

---

# NEURAL OPERATOR INDUCED GAUSSIAN PROCESS FRAMEWORK FOR PROBABILISTIC SOLUTION OF PARAMETRIC PARTIAL DIFFERENTIAL EQUATIONS

---

**Sawan Kumar**

Department of Applied Mechanics  
Indian Institute of Technology Delhi  
sawan.kumar@am.iitd.ac.in

**Rajdip Nayek**

Department of Applied Mechanics  
Indian Institute of Technology Delhi  
rajdipn@am.iitd.ac.in

**Souvik Chakraborty**

Department of Applied Mechanics  
Yardi School of Artificial Intelligence (ScAI)  
Indian Institute of Technology Delhi  
souvik@am.iitd.ac.in

## ABSTRACT

The study of neural operators has paved the way for the development of efficient approaches for solving partial differential equations (PDEs) compared with traditional methods. However, most of the existing neural operators lack the capability to provide uncertainty measures for their predictions, a crucial aspect, especially in data-driven scenarios with limited available data. In this work, we propose a novel Neural Operator-induced Gaussian Process (NOGaP), which exploits the probabilistic characteristics of Gaussian Processes (GPs) while leveraging the learning prowess of operator learning. The proposed framework leads to improved prediction accuracy and offers a quantifiable measure of uncertainty. The proposed framework is extensively evaluated through experiments on various PDE examples, including Burger's equation, Darcy flow, non-homogeneous Poisson, and wave-advection equations. Furthermore, a comparative study with state-of-the-art operator learning algorithms is presented to highlight the advantages of NOGaP. The results demonstrate superior accuracy and expected uncertainty characteristics, suggesting the promising potential of the proposed framework.

**Keywords** Gaussian Process, neural operators, uncertainty quantification, probabilistic machine learning.

## 1 Introduction

Neural Operators (NO)[1, 2, 3] have recently emerged as a viable alternative to traditional numerical methods for solving partial differential equations. Some recent applications of NO in computational mechanics include weather modeling [4, 5], medical diagnosis using elastography [3], fracture propagation [6], and fault diagnosis [7], among others. However, a major bottleneck in the widespread adoption of NO in critical engineering systems (e.g., nuclear reactors) resides in the fact that they fail to quantify the uncertainty in the trained model. This has been recognized by the research community, and some efforts in this direction can be found in [8, 9]. An alternative to the neural operators is to employ probabilistic machine learning algorithms such as the Gaussian Process (GP) for solving partial differential equations. Unlike NO, GP is inherently probabilistic and hence, can seamlessly quantify the uncertainty in the model. Recent developments on the use of GP for solving PDEs include [10, 11, 12]. Unfortunately, the developments on GP and NO are carried out somewhat in parallel. To that end, the objective of this paper is to investigate how NO and GP can benefit from each other, and if a fusion of the two can result in a superior model.

Neural operators learn and map two infinite-dimensional function spaces. Some neural operators that have been studied in the past and gained popularity are Wavelet Neural Operator (WNO) [3], Physics Informed WNO [13], DeepONet [2, 14, 15], Fourier Neural Operator (FNO) [1], PCA-Net [16, 17], Graph Neural Operators (GNO) [18] and many

more. Compared to classical PDE solvers, neural operators have significant computational advantages; once trained, they can make predictions for different initial and boundary conditions efficiently. The performance of both WNO and FNO is promising when compared with the state-of-the-art operator learning methods in terms of prediction accuracy. However, FNO utilizes the Fast Fourier Transform (FFT), resulting in a loss of spatial information. Consequently, frequency-localized basis functions in FFT pose challenges in learning complex domain problems. To address this variants such as geometric-FNO [19] and geometry-informed neural operators [20] have been proposed. Alternatively, WNO utilizes wavelet transform which retains space-frequency information and can handle signals with spikes and discontinuity, thus enabling superior learning of the patterns in images. Some recent improvements of the classical WNO include [21, 22, 23, 24, 13]. However, as previously stated, both WNO and FNO are deterministic and cannot yield the predictive uncertainty that is essential for decision-making in safety-critical systems.

Data in real-world scenarios are accompanied by inherent noise and sources of uncertainty. These sources include errors stemming from the limitation of precision of the measuring instrument, incorrect measurement of observations, missing data, etc. Additionally, models may be built upon simplifying assumptions that may not fully capture the complexity of the underlying phenomena, which also contributes to the uncertainty. This is often addressed using probabilistic frameworks like Gaussian Process Regression (GPR) [25, 26, 27] and Bayesian neural operator frameworks [28, 29, 30]. GPR is a non-parametric machine learning model, and by design, learns distribution over the functions. Similar to neural networks and neural operators, GPR is also a popular choice among researchers for solving partial differential equations. This includes the development of physics-informed GPR for linear and nonlinear systems [31, 32, 33]. However, GPR suffers from the curse of dimensionality and does not scale with an increase in training samples; hence, GPR is less popular as compared to neural network-based approaches.

The developments on NO and GP are mostly carried out in parallel. To that end, the objective of this paper is to propose a novel architecture referred to as the Neural Operator induced Gaussian Process (NOGaP) that attempts to bridge the gap between the NO and GP-based approaches. NOGaP ingeniously combines the strengths of Gaussian process regression with NO. Specifically, the high-level idea is to use NO as a mean function within the GP framework. We hypothesize that with NO as the mean function, NOGaP will be able to exploit the accuracy of NO and the probabilistic nature of the GP. Without loss of generality, we use the recently proposed WNO as the mean function with the proposed approach. Additionally, to address the challenge associated with the scalability of GP, we propose to use the Kronecker-product-based approach proposed in [34, 35].

The proposed framework has the following key features:

- **Accuracy:** NOGaP, a fusion of NO and GP, is anticipated to offer superior accuracy compared to both NO and GP individually. This has been empirically illustrated in the numerical section.
- **In-built uncertainty quantification capability:** NOGaP possesses built-in probabilistic capabilities, allowing it to assess the epistemic uncertainty due to limited and noisy datasets. This feature is crucial, especially for ensuring the reliability of the model in critical engineering systems.
- **Scalability:** Leveraging the Kronecker product property, NOGaP exhibits scalability, enabling its efficient handling of large datasets. This marks a departure from conventional GPs, which often encounter scalability issues due to the curse of dimensionality.

The remainder of the paper is arranged as follows: Section 2 elaborates on the mathematical formulation of the proposed framework. Section 3 explores the different numerical examples on which the NOGaP is tested along with the results. Finally, the paper is concluded in the last section 4 with results, observations and future work.

