness.

In light of this, we propose a Bayesian probabilistic framework for SR of physical simulations, which we call ProbSR. By utilising the statistical Finite Element Method (statFEM) [9, 10], coupled with principles from energy-based generative modelling [11], a conditional distribution of HR predictions given corresponding LR data is obtained. This methodology obviates the need for the training of a complicated upscaling neural network, thereby eliminating the need for large labelled HR datasets. Instead we directly model the residual between HR and LR fields with simpler network architectures. Moreover, the probabilistic nature of the proposed framework inherently facilitates UQ, enhancing the interpretability and reliability of SR predictions. By bypassing the computationally expensive process of simulating full HR solutions, our approach achieves efficient SR while maintaining the robustness necessary for engineering applications. The rest of this paper is organized as follows: in Section 2. we go over the statFEM methodology as well as Langevin sampling; in Section 3. we introduce the proposed methodology by discussing the model structure, the downscaling network, and the training/inference procedures; in Section 4. we show numerical results demonstrating the effectiveness of the proposed method.

## 2. THEORETICAL BACKGROUND

In this section, we introduce the relevant background, namely statFEM and Langevin dynamics for MCMC sampling of un-normalised probability distributions.

### 2.1. Statistical Finite Element Method (statFEM)

The Statistical Finite Element Method (statFEM) is an extension of classical FEM that models physical fields probabilistically. Unlike traditional FEM, which provides deterministic solutions based on discretised governing equations, statFEM focuses on incorporating observational data from engineering systems into finite element models through statistical reasoning by means of Bayesian updating. In our proposed SR method, we make use of statFEM to provide a probabilistic description of the HR field in the form of a prior distribution.

Consider a standard second-order elliptic PDE

$$\begin{aligned} -\nabla^2 u(x) &= f(x) & x \in \Omega, \\ u(x) &= 0 & x \in \partial\Omega, \end{aligned} \tag{1}$$

where  $u(x)$  represents the solution field,  $\Omega$  and  $\partial\Omega$  denote the PDE domain and boundary, and  $f(x)$  is a source term. Following the statFEM approach, we replace the right-hand side term  $f(x)$  with a Gaussian Process (GP) prior to represent model misspecification, such that  $\xi(x) \sim \mathcal{GP}(f(x), k(x, x'))$ , where the

kernel function  $k(x, x')$  models the spatial correlations of the misspecification. This yields the stochastic PDE

$$-\nabla^2 u(x) = \xi(x). \quad (2)$$

We discretise (2) with a FEM basis function expansion,  $\hat{u}(x) = \sum_{i=1}^M u_i \phi_i(x)$ , along with test functions, to obtain the prior distribution,

$$p(\mathbf{u}) = \mathcal{N}(\mathbf{u}; \mathbf{A}^{-1} \mathbf{b}, \mathbf{A}^{-1} \mathbf{G} \mathbf{A}^{-T}), \quad (3)$$

where  $\mathbf{u} = (u_1, \dots, u_M)^T$ ,  $\mathbf{A}_{ij} = \langle \nabla \phi_i, \nabla \phi_j \rangle$ ,  $\mathbf{b}_j = \langle f, \phi_j \rangle$  and  $\mathbf{G}_{ij} = \langle \phi_i, \langle k(\cdot, \cdot), \phi_j \rangle \rangle$ . GP introduces uncertainty into the PDE solution and is itself characterised by kernel parameters. Our proposed SR method employs this derived prior as the HR field's prior distribution.

## 2.2. Langevin Dynamics

Langevin dynamics [12] has emerged as a popular sampling method, which combines elements of both deterministic and stochastic processes to efficiently explore complex probability distributions. This method is based on the Langevin Stochastic Differential Equation (SDE), and holds significant interest across various scientific fields. The Langevin SDE is expressed as

$$dX_t = -\nabla U(X_t) dt + \sqrt{2} dB_t, \quad (4)$$

where  $U(\cdot)$  represents a potential function and  $(B_t)_{t \geq 0}$  is Brownian motion. When using Langevin SDEs to sample from distribution  $\pi(x)$ , one can set

$$U(x) = -\log \pi(x). \quad (5)$$

Since this is a continuous-time SDE and cannot be directly simulated, discretisation schemes are necessary. Applying a first-order Euler-Maruyama discretization [13], we obtain

$$X_{k+1} = X_k + \gamma \nabla \log \pi(X_k) + \sqrt{2\gamma} W_{k+1}, \quad (6)$$

where  $(W_k)_{k \geq 0}$  are drawn from the standard normal distribution and  $\gamma$  is a user-selected step size. Our proposed method uses Langevin dynamics as a sampling scheme to obtain draws from the distribution of HR predictions, thereby performing UQ.