## 2 Neural Operator induced Gaussian Process (NOGaP)

Let us define the hypothesis space, where  $\mathcal{Z} \subset \mathbb{R}^d$  represents the input space and  $\mathcal{Y} \subset \mathbb{R}^o$  denotes the output space. Suppose we have  $N$  distinct pairs of observations with inputs,  $\mathbf{Z} = (\mathbf{x}_1, \dots, \mathbf{x}_N, \mathbf{a}_1, \dots, \mathbf{a}_N)$ , and a corresponding set of outputs,  $\mathbf{Y} = (\mathbf{y}_1, \dots, \mathbf{y}_N)$ . The model takes the form:

$$\begin{aligned} \mathbf{y}(\mathbf{z}) &= \mathbf{f}_o(\mathbf{z}) + \boldsymbol{\epsilon} \\ \boldsymbol{\epsilon} &\sim \mathcal{N}(\mathbf{0}, \sigma_o^2 \mathbf{I}) \end{aligned} \quad (1)$$

where  $\mathbf{y}$  are the noisy observations and  $\boldsymbol{\epsilon} \in \mathbb{R}^o$  is modeled as independent zero-mean Gaussian white noise with variance  $\sigma_o^2$ . The vector-valued function  $\mathbf{f}_o$  is assumed to follow a GP prior:

$$\mathbf{f}_o(\mathbf{z}) \sim \mathcal{GP}(\mathbf{m}_o(\mathbf{z}; \boldsymbol{\theta}_m), \mathbf{K}_o(\mathbf{z}, \mathbf{z}'; \boldsymbol{\theta}_k)) \quad (2)$$

Here we have *vector-valued* mean function  $\mathbf{m}_o : \mathbb{R}^d \mapsto \mathbb{R}^o$  and a *matrix-valued* covariance function  $\mathbf{K}_o : \mathcal{Z} \times \mathcal{Z} \mapsto \mathbb{R}^{o \times o}$ , with vector-valued parameters  $\boldsymbol{\theta}_m$  and  $\boldsymbol{\theta}_k$  where  $\boldsymbol{\theta}_m$  represents parameters of the mean function and  $\boldsymbol{\theta}_k$  denotes

kernel parameters. Here in our formulation, the multioutput kernel  $\mathbf{K}_o$  is assumed to be separable [36]. It can be expressed as:

$$\begin{aligned} (\mathbf{K}_o)_{dd'} &= k_x(\mathbf{z}, \mathbf{z}') \cdot k_f(d, d') \\ \mathbf{K}_o &= \mathbf{K}_x \otimes \mathbf{K}_f \end{aligned} \quad (3)$$

where  $\otimes$  denotes the Kronecker product,  $\mathbf{K}_o$  is the multioutput kernel in  $\mathbb{R}^{N_o \times N_o}$ ,  $k_f$  is a covariance function specifying inter-output similarities and  $k_x$  is a covariance function over inputs. In case the inputs are available on a multidimensional cartesian grid, we can further decompose the covariance matrix  $\mathbf{K}_x$  as the Kronecker product  $\mathbf{K}_{x1} \otimes \dots \otimes \mathbf{K}_{xm}$ . Exploiting the Kronecker product property can further improve computational efficiency.

## 2.1 Mean Function: Neural Operator

There are many available choices for the mean function in GP, each with its advantages and limitations. For instance, a linear mean function can capture linear trends in the data, while a non-linear mean function, such as a neural network, can handle more complex relationships and nonlinear patterns. By using a mean function that aligns with the characteristics of the problem, the GP can better model the underlying process as it only has to learn the discrepancy between the ground truth and the values computed using the mean function. This results in overall improved predictions. Thus, selecting a suitable mean function in GP is critical in improving the prediction capability, resulting in efficient learning of the underlying function. Here in this work, we have used neural operators as the mean function. The core idea of neural operators is to leverage neural networks to learn the underlying function space representation. Neural operators represent a composite function consisting of the encoder, neural network, and decoder. The encoder lifts the dimensionality of the input to a high dimensional latent space. The decoder projects the computed output back to the function space. Consider a fixed domain  $D \in (a, b)$  and  $x \in D$ . For a source term of a PDE  $f(x, t) : D \times \mathcal{T} \mapsto \mathbb{R}$ , with initial condition  $u(x, 0) : D \mapsto \mathbb{R}$  and the boundary condition is  $u(\partial D, t) : D \times \mathcal{T} \mapsto \mathbb{R}$ . The solution space is  $u(x, t) : D \times \mathcal{T} \mapsto \mathbb{R}$ , where  $t$  is the time coordinate. For a Banach space with  $\mathcal{A} := \mathcal{C}(D; \mathbb{R}^{d_1})$  and  $\mathcal{U} := \mathcal{C}(D; \mathbb{R}^{d_2})$ , suppose that

$$\mathcal{D} : \mathcal{A} \times \theta_{NN} \mapsto \mathcal{U}; \quad \theta_{NN} = \{\mathcal{W}, \mathbf{b}\} \quad (4)$$

for arbitrarily any non-linear neural operator  $\mathcal{D}$ , where  $\theta_{NN}$  is the parameter space of the neural network. If we have access to  $N$  number of pairs  $\{(a_j, u_j)\}_{j=1}^N$ , then we can approximate  $\mathcal{D}$  using the pairs of  $N$  data points. Although the proposed approach can accommodate any neural operator as the mean function, we restrict ourselves to the scenario where Wavelet Neural Operator (WNO) [3] is used as the mean function.

WNO exploits the kernel integration proposed in [18] in combination with wavelet transformation. WNO takes  $N$  input functions discretized on a grid and lifts them to a higher dimension using a densely connected layer. The output from the densely connected layer is then fed into the wavelet block, and its dimensions are preserved after the final operations of the wavelet block. Mathematically, the wavelet blocks are represented as:

$$f_{j+1}(x) = \sigma((\mathcal{K}(a(x); \phi \in \beta) * v_j(x)) + bv_j(x)), \quad j = 1, \dots, J, \quad (5)$$

where  $\sigma$  represents a non-linear activation function,  $\mathcal{K}$  denotes the wavelet integral operator,  $v_j(x)$  corresponds to the transformed function from the preceding layer, and  $b$  signifies a linear transformation. The wavelet coefficients are derived through a discrete wavelet transform (DWT) and an inverse discrete wavelet transform (IDWT) of the mother wavelet. For further details, the interested reader can refer [3].

## 2.2 Likelihood of the data

Consider  $N$  distinct observations with inputs,  $\mathbf{Z} = (\mathbf{x}_1, \dots, \mathbf{x}_N, \mathbf{a}_1, \dots, \mathbf{a}_N)^T$ , and a corresponding set of outputs,  $\mathbf{Y} = (\mathbf{y}_1, \dots, \mathbf{y}_N)^T$  as before. For brevity of representation, we consider the vectorized form of  $\mathbf{Y}$ ,  $\mathbf{q} = (y_{11}, \dots, y_{N1}, \dots, y_{1o}, \dots, y_{No})$ , where  $y_{ij}$  denotes the  $j$ -th component of the output  $\mathbf{y}_i$  corresponding to the  $i$ -th input  $\mathbf{x}_i$ . With this setup, the likelihood can be expressed as follows:

$$\mathbf{q} \mid \mathbf{f}_o, \mathbf{Z}, \theta_m, \theta_k \sim \mathcal{N}(\mathbf{f}_o(\mathbf{Z}), \sigma_o^2 \mathbf{I}) \quad (6)$$

Given the fact that the mean function is parameterized using the WNO, we set  $\theta_m = \theta_{NN}$ , and the mean function computed using WNO is represented using  $\mathbf{h}$ . In practice, we work with the Negative log-marginal likelihood (NLML),

$$\begin{aligned} \log p(\mathbf{q} \mid \mathbf{Z}, \theta) &= -\frac{1}{2}(\mathbf{q} - \mathbf{h}(\mathbf{Z}; \theta_{NN}))^T (\mathbf{K}_o(\mathbf{Z}, \mathbf{Z}'; \theta_k) + \sigma_o^2 \mathbf{I})^{-1} (\mathbf{q} - \mathbf{h}(\mathbf{Z}; \theta_{NN})) \\ &\quad - \frac{1}{2} \log |\mathbf{K}_o(\mathbf{Z}, \mathbf{Z}'; \theta_k) + \sigma_o^2 \mathbf{I}| + \text{constant}, \end{aligned} \quad (7)$$

where the  $\mathbf{I}$  denotes an identity matrix. The unknown hyperparameters  $\theta = \{\theta_{NN}, \theta_k, \sigma_o\}$  are determined by minimizing the negative log-marginalized likelihood (NLML) as given by eq. (8). The optimization problem to find the parameter  $\theta$  is expressed as follows:

$$\hat{\theta} = \arg \min_{\theta} -\log p(\mathbf{q} \mid \mathbf{Z}, \theta) \quad (8)$$

Both the parameters of the mean and covariance are optimized simultaneously. Once the training is done the hyperparameter  $\theta$  is obtained. More thorough and rigorous details can be found in Murphy [37] and Rasmussen and Williams [38]. An algorithm depicting the steps involved in training the proposed model is shown in Algorithm 1.

---

**Algorithm 1** Training: NOGaP model

---

**Input:**  $N$  training samples of input-output pairs  $\{z, u(x)\}$  where spatial points  $x \in \text{Domain}$ , and network hyperparameters.

- 1: Choose a suitable covariance function  $k(z_i, z_j)$  ▷ Eq. (12)
- 2: Set mean function as **WNO**
- 3: Initialize the kernel and mean function's hyperparameters  $\theta$
- 4: **for** Number of iteration =  $n$  **do**
- 5:   Compute the GP output  $\hat{u}(x)$  for input  $z$  ▷ Eq. (10a)
- 6:   Minimize the negative marginal log-likelihood loss  $\mathcal{L}(u, \hat{u})$  ▷ Eq. (8)
- 7:   Compute the gradients of the loss with respect to the model's hyperparameters
- 8:   Update the model's hyperparameters using a suitable optimization algorithm
- 9: **end for**
- 10: Save the trained model for later prediction

**Output:** Trained NOGaP model with optimized mean function and covariance kernel hyperparameters ( $\theta$ )

---

### 2.3 The Predictive Inference

Since we are working with Gaussian likelihood we can get the predictive distribution in its closed form. We get the following joint distribution for a new data sample  $z^*$ .

$$\begin{pmatrix} q \\ f_o^* \end{pmatrix} \sim \mathcal{N} \left( \begin{pmatrix} h(Z; \theta_{NN}) \\ h(z^*; \theta_{NN}) \end{pmatrix}, \begin{pmatrix} K_o(Z, Z; \theta_k) + \sigma_o^2 \mathbf{I} & K_o(Z, z^*; \theta_k) \\ K_o(z^*, Z; \theta_k) & K_o(z^*, z^*; \theta_k) \end{pmatrix} \right) \quad (9)$$

Here, the observed training output is denoted by  $q$ , the predictions for the test sample by  $f_o^*$ , and the mean function values for the training samples and the test sample by  $h(Z; \theta_{NN})$  and  $h(z^*; \theta_{NN})$ . Computing conditional distribution from the joint distribution is straightforward, refer [39]. Inference can be done using standard GP formulas for the mean and variance of the predictive distribution. By conditioning the joint distribution on  $q$  the posterior predictive distribution of  $f_o(z^*)$  can be expressed as:

$$p(f_o^* | Z, z^*, q) = \mathcal{N}(f_o^* | \bar{m}^*, \text{cov}(f_o^*)), \quad (10a)$$

$$\bar{m}^* = h^*(z^*; \theta_{NN}) + K_o(z^*, Z; \theta_k)(K_o(Z, Z; \theta_k) + \sigma_o^2 \mathbf{I})^{-1}(q - h(Z; \theta_{NN})), \quad (10b)$$

$$\text{cov}(f_o^*) = K_o(z^*, z^*; \theta_k) - K_o(z^*, Z; \theta_k)(K_o(Z, Z; \theta_k) + \sigma_o^2 \mathbf{I})^{-1}K_o(Z, z^*; \theta_k) \quad (10c)$$

where,  $\bar{m}^*$  represents the predictive mean of  $f_o^*$ . The covariance of  $f_o^*$ , denoted as  $\text{cov}(f_o^*)$ , is computed using the Eq. (10c). An algorithm depicting the steps involved in inference is shown in Algorithm 2.

### 2.4 Choice of the Covariance function

One key aspect of any GP model is the choice of the covariance kernel. Out of a vast range of available kernels, the Matérn family of kernels is quite popular as it offers a diverse variety of smoothness for modeling signals, and therefore it has been the choice for the numerical examples covered in Section 3. Mathematically, the family of Matérn kernels for  $\mathcal{X} \subset \mathbb{R}^d$ , constants  $\alpha > 0$  and  $h > 0$  can be expressed as

$$k_{\alpha, h}(z, z') = \frac{1}{2^{\alpha-1} \Gamma(\alpha)} \left( \frac{\sqrt{2\alpha} \|z - z'\|}{h} \right)^\alpha K_\alpha \left( \frac{\sqrt{2\alpha} \|z - z'\|}{h} \right), \quad z, z' \in \mathcal{X} \quad (11)$$

where  $\Gamma$  is the gamma function, and  $K_\alpha$  is the modified Bessel function of the second kind of order  $\alpha$ . For specific values of  $\alpha$ , we get different Matérn kernels namely, Matérn-1/2, Matérn-3/2, and Matérn-5/2. The following equation shows the mathematical expression for the Matérn-5/2:

$$k_{5/2, h}(z, z') = \left( 1 + \frac{\sqrt{5} \|z - z'\|}{h} + \frac{5 \|z - z'\|^2}{3h^2} \right) \exp \left( -\frac{\sqrt{5} \|z - z'\|}{h} \right). \quad (12a)$$

**Algorithm 2** Inference: NOGaP model

**Input:**  $N^*$  test samples  $z = \{x, a(x)\}$  where spatial points  $x \in \text{Domain}$ , and trained network hyperparameters  $\theta$ .

- 1: **for** each test sample **do**
- 2:   Compute the GP output  $\hat{u}(x)$  for input  $z$  using the trained GP model. ▷ Refer to Eq. (10a)
- 3:   Store the prediction  $\hat{u}(x)$  corresponding to the input  $z$ .
- 4: **end for**