## 3. PROPOSED METHODOLOGY

In this section we focus on our proposed methodology. We firstly discuss the probabilistic model structure. We then introduce the downscaling neural network which appears in the probabilistic learning formulation. Finally, we show how our overall SR framework is trained and used for inference.

### 3.1. Model Structure

We will derive the proposed methodology on the Poisson equation, but the method is readily extendable to a wide range of differential equations. We consider again (1), with varied forcing terms  $f_\theta(\mathbf{x})$  parameterised by  $\theta$ . We assume a prior distribution for the HR predictions following the structure of the statFEM prior (3)

$$p(\mathbf{u}^h | \theta) = \mathcal{N}(\mathbf{u}^h; \mathbf{A}^{-1} \mathbf{b}_\theta, \mathbf{A}^{-1} \mathbf{G} \mathbf{A}^{-T}), \quad (7)$$

where  $\mathbf{u}^h \in \mathbb{R}^{N_h}$ , which is the vector of FEM coefficients for the HR solution field. In the proposed method, we assume  $\mathbf{G} = \sigma^2 \mathbf{I}$ . In numerical solvers, the HR solution is computed by solving  $\mathbf{A} \mathbf{u}^h = \mathbf{b}_\theta$ , involving the computationally expensive solving of a system of equations with cost scaling rapidly

with the resolution. However, our proposed method circumvents the need for directly solving the linear equation, thereby enabling a more scalable computation of HR solutions. Next, we define the probability of the LR data given the latent HR solution through a downscaling network  $\mathcal{H}_\phi : \mathbb{R}^{N_h} \rightarrow \mathbb{R}^{N_l}$  as

$$p(\mathbf{u}^l | \mathbf{u}^h, \phi) = \mathcal{N}(\mathbf{u}^l; \mathcal{H}_\phi(\mathbf{u}^h), \epsilon^2 \mathbf{I}), \quad (8)$$

where  $\mathbf{u}^l \in \mathbb{R}^{N_l}$ . The training dataset is comprised of only LR data generated from the given equation using various forcing terms  $f_\theta(x)$ . We select  $\phi$  by maximizing the marginal likelihood of LR data,

$$p(\mathbf{u}^l | \phi, \theta) = \int p(\mathbf{u}^l | \mathbf{u}^h, \phi) p(\mathbf{u}^h | \theta) d\mathbf{u}^h. \quad (9)$$

Using Fisher's identity, the gradient of the log-marginal likelihood is given by

$$\nabla_\phi \log p(\mathbf{u}^l | \phi, \theta) = \int \left[ \nabla_\phi \log p(\mathbf{u}^l | \mathbf{u}^h, \phi) \right] p(\mathbf{u}^h | \mathbf{u}^l, \phi, \theta) d\mathbf{u}^h. \quad (10)$$

Approximating this integral using Monte Carlo sampling, we have

$$\nabla_\phi \log p(\mathbf{u}^l | \phi, \theta) \approx \frac{1}{M} \sum_{m=1}^M \nabla_\phi \log p(\mathbf{u}^l | \mathbf{u}^{h,(m)}, \phi) \quad \mathbf{u}^{h,(m)} \sim p(\mathbf{u}^h | \mathbf{u}^l, \phi, \theta). \quad (11)$$

Since directly sampling from  $p(\mathbf{u}^h | \mathbf{u}^l, \phi, \theta)$  is intractable, we use Langevin sampling (6) to obtain posterior samples as

$$\mathbf{u}_{k+1}^h = \mathbf{u}_k^h + \gamma \nabla_{\mathbf{u}^h} \log p(\mathbf{u}_k^h | \mathbf{u}^l, \phi, \theta) + \sqrt{2\gamma} W_{k+1}, \quad (12)$$

and selecting  $m = 1, \dots, M$  samples from this chain. We can expand the log gradient term as

$$\nabla_{\mathbf{u}^h} \log p(\mathbf{u}^h | \mathbf{u}^l, \phi, \theta) = \nabla_{\mathbf{u}^h} \log p(\mathbf{u}^l | \mathbf{u}^h, \phi, \theta) + \nabla_{\mathbf{u}^h} \log p(\mathbf{u}^h | \theta). \quad (13)$$