**Output:** Predictions  $\hat{u}(x)$  corresponding to inputs  $z$  for all test samples.

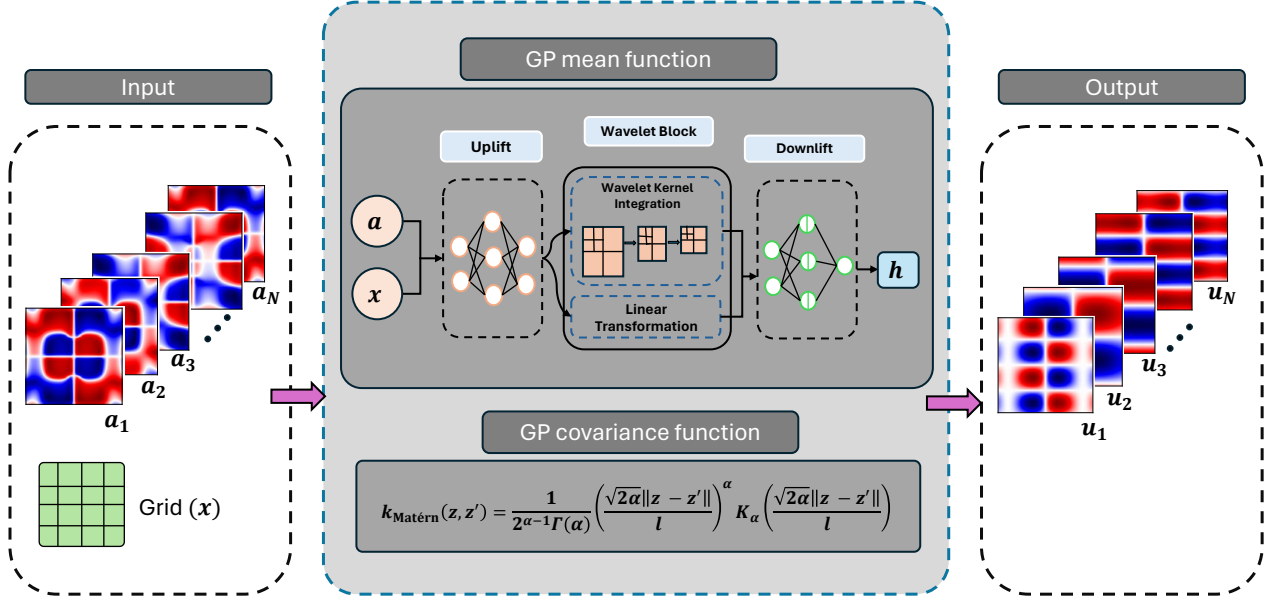


Figure 1: **Schematic of NOGaP:** The illustration depicts the NOGaP framework, featuring a Mean and Covariance block. The computation of the mean function is facilitated by the Wavelet Neural Operator (WNO) block. NOGaP operates by utilizing  $N$ -training samples comprising input-output pairs alongside the grid. The corresponding  $N$  outputs encompass both the mean of the predicted samples and the associated standard deviation.

### 3 Numerical examples

This section explores a variety of case studies ranging from fluid dynamics, gas dynamics, and phase-field modeling on which the proposed NOGaP framework is tested for its efficacy. The examples include (a) 1D Burger's equation, (b) 1D wave-advection equation, (c) 2D Darcy flow equation with a notch in the triangular domain, and (d) 2D non-homogeneous Poisson equation. The architecture of NOGaP consists of two key blocks: mean function and covariance function. The Daubechies family of mother wavelets is used for the mean function block with the GeLU which is differentiable at all points [40]. The information related to the architecture of the framework is summarised in Table 1. The framework's performance is summarised in the Table 2, displaying a comparison of relative Mean Square Errors (MSE) and the comparison with WNO, GP (zero mean) and Bayesian WNO (B-WNO) for various examples. The results obtained from NOGaP demonstrate that the proposed framework shows improved accuracy and can learn the desired mapping between input and output space effectively. The proposed NOGaP framework also provides the uncertainty estimates associated with the predictions. Further details of individual cases are illustrated in the following subsection.

#### 3.1 1D Burger Equation

The 1D Burgers equation is a well-known partial differential equation that models one-dimensional flows in various fields, including fluid mechanics, gas dynamics, and traffic flow. One of the interesting features of this equation is that it exhibits both diffusive and advective behavior, making it a useful tool for studying a wide range of phenomena. The 1D Burger equation with periodic boundary is mathematically expressed as,

Table 1: NOGaP’s architecture for each numerical example; Matérn kernel is used for covariance function in all cases.

| Examples                            | Data size |         | Wavelet | NN dimensions (mean function) |      |     |       |
|-------------------------------------|-----------|---------|---------|-------------------------------|------|-----|-------|
|                                     | Training  | Testing |         | FNN1                          | FNN2 | CNN | level |
| Burgers <sup>(3.1)</sup>            | 1000      | 50      | db6     | 64                            | 128  | 4   | 8     |
| Advection <sup>(3.2)</sup>          | 1000      | 50      | db8     | 96                            | 128  | 4   | 4     |
| Darcy (triangular) <sup>(3.3)</sup> | 500       | 50      | db6     | 32                            | 192  | 4   | 6     |
| Poisson <sup>(3.4)</sup>            | 500       | 50      | db4     | 64                            | 132  | 4   | 4     |

$$\begin{aligned}
\partial_t u(x, t) + 0.5 \partial_x u^2(x, t) &= \nu \partial_{xx} u(x, t), & x \in (0, 1), t \in (0, 1] \\
u(x = 0, t) &= u(x = 1, t), & x \in (0, 1), t \in (0, 1] \\
u(x, 0) &= u_0(x), & x \in (0, 1).
\end{aligned} \tag{13}$$

$\nu$  here is the viscosity of the flow, which is positive and greater than zero, and  $u_0(x)$  is the initial condition. For generating the initial conditions  $u_0(x)$ , a Gaussian random field is used where  $u_0(x) \sim \mathcal{N}(0, 625(-\Delta + 25I)^{-2})$ . The aim is to learn the mapping from input and output function spaces for the 1D Burger’s PDE that is  $u(x, t = 0) \mapsto u(x, t = 1)$ . We consider  $\nu = 0.1$  and a spatial resolution of 512. The dataset is taken from Ref. [1]. **Results:** The predictions for three different initial conditions are presented in Figure 2. The relative MSE values of the predictions are mentioned in Table 2, revealing that the proposed framework exhibits favorable performance for the 1D Burger’s problem. Notably, it provides an improvement over the standalone WNO, GPR (zero mean) and B-WNO predictions. A visual inspection of the plot confirms the agreement between the ground truth obtained from the numerical solver and the predictions generated by the NOGaP framework. Additionally, the plot includes the corresponding 95% confidence intervals, providing a measure of uncertainty associated with the predictions. Further, a case study is carried out by considering different training sample sizes (see Figure 3). The mean response for all the cases is found to have an excellent match with the ground truth. The predictive uncertainty, on the other hand, reduces with an increase in the number of training samples. This indicates a reduction in the predictive uncertainty with an increase in training samples.

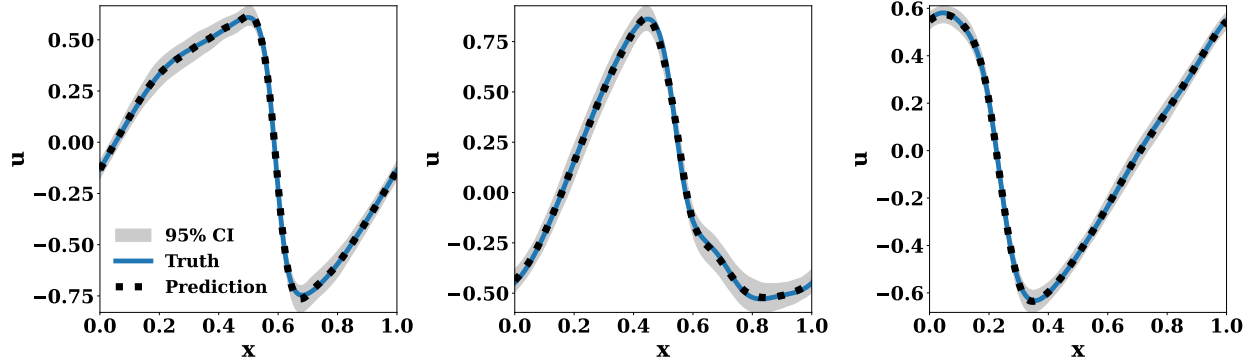


Figure 2: **1D Burger equation with periodic boundary condition.** This figure shows the ground truth and predicted solution for different initial conditions along with the 95% confidence interval (grey region) associated with each prediction. The results clearly show that NOGaP can make excellent predictions for this example.

### 3.2 1D Wave Advection Equation

The wave advection equation is a type of hyperbolic partial differential equation (PDE) that is commonly used in physics and engineering to describe the movement of a scalar under a known velocity field. The equation is often used to model the behavior of waves in various media, such as sound waves in air or water waves in the ocean. In this example, periodic boundary conditions are applied to simulate the behavior of the medium over time. When the wave advection equation is considered with periodic boundary conditions, it can be expressed mathematically as:

$$\begin{aligned}
\partial_t u(x, t) + \nu \partial_x u(x, t) &= 0, & x \in (0, 1), t \in (0, 1) \\
u(x - \pi) &= u(x + \pi), & x \in (0, 1).
\end{aligned} \tag{14}$$

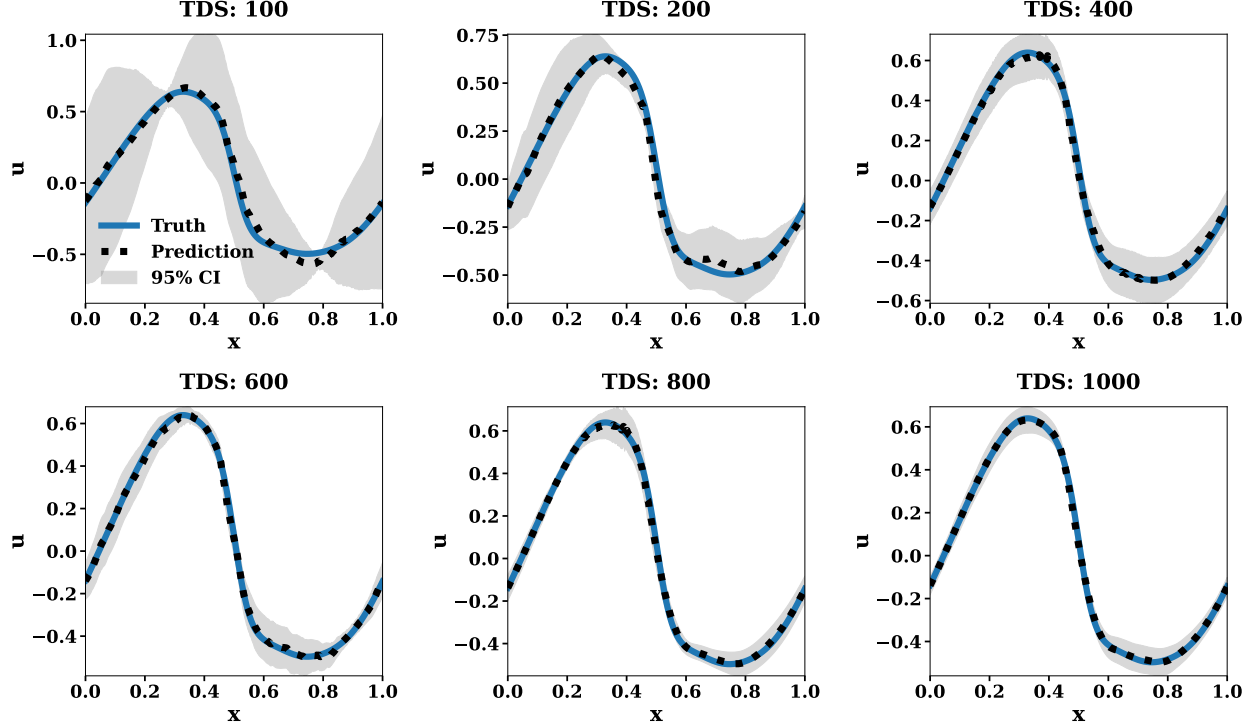


Figure 3: **NOGaP** predictions for 1D Burgers' equation for different sizes of training data samples (TDS). This figure shows the ground truth and the predicted solution along with the 95% confidence interval (CI). It can be observed that the predictive uncertainty decreases with an increase in training sample sizes, thus reducing the CI. It is the expected behavior in GPR.

Here  $\nu$  is positive and greater than zero,  $u$  representing the speed of the flow and the boundary condition is  $2\pi$  periodic. The initial condition is chosen as,

$$u(x, 0) = h1_{\{c-\frac{\omega}{2}, c+\frac{\omega}{2}\}} + \sqrt{\max(h^2 - (a(x-c))^2, 0)}. \quad (15)$$

here the variable  $\omega$  represents the width and  $h$  represents the height of the square wave. The wave is centred around at  $x = c$  and the values of  $\{c, \omega, h\}$  are selected from  $[0.3, 0.7] \times [0.3, 0.6] \times [1, 2]$ . For this example, the value of  $\nu$  is set to 1 and we are learning the mapping from the initial condition to the final condition at  $t = 0.5$  that is  $u(x, t = 0) \mapsto u(x, t = 0.5)$ . The spatial resolution grid is 40 and we have considered 1000 instances of training samples.

**Results:** Figure 4 depicts the predictions obtained from the NOGaP for different initial conditions. The Table 2 shows the relative MSE value from NOGaP and other frameworks. The RMSE value for NOGaP is better compared to WNO, GP (zero mean) and B-WNO. Similar to the previous example it can be observed that the NOGaP model can make predictions effectively for this example. Further, a case study is carried out by considering different training sample sizes (see Figure 5). The mean response for all the cases is found to have an excellent match with the ground truth. The predictive uncertainty, on the other hand, reduces with an increase in the number of training samples. This indicates a reduction in the predictive uncertainty with an increase in training samples.

Table 2: Relative Mean Square Error (RMSE) between the ground truth and the predicted results.

|                       | PDEs                          |                              |                               |                              |
|-----------------------|-------------------------------|------------------------------|-------------------------------|------------------------------|
|                       | Burger                        | Wave Advection               | Darcy (triangular)            | Poisson                      |
| <b>NOGaP</b>          | $\approx 4.097 \pm 0.307 \%$  | $\approx 0.232 \pm 0.01 \%$  | $\approx 5.798 \pm 0.407 \%$  | $\approx 0.278 \pm 0.033 \%$ |
| <b>WNO</b>            | $\approx 6.556 \pm 0.794 \%$  | $\approx 0.599 \pm 0.127 \%$ | $\approx 5.589 \pm 0.715 \%$  | $\approx 3.710 \pm 0.593 \%$ |
| <b>B-WNO</b>          | $\approx 21.434 \pm 0.784 \%$ | Did not converge             | $\approx 11.714 \pm 0.987 \%$ | $\approx 9.042 \pm 0.484 \%$ |
| <b>GP (zero mean)</b> | $\approx 4.581 \pm 0.111 \%$  | $\approx 5.186 \pm 0.509 \%$ | $\approx 11.004 \pm 0.461 \%$ | $\approx 1.503 \pm 0.245 \%$ |