The first term in (13) is the gradient of the likelihood, which can be computed using automatic differentiation (autograd) based on (8). The second term is the gradient of the prior, given in (7). After rearranging the equation, we avoid the inversion of matrix  $\mathbf{A}$  to speed up the procedure,

$$\log p(\mathbf{u}^h | \theta) \propto -\frac{1}{2} (\mathbf{A}\mathbf{u}^h - \mathbf{b}_\theta)^T \mathbf{G}^{-1} (\mathbf{A}\mathbf{u}^h - \mathbf{b}_\theta). \quad (14)$$

Thus, the gradient of log prior is

$$\nabla_{\mathbf{u}^h} \log p(\mathbf{u}^h | \theta) = -\mathbf{A}^T \mathbf{G}^{-1} (\mathbf{A}\mathbf{u}^h - \mathbf{b}_\theta). \quad (15)$$

This expression only includes multiplication of the matrix  $\mathbf{A}$ , and  $\mathbf{G}^{-1}$  is diagonal, which is easy to compute. By combining the gradients of the log-likelihood and log-prior into (13), we can perform Langevin sampling according to (6) to obtain the necessary posterior samples.

### 3.2. Downscaling Network

The architecture of the downscaling network consists of downsampling CNNs designed to learn the residuals between the ground truth LR data and the bicubic downsampled HR data. This approach enables the network to capture fine-grained details lost during conventional downsampling processes. The network is formulated as follows,

$$\mathbf{u}^l = \mathcal{H}_\phi(\mathbf{u}^h) = \text{Bicubic Interpolation}^1(\mathbf{u}^h) + \mathcal{F}_\phi(\mathbf{u}^h), \quad (16)$$

where  $\mathcal{F}_\phi : \mathbb{R}^{N_h} \rightarrow \mathbb{R}^{N_l}$ , which is the downscaling CNN composed of multiple convolutional layers with nonlinear activation functions and max-pooling layers, designed to progressively extract hierarchical features from the input HR data while reducing spatial dimensions.

<sup>1</sup>The bicubic interpolation function used is from package *torch.nn.functional.interpolate* [14].

### 3.3. Training and Inference Procedure

#### 3.3.1. Training

During the training process, the network parameters of the downscaling model,  $\phi$ , need to be optimised. We first need to generate some LR trainset by FEM using different forcing terms  $f_\theta$ . The following pseudocode outlines the training procedure,

---

#### Algorithm 1: PDE Super-resolution Algorithm

---

**input** : Epoch number  $T$ , learning rate  $\eta$ , initial downscaling model parameters  $\phi_0$ , high-resolution prior  $p(\mathbf{u}^h|\theta)$ , low-resolution data  $\{\mathbf{u}^{l,(i)}\}_{i=1}^n$ , batch size  $m$ , posterior sampling step number  $K$  and step size  $s$ .  
**output**: Downscaling model parameters  $\phi_T$

**for**  $t = 0 : T - 1$  **do**

- Mini-Batch**: Sample low-resolution examples  $\{\mathbf{u}^{l,(i)}\}_{i=1}^m$ ;
- Posterior Sampling**: For each  $\mathbf{u}^{l,(i)}$ , sample  $\mathbf{u}^{h,(i)} \sim p(\mathbf{u}^h|\mathbf{u}^l, \phi, \theta)$  via Langevin dynamics following (12);
- Updating parameters**:  $\phi_{t+1} = \phi_t + \eta \frac{1}{m} \sum_{i=1}^m \nabla_\phi \log p(\mathbf{u}^{l,(i)}|\mathbf{u}^{h,(i)}, \phi)$

---

#### 3.3.2. Inference

During the inference phase, we generate HR solutions given a specific LR input by sampling from the posterior  $p(\mathbf{u}^h|\mathbf{u}^l, \phi, \theta)$ . Thus, once we have trained the downscaling network  $\mathcal{H}_\phi$ , the SR task can be performed following (12) until stationarity is reached. This approach allows us to estimate both the mean and covariance of our predictions, thereby enabling UQ within the framework.