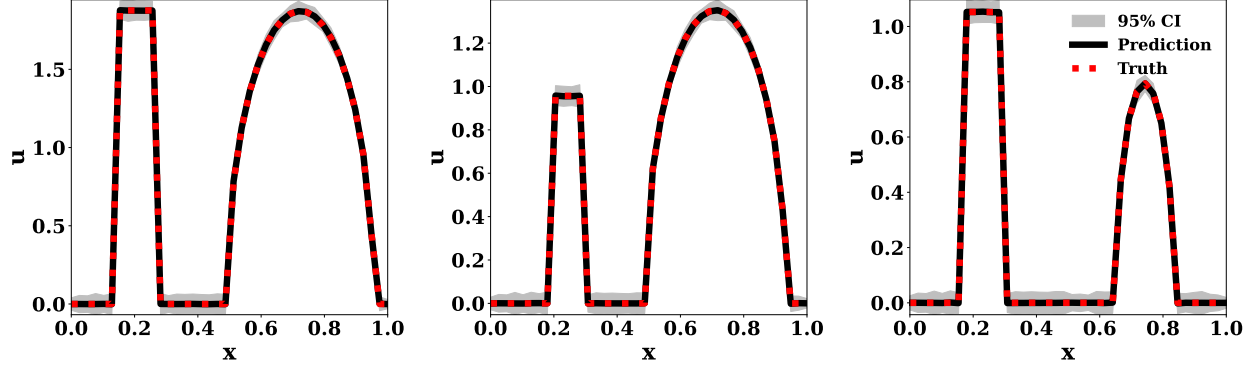


Figure 4: **1D Wave advection equation with periodic boundary condition.** The plot illustrates the predictions of the NOGaP model, which effectively learns the mapping between input and output function spaces. The predicted solutions for three different initial conditions obtained by the proposed NOGaP framework for  $u(x, t)$  at  $t = 0.5$  are plotted, along with a 95 % confidence interval.

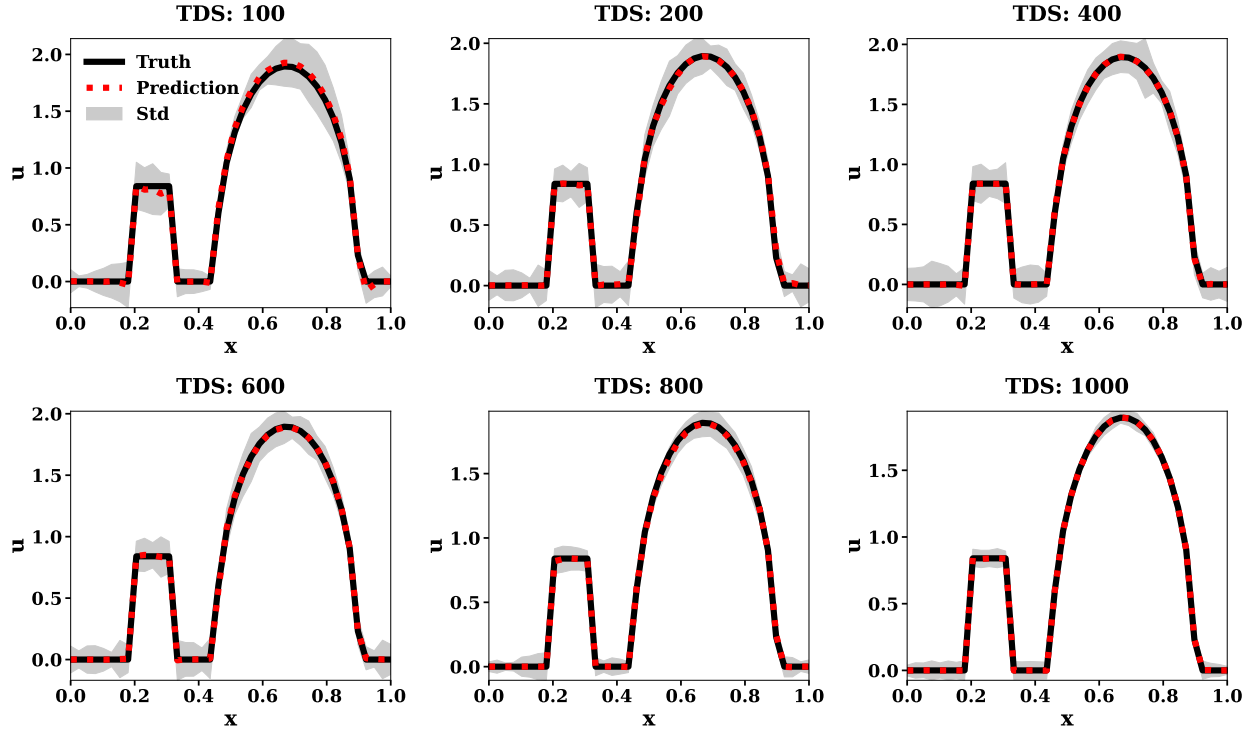


Figure 5: **NOGaP predictions for 1D Wave advection equation for different training sample sizes.** The plot shows the ground truth and predicted solution along with the 95% CI. It can be observed that the predictive uncertainty decreases with an increasing number of training sample sizes.

### 3.3 2D Darcy flow equation with a notch in triangular domain

The 2D Darcy flow equation is a popular tool utilized for modeling the flow of gases or liquids through a porous medium. This equation is particularly useful in fields such as environmental engineering, geology, and hydrology, where understanding fluid dynamics in porous media is essential. Without the time component, the 2D Darcy flow equation can be represented as a second-order nonlinear elliptic partial differential equation (PDE). The equation is given as,

$$\begin{aligned} -\nabla \cdot (a(x, y) \nabla u(x, y)) &= f(x, y), \quad x, y \in (0, \mathbb{R}) \\ u(x, y) &= u_0(x, y), \quad x, y \in \partial(0, \mathbb{R}) \end{aligned} \quad (16)$$

here we have  $u(x, y) = u_0(x, y)$  is the initial condition,  $a(x, y)$  is the permeability field,  $u(x, y)$  is the pressure,  $f(x, y)$  is a source function, and  $\nabla u(x, y)$  is the pressure gradient. For this example, the setup and dataset are taken from Ref. [1]. In this example, a complex boundary condition is considered that is a triangular domain with a notch. The dataset is generated using Gaussian Process (GP) as follows:

$$\begin{aligned} u(x) &\sim \mathcal{GP}(0, k(x, x')), \\ k(x, x') &= \exp\left(-\frac{(x - x')^2}{2l^2}\right), \quad l = 0.2, \quad x, x' \in [0, 1] \end{aligned} \quad (17)$$

A notch is introduced in the flow medium with the triangular geometry. The forcing function  $f(x, y)$  is set to -1 and the permeability field  $a(x, y)$  is set to 0.1. The objective is to learn the mapping from the input (boundary conditions) and output (pressure field) function spaces that are  $u(x, y)|_{\partial\omega} \mapsto u(x, y)$ . The spatial resolution is taken to be  $26 \times 26$ .

**Results:** The Figure 6 shows the prediction obtained from the proposed NOGaP model. It displays predictions corresponding to one of the boundary conditions. The corresponding ground truth is obtained from the numerical solver and the mean predictions from the NOGaP. The plot suggests that the proposed framework can make accurate predictions and is able to approximate the true solution. From the Table 2 we can observe, that there isn't much difference between the RMSE value for both NOGaP and WNO but since the former provides an uncertainty measure over the predictions it becomes more useful and reliable. However, the predictions are significantly improved when compared with GP (zero mean) and B-WNO. Through visual inspection, one can hardly find any discernable difference between the ground truth and the predictions.