## 4. NUMERICAL EXPERIMENTS

We now present the numerical experiments and comparisons of ProbSR in a series of test cases.

### 4.1. Problem setup

Take the Poisson problem, where both Dirichlet and Neumann boundary conditions are present,

$$-\nabla^2 u(\mathbf{x}) = f_\theta(\mathbf{x}) \quad x, y \in [-3, 3], \quad (17)$$

$$u(\mathbf{x}) = d \quad y \in \{-3, 3\}, \quad (18)$$

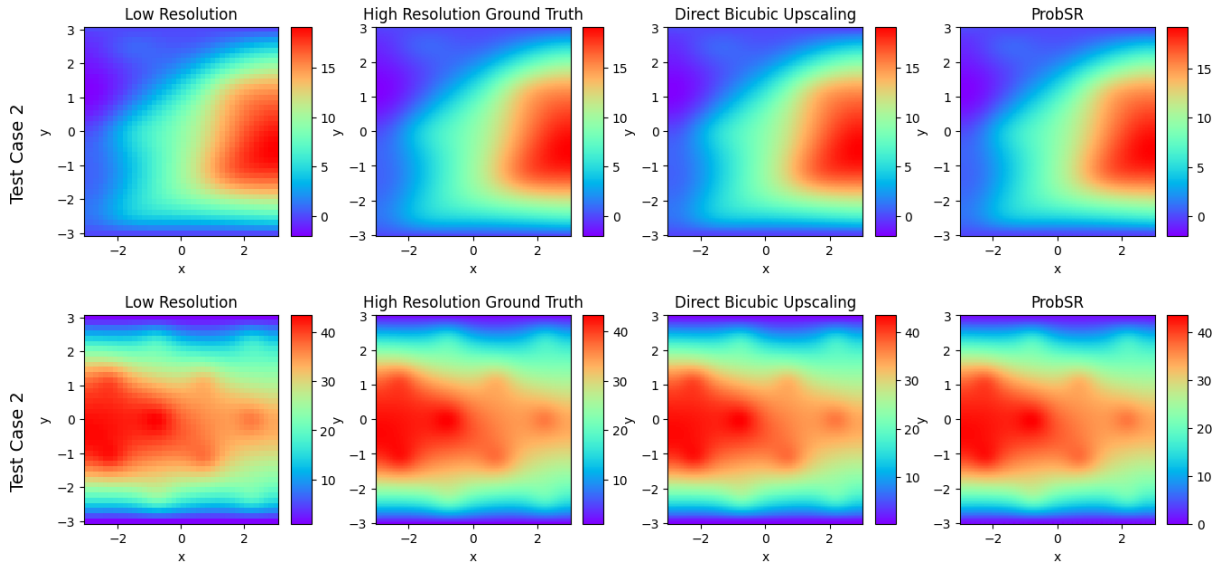
$$\partial_{\mathbf{x}}(u(\mathbf{x})) = 0 \quad x \in \{-3, 3\}, \quad (19)$$

where  $f_\theta(\mathbf{x}) = a \sin(bx) \cos(cy) + b \cos(ax) \sin(cy) + ce^{a \cos(bx) \sin(cy)} + \frac{ax^3 - by^3}{x^2 + cy^2 + 1}$ . In this case,  $a, b, c$ , are in the forcing term, and  $d$  is in the Dirichlet boundary condition. A 4-fold SR task is investigated, from resolution  $N_l = l \times l$  and  $N_h = 4l \times 4l$ , where  $l$  is the number of points along the side of the square domain. We use  $l = 40$  in our experiment. The LR training dataset is generated using various source terms and boundary conditions, with  $a \sim \mathcal{U}(-4, 4)$ ,  $b \sim \mathcal{U}(-3, 3)$ ,  $c \sim \mathcal{U}(0, 3)$ ,  $d \sim \mathcal{U}(-2, 2)$ . A total of 1000 training data is randomly generated, which are split into training and testing sets following an 80/20 ratio.

### 4.2. Results

A comparative analysis is conducted between the proposed ProbSR method and the conventional bicubic interpolation. The results for two representative test cases are depicted in Figure 1. In Test Case 1, the parameters are set as  $a = b = -2.5$ ,  $c = 1$ ,  $d = 0$ , whereas in Test Case 2, the parameters are specified as  $a = -3.6$ ,  $b = 2.7$ ,  $c = 2.1$ ,  $d = 1$ . The Mean Squared Error (MSE) of both SR methods is reported in Table 1, detailing the errors for the two individual test cases as well as the mean error across all test

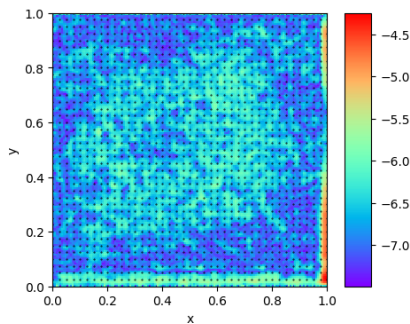
cases. The results indicate that the proposed method outperforms bicubic interpolation. We attribute this improvement to the incorporation of physical constraints in the prior distribution, which enforces consistency with the governing equation  $\mathbf{Ax} = \mathbf{b}$  and improves the accuracy of the reconstructed HR predictions. Next, we assess the UQ of the HR predictions for Test Case 1. Specifically, we plot the



**Figure 1:** The LR solution, HR ground truth solution and SR results of two test cases using direct bicubic upscaling and ProbSR.

logarithm of the standard deviation of the Langevin chain in Figure 2, where lower variance is observed at certain locations corresponding to the original LR input points. To further investigate this phenomenon, we overlay the LR grid points ( $40 \times 40$ ) as black dots onto the HR uncertainty field. This suggests that in the SR task, the model exhibits higher confidence in its predictions at points mapped from the original resolution, while greater uncertainty arises in the newly interpolated regions.

Finally, the computational efficiencies of the proposed method and traditional numerical methods for solving PDEs are compared. Consider a scenario where a  $4l \times 4l$  field is required. We compare the computational cost of the traditional method, which involves solving the linear system  $\mathbf{Au} = \mathbf{b}$ , with the ProbSR approach, which first generates a LR solution on an  $l \times l$  mesh and subsequently upscales it to a  $4l \times 4l$  mesh. The efficiencies at different resolutions, ranging from  $120 \times 120$  to  $320 \times 320$ , are detailed in Table 2. The corresponding plot of running times against required resolutions is shown in Figure 3. It is evident that traditional solvers perform better at lower resolutions, but as resolution increases, their computational cost grows significantly faster than that of the proposed method. This is due to the need to



**Figure 2:** Logarithm of the standard deviation field overlaid with black dots representing the original  $40 \times 40$  grid.

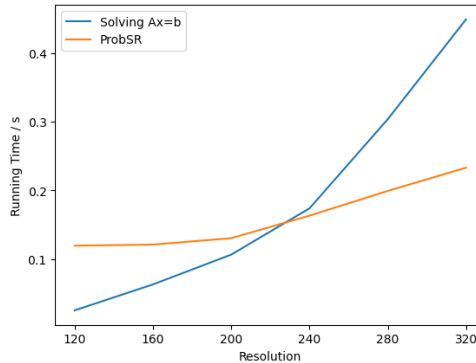
Test Case	Bicubic Interpolation	ProbSR
1	0.0470	0.0412
2	0.4663	0.4367
Mean	0.1151	0.1092

**Table 1:** MSE of SR results of bicubic interpolation and ProbSR of two test cases and the mean of all test cases.

solve the linear system  $\mathbf{A}\mathbf{u} = \mathbf{b}$ , which scales less favorably than the proposed method as the resolution increases. As a result, the proposed approach is more computationally efficient than traditional numerical methods for high-resolution problems.

Method	Resolution					
	$120 \times 120$	$160 \times 160$	$200 \times 200$	$240 \times 240$	$280 \times 280$	$320 \times 320$
Numerical solvers	0.0257 s	0.0634 s	0.1067 s	0.1740 s	0.3036 s	0.4487 s
ProbSR	0.1197 s	0.1213 s	0.1306 s	0.1634 s	0.1993 s	0.2331 s