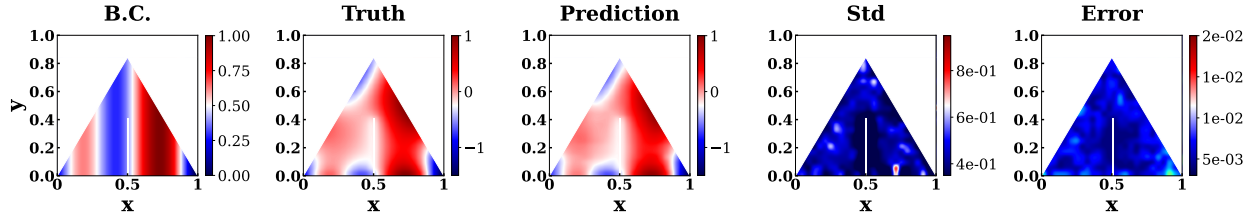


Figure 6: **2D Darcy flow with a notch in the triangular domain.** The figure shows the prediction obtained from the proposed NOGaP framework for one of the representative test samples. The plot shows the boundary condition, the ground truth, the mean and standard deviation of the prediction and the corresponding error.

### 3.4 2D Non-homogeneous Poisson's Equation

The non-homogeneous Poisson equation is an elliptic partial differential equation (PDE) widely applicable in diverse physical scenarios. It serves as the fundamental mathematical model for various problems, such as the steady-state heat diffusion and the electric field generated by a specified electric charge. The Poisson equation, incorporating a source term  $f(x, y)$  and periodic boundary conditions, is mathematically expressed as:

$$\begin{aligned} \partial_{xx}u + \partial_{yy}u &= f(x, y), \quad x \in [-1, 1], \quad y \in [-1, 1] \\ u(x = -1, y) &= u(x = 1, y) = u(x, y = -1) = u(x, y = 1) = 0 \end{aligned} \quad (18)$$

where,  $f(x, y)$  is the source function,  $u(x, y)$  is the solution of the equation over the 2D domain. The objective is to learn the mapping between the input and the output space, that is  $f(x, y) \mapsto u(x, y)$ . To establish the ground truth solution, we opt for an analytical solution represented as  $u(x, y) = \alpha \sin(\pi x)(1 + \cos(\pi y)) + \beta \sin(2\pi x)(1 - \cos(2\pi y))$ . For generating input data, we employ an analytical form for the source function, defined as  $f(x, y) = 16\beta\pi^2(\cos(4\pi y) \sin(4\pi x) + \sin(4\pi x)(\cos(4\pi y) - 1)) - \alpha\pi^2(\cos(\pi y) \sin(\pi x) + \sin(\pi x)(\cos(\pi y) + 1))$ . This expression for  $f(x, y)$  is derived by substituting the analytical expression for  $u(x, y) = \alpha \sin(\pi x)(1 + \cos(\pi y)) + \beta \sin(2\pi x)(1 - \cos(2\pi y))$  into Equation 18. Consequently, the solution implicitly satisfies the boundary conditions. The training instances are generated by varying the parameters  $\alpha$  and  $\beta$  such that  $\alpha \sim \text{Unif}(-2, 2)$  and  $\beta \sim \text{Unif}(-2, 2)$ . The spatial resolution of  $33 \times 33$  is taken for this example.

**Results:** The Figure 7 shows the predictions obtained from the proposed NOGaP model. It displays predictions corresponding to one of the source functions. The plot consists of the ground truth obtained from the numerical solver

and the mean predictions from the NOGaP framework. The plot suggests that the proposed framework can make predictions effectively. From Table 2, we can observe that the Relative Mean Square Error (RMSE) value is significantly improved over WNO. Thus, we can conclude that NOGaP is an improvement over the standalone WNO, GPR (zero mean) and B-WNO. Moreover, visually, the predicted and true solutions are almost indiscernible.

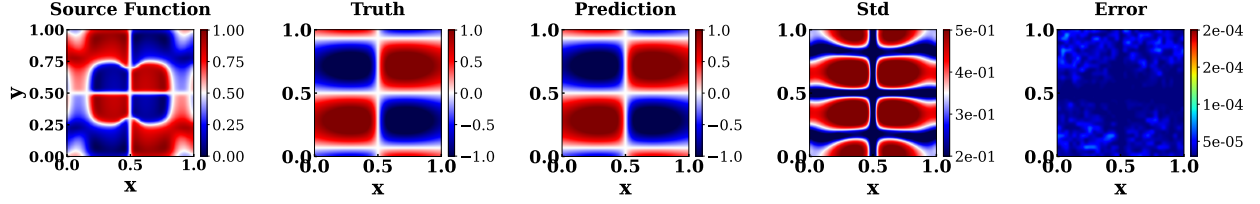


Figure 7: **2D Non-homogeneous Poisson's equation**. The figure shows the prediction obtained from the proposed NOGaP framework for one of the representative test samples. The plot shows the source function, the ground truth, the mean and standard deviation of the prediction and the corresponding error.

## 4 Conclusion

This paper introduced a novel framework called NOGaP for effectively learning solutions to non-linear PDEs. By harnessing the capabilities of GP, the proposed framework not only provides accurate predictions but also provides a measure of uncertainty. By incorporating mean functions to be WNO, the NOGaP framework demonstrates effective performance in learning the solutions of non-linear PDEs. Compared to a standalone WNO, GPR (zero mean) and B-WNO, the proposed framework offers distinct advantages in terms of improved prediction capabilities.

To validate the effectiveness of the framework, four numerical examples were considered, and the predictions were compared against the ground truth obtained using numerical solvers. The results demonstrate that the NOGaP framework achieves improved accuracy, with the predictions closely matching the ground truth solutions. The following observation has been made from the results obtained:

- The predictions obtained from the NOGaP framework demonstrate a strong agreement with the ground truth, providing a good approximation of the true solutions.
- The selection of a suitable mean function poses a significant challenge in constructing effective frameworks involving GP. However, by incorporating the Wavelet Neural Operator (WNO) as the mean function, the NOGaP framework achieves improved accuracy over standalone WNO, GP (zero mean) and B-WNO.
- Additionally, the NOGaP framework is convenient to implement and applicable to a wide variety of examples, making it a versatile tool for solving non-linear partial differential equations.

The results suggest that the proposed NOGaP can make predictions effectively and efficiently. We have used vanilla kernels as the choice of covariance function, but it will be interesting to note the effect of other possible kernels on the proposed NOGaP framework. Working with GPs is computationally expensive, especially with large and high dimensional datasets, thus further study can be done to bring down the computational cost.

## Acknowledgements

SK acknowledges the support received from the Ministry of Education (MoE) in the form of Research Fellowship. RN acknowledge the financial support received from the Science and Engineering Research Board (SERB), India via grant no. SRG/2022/001410. SC acknowledge the financial support received from the Ministry of Port and Shipping via letter no. ST-14011/74/MT (356529). Faculty Seed grants received from IIT Delhi are also acknowledged.

## Code availability

On acceptance, all the source codes to reproduce the results in this study will be made available to the public on GitHub by the corresponding author.