**Table 2:** Running time of numerical solvers and ProbSR for generating different resolutions of PDE solution



**Figure 3:** Running time of two different methods for generating different resolutions of PDE solution.

## 5. CONCLUSIONS AND FUTURE WORK

This paper introduces ProbSR, a probabilistic framework for enhancing the resolution of physical system simulations. A notable advantage of the proposed method is its independence from HR training data, reducing the computational cost of training. Furthermore, as only a simple downscaling neural network is needed, this approach obviates the need for training a complex upscaling neural network similar to existing literatures, thereby simplifying the overall model architecture. The experimental results demonstrate that ProbSR outperforms traditional bicubic interpolation by leveraging physics-informed priors, which constrain the HR predictions to better satisfy the underlying governing equations. Furthermore, UQ analysis, conducted via Langevin dynamics sampling, reveals that the model exhibits lower uncertainty at locations corresponding to the original LR grid points, while interpolated regions show higher uncertainty. This finding underscores the reliability of ProbSR in preserving known physical information while acknowledging uncertainty in newly generated HR regions. Furthermore, the proposed method demonstrates a computational advantage over traditional numerical methods when high-resolution data is required. This efficiency makes ProbSR a viable alternative for high-fidelity physical simulations, particularly in applications requiring repeated evaluations. Future work will focus on extending ProbSR to more complex physical systems, incorporating additional domain-specific constraints, and improving the efficiency of the Langevin sampling process by introducing preconditioners.

## ACKNOWLEDGMENTS

AGD was supported by Splunk Inc. [G106483] Ph.D scholarship funding. CD and MG were supported by EPSRC grant EP/T000414/1. AV is supported through the EPSRC ROSEHIPS grant EP/W005816/1. MG was supported by a Royal Academy of Engineering Research Chair, and Engineering and Physical Sciences Research Council (EPSRC) grants EP/T000414/1, EP/W005816/1, EP/V056441/1, EP/V056522/1, EP/R018413/2, EP/R034710/1, and EP/R004889/1.

## REFERENCES

- [1] P. Qin, A. Ricci, and B. Blocken. Cfd simulation of aerodynamic forces on the driver car model: impact of computational parameters. *Journal of Wind Engineering and Industrial Aerodynamics*, 248:105711, 2024.
- [2] C. Hoffmann, A. Sameh, and A. Grama. High-fidelity simulation of large-scale structures. *Lecture Notes in Computer Science*, pages 664–671, 2005.
- [3] W. Wu, C. Bonneville, and C. J. Earls. A principled approach to design using high fidelity fluid-structure interaction simulations. *Finite Elements in Analysis and Design*, 194:103562, 2021.
- [4] C. Dong, C. Loy, K. He, and X. Tang. Image super-resolution using deep convolutional networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(2):295–307, 2016.
- [5] W. Shi, J. Caballero, F. Huszár, J. Totz, A. P. Aitken, R. Bishop, D. Rueckert, and Z. Wang. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1874–1883, 2016.
- [6] C. Wang, E. Bentivegna, W. Zhou, L. Klein, and B. Elmegeen. Physics-Informed Neural Network Super Resolution for Advection-Diffusion Models, 2020.
- [7] H. Gao, L. Sun, and J. Wang. Super-resolution and denoising of fluid flow using physics-informed convolutional neural networks without high-resolution labels. *Physics of Fluids*, 33(7):073603, 2021.
- [8] M. Bode, M. Gauding, Z. Lian, D. Denker, M. Davidovic, K. Kleinheinz, J. Jitsev, and H. Pitsch. Using physics-informed enhanced super-resolution generative adversarial networks for subfilter modeling in turbulent reactive flows. *Proceedings of the Combustion Institute*, 38:2617–2625, 2021.
- [9] M. Girolami, E. Febrianto, G. Yin, and F. Cirak. The statistical finite element method (statfem) for coherent synthesis of observation data and model predictions. *Computer Methods in Applied Mechanics and Engineering*, 375:113533, 2021.
- [10] C. Duffin, E. Cripps, T. Stemler, and M. Girolami. Statistical finite elements for misspecified models. *Proceedings of the National Academy of Sciences*, 118(2), 2021.
- [11] B. Pang, T. Han, E. Nijkamp, S. Zhu, and Y. Wu. Learning latent space energy-based prior model. In *Proceedings of the 34th International Conference on Neural Information Processing Systems*, Red Hook, NY, USA, 2020.
- [12] A. Durmus and É. Moulines. High-dimensional bayesian inference via the unadjusted langevin algorithm. *Bernoulli*, 25, 2019.
- [13] P. E. Kloeden and E. Platen. *Numerical Solution of Stochastic Differential Equations*. Springer, 1992.
- [14] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32, 2019.