## References

- [1] Zongyi Li, Nikola Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Fourier neural operator for parametric partial differential equations. *arXiv preprint arXiv:2010.08895*, 2020.
- [2] Lu Lu, Pengzhan Jin, and George Em Karniadakis. Deeponet: Learning nonlinear operators for identifying differential equations based on the universal approximation theorem of operators. *arXiv preprint arXiv:1910.03193*, 2019.
- [3] Tapas Tripura and Souvik Chakraborty. Wavelet neural operator for solving parametric partial differential equations in computational mechanics problems. *Computer Methods in Applied Mechanics and Engineering*, 404:115783, 2023.
- [4] Jaideep Pathak, Shashank Subramanian, Peter Harrington, Sanjeev Raja, Ashesh Chattopadhyay, Morteza Mardani, Thorsten Kurth, David Hall, Zongyi Li, Kamyar Azizzadenesheli, Pedram Hassanzadeh, Karthik Kashinath, and Animashree Anandkumar. Fourcastnet: A global data-driven high-resolution weather model using adaptive fourier neural operators, 2022.
- [5] Kenghong Lin, Xutao Li, Yunming Ye, Shanshan Feng, Baoquan Zhang, Guangning Xu, and Ziyi Wang. Spherical neural operator network for global weather prediction. *IEEE Transactions on Circuits and Systems for Video Technology*, 2023.
- [6] Somdatta Goswami, Minglang Yin, Yue Yu, and George Em Karniadakis. A physics-informed variational deeponet for predicting crack path in quasi-brittle materials. *Computer Methods in Applied Mechanics and Engineering*, 391:114587, March 2022.
- [7] Jyoti Rani, Tapas Tripura, Hariprasad Kodamana, Souvik Chakraborty, and Prakash Kumar Tamboli. Fault detection and isolation using probabilistic wavelet neural operator auto-encoder with application to dynamic processes. *Process Safety and Environmental Protection*, 173:215–228, 2023.
- [8] Shailesh Garg and Souvik Chakraborty. Vb-deeponet: A bayesian operator learning framework for uncertainty quantification. *Engineering Applications of Artificial Intelligence*, 118:105685, 2023.
- [9] Emilia Magnani, Nicholas Kramer, Runa Eschenhagen, Lorenzo Rosasco, and Philipp Hennig. Approximate bayesian neural operators: Uncertainty quantification for parametric pdes. *ArXiv*, abs/2208.01565, 2022.
- [10] Yifan Chen, Bamdad Hosseini, Houman Owhadi, and Andrew M Stuart. Solving and learning nonlinear pdes with gaussian processes, 2021.
- [11] Pau Battle, Matthieu Darcy, Bamdad Hosseini, and Houman Owhadi. Kernel methods are competitive for operator learning, 2023.
- [12] Yifan Chen, Houman Owhadi, and Florian Schafer. Sparse cholesky factorization for solving nonlinear pdes via gaussian processes. *ArXiv*, abs/2304.01294, 2023.
- [13] Navaneeth N, Tapas Tripura, and Souvik Chakraborty. Physics informed wno, 2023.
- [14] Lu Lu, Pengzhan Jin, Guofei Pang, Zhongqiang Zhang, and George Em Karniadakis. Learning nonlinear operators via deeponet based on the universal approximation theorem of operators. *Nature Machine Intelligence*, 3(3):218–229, 2021.
- [15] Sifan Wang, Hanwen Wang, and Paris Perdikaris. Learning the solution operator of parametric partial differential equations with physics-informed deeponets. *Science advances*, 7(40):eabi8605, 2021.
- [16] Jan S Hesthaven and Stefano Ubbiali. Non-intrusive reduced order modeling of nonlinear problems using neural networks. *Journal of Computational Physics*, 363:55–78, 2018.
- [17] Kaushik Bhattacharya, Bamdad Hosseini, Nikola B Kovachki, and Andrew M Stuart. Model reduction and neural networks for parametric pdes. *The SMAI journal of computational mathematics*, 7:121–157, 2021.
- [18] Zongyi Li, Nikola Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Neural operator: Graph kernel network for partial differential equations, 2020.
- [19] Zongyi Li, Daniel Zhengyu Huang, Burigede Liu, and Anima Anandkumar. Fourier neural operator with learned deformations for pdes on general geometries, 2022.
- [20] Zongyi Li, Nikola Kovachki, Chris Choy, Boyi Li, Jean Kossaifi, Shourya Otta, Mohammad Amin Nabian, Maximilian Stadler, Christian Hundt, Kamyar Azizzadenesheli, et al. Geometry-informed neural operator for large-scale 3d pdes. *Advances in Neural Information Processing Systems*, 36, 2024.
- [21] Tapas Tripura and Souvik Chakraborty. A foundational neural operator that continuously learns without forgetting, 2023.

- [22] Akshay Thakur, Tapas Tripura, and Souvik Chakraborty. Multi-fidelity wavelet neural operator with application to uncertainty quantification, 2023.
- [23] Jyoti Rani, Tapas Tripura, Hariprasad Kodamana, and Souvik Chakraborty. Generative adversarial wavelet neural operator: Application to fault detection and isolation of multivariate time series data, 2024.
- [24] N Navaneeth and Souvik Chakraborty. Waveformer for modelling dynamical systems, 2023.
- [25] Markus Lange-Hegermann. Algorithmic linearly constrained gaussian processes, 2019.
- [26] Laura P Swiler, Mamikon Gulian, Ari L Frankel, Cosmin Safta, and John D Jakeman. A survey of constrained gaussian process regression: Approaches and implementation challenges. *Journal of Machine Learning for Modeling and Computing*, 1(2), 2020.
- [27] Mamikon Gulian, Ari Frankel, and Laura Swiler. Gaussian process regression constrained by boundary value problems. *Computer Methods in Applied Mechanics and Engineering*, 388:114117, 2022.
- [28] Liu Yang, Xuhui Meng, and George Em Karniadakis. B-pinns: Bayesian physics-informed neural networks for forward and inverse pde problems with noisy data. *Journal of Computational Physics*, 425:109913, January 2021.
- [29] Shailesh Garg and Souvik Chakraborty. Variational bayes deep operator network: A data-driven bayesian solver for parametric differential equations, 2022.
- [30] Zongren Zou, Xuhui Meng, Apostolos F Psaros, and George Em Karniadakis. Neuraluq: A comprehensive library for uncertainty quantification in neural differential equations and operators, 2022.
- [31] Marvin Pförtner, Ingo Steinwart, Philipp Hennig, and Jonathan Wenger. Physics-informed gaussian process regression generalizes linear pde solvers, 2023.
- [32] Yifan Chen, Houman Owhadi, and Florian Schäfer. Sparse cholesky factorization for solving nonlinear pdes via gaussian processes, 2023.
- [33] Andreas Besginow and Markus Lange-Hegermann. Constraining gaussian processes to systems of linear ordinary differential equations, 2022.
- [34] Ilias Bilonis, Nicholas Zabaras, Bledar A Konomi, and Guang Lin. Multi-output separable gaussian process: Towards an efficient, fully bayesian paradigm for uncertainty quantification. *Journal of Computational Physics*, 241:212–239, 2013.
- [35] Edwin V Bonilla, Kian Chai, and Christopher Williams. Multi-task gaussian process prediction. *Advances in neural information processing systems*, 20, 2007.
- [36] Mauricio A. Alvarez, Lorenzo Rosasco, and Neil D. Lawrence. Kernels for vector-valued functions: a review, 2012.
- [37] Kevin P Murphy. *Probabilistic machine learning: an introduction*. MIT press, 2022.
- [38] Carl Edward Rasmussen. *Gaussian Processes in Machine Learning*, pages 63–71. Springer Berlin Heidelberg, Berlin, Heidelberg, 2004.
- [39] C Bishop. Pattern recognition and machine learning. *Springer google schola*, 2:531–537, 2006.
- [40] Dan Hendrycks and Kevin Gimpel. Gaussian error linear units (gelus), 2023.