用于不确定性量化的随机先验小波神经算子

预印本

Shailesh Garg
应用力学系 印度理工学院德里分校 豪兹卡斯，新德里 110016，印度。shaileshgarg96@gmail.com

Souvik Chakraborty
应用力学系 亚迪人工智能学院（YScAI）印度理工学院德里分校 豪兹卡斯，新德里 110016，印度。souvik@am.iitd.ac.in

arXiv:2302.01051v1 [stat.ML] 2 Feb 2023

2023年2月3日

摘要

本文提出一种新颖的数据驱动算子学习框架，称为随机先验小波神经算子（RP-WNO）。所提出的 RP-WNO 是近期提出的小波神经算子的扩展。后者具有出色的泛化能力，但无法估计其预测所伴随的不确定性。与原始小波神经算子不同，RP-WNO 内置了不确定性量化模块，因此预计对科学家和工程师都极为有用。RP-WNO 使用随机先验网络，这类网络能够纳入先验信息，而且与贝叶斯对应方法相比，在大型、复杂的深度学习架构中更易于实现。本文通过求解四个示例来检验所提出的框架，所得结果充分证明了该框架的有效性。

关键词 不　确定性量化 · 小波神经算子 · 随机先验网络 · 神经网络　s

1 引言

不确定性是任何源自物理过程的数据集不可分割的一部分。所考虑的物理过程可以是自然现象，例如天气现象 [1, 2], ，也可以是人为过程，例如建筑物挠度 [3–5]。这些不确定性 [6, 7] 可能表现为（i）偶然不确定性、（ii）认知不确定性，或两者兼有。偶然不确定性主要源于物理过程固有的随机性。使用物联网收集的数据集中的测量噪声也可能是偶然不确定性的来源。由于偶然不确定性的来源无法消除，因此它通常被称为不可约不确定性。另一种可能存在的不确定性是认知不确定性，它源于对所考虑物理过程的认知不完整。理论上，只要收集足够的数据，人们便能加深对物理过程的理解，因此认知不确定性也称为可约不确定性。尽管理论上只要有足够的数据，认知不确定性就可以降低，但实际上，由于存储容量有限、数据收集时间有限等约束，它永远无法被彻底消除。

如今，在研究任何物理过程时，了解其以各种形式表现出的相关不确定性，可以为用户提供额外的信心，尤其是在决策任务中。了解偶然不确定性可以帮助用户判断系统偏离均值状态的程度，而了解认知不确定性则可以让用户评估所选数据集是否足以支持决策。现在，让我们暂时聚焦于为任何物理过程生成数据集这一问题。由于物理约束（如时间和现场存储容量）以及科学约束（如传感器技术或传感器布置问题），现实中可获取的各种物理过程数据十分有限。因此，用户需要通过人工数据对这些数据进行扩充，以增加数据总量。为生成这些人工数据，我们需要开发合适的计算高效算法，使其能够学习物理系统观测到的输入与输出之间的映射。此外，在可靠性分析 [8–10]或设计

优化 [11–13], 中，当需要进行重复模拟时，这些算法可以作为代理模型，以降低与详细分析相关的计算成本。

在现有文献中，过去人们曾使用多项式混沌展开 [14–16], 响应面法 [17–21], 和高斯过程 [22–25] 等技术来构建代理模型；但近年来，研究人员开始深入探索基于深度学习的技术。深度学习 [26, 27] 算法凭借其强大的可修改性、低预测成本和灵活的实现方式，已在众多领域 [28–36] 得到广泛应用。如今，基于深度学习的数据驱动算子学习算法能够学习物理过程输入与输出之间的映射。近期提出的小波神经算子 [37, 38] (WNO)、DeepONet [9, 39, 40], 和傅里叶神经算子 [41] 都是颇受欢迎的算子学习算法；尽管它们是数据驱动算法，却展现出卓越的泛化能力。物理信息神经网络 [42], 等函数逼近算法直接利用系统的控制方程，也可用于此目的。然而，这类方法的应用受到限制，因为要成功收敛到解，必须全面了解物理规律。数据驱动算法则不受此类限制。

理论上，只要数据集足够大，估计偶然不确定性就很简单，而使用合适的代理模型即可满足这一条件。然而，估计认知不确定性并非易事，需要专门为此设计算法。在深度学习中，可以通过贝叶斯方法或频率学派方法实现不确定性量化。前者利用贝叶斯推断 [43]进行不确定性量化。在贝叶斯方法中，基于蒙特卡洛的方法 [44,45] 最为准确，但面对复杂的深度神经网络架构时，这类方法并不实用。另一种流行的贝叶斯方法是采用基于变分推断 [46,47] 的方法。虽然这类方法可以应用于复杂的神经网络架构，但只能得到近似解，而且对于多模态后验分布，甚至可能无法收敛。另一方面，频率学派方法通常更容易实现，因为它们基于确定性框架。在频率学派方法中，集成训练 [48] 是最流行的不确定性估计方法之一。然而，在普通的集成训练中，没有纳入先验信息的机制。

本文提出了一种新颖的随机先验小波神经算子（RP-WNO）。所提出的 RP-WNO 利用随机先验网络的概念 [49]，能够估计其预测所关联的认知不确定性（预测不确定性），从而帮助用户做出有充分依据的决策。随机先验网络能够融入先验信息，同时又处于频率学派框架中，因此是该框架的理想选择。本文通过四个示例展示了所提算法的有效性：既能保持 WNO 的泛化能力，又能实现此前在 WNO 架构中无法实现的不确定性量化。

本文其余部分安排如下。第2节讨论 WNO 并介绍所提出的 RP-WNO。第3节通过多个示例展示所提框架的有效性，第4节总结全文。

2 所提出的框架

算子是一种数学实体，可将 $\mathcal{I}$ 维输入函数空间映射到 $\mathcal{O}$ 维输出函数空间。该关系可描述为：

$$f_{\mathcal{I}} \in \mathbb{R}^{\mathcal{I}} \xrightarrow{\mathbb{O}} f_{\mathcal{O}} \in \mathbb{R}^{\mathcal{O}}, \tag{1}$$

其中，$\mathbb{O}$ 表示将输入函数映射到输出函数的算子。在基于深度学习的算子学习方案中，目标是通过学习两个函数空间之间的映射来学习算子。在小波神经算子中，结合使用小波变换和卷积神经网络来学习映射，进而学习算子 $\mathbb{O}$。

2.1 小波神经算子

小波神经算子（WNO）在 [37], 中提出，是一种基于深度学习的算子学习方案，通过在其架构中使用小波块来实现目标。任何基于深度学习的方案，其效果都深受训练数据质量的影响。与包含大量噪声和冗余信息的数据相比，信息精炼且密集的数据更容易学习。这是因为学习密集数据集所需的权重和偏置更少，从而使反向传播更简单、更精确。为此，小波神经算子采用小波变换，从输入数据中提取相关信息，同时降低维度，因此更容易学习输入与输出之间的映射。

小波神经算子将 n 输入函数及其离散化所在的网格作为输入，并将它们提升到更高的维度，使用个全连接层。对于一个二维区域，使用具有 r 行的网格进行离散化，c

列，以及 s 个样本，上述数据集维度演化过程可视化如下，

$$\mathbf{I} \in \mathbb{R}^{s\times r\times c\times n+2} \xrightarrow{\text{Dense layer}} \mathbf{O}_D \in \mathbb{R}^{s\times r\times c\times p}, \tag{2}$$

其中 $\mathbf{I}$ 是小波神经算子的输入，$\mathbf{O}_D$ 是具有 p 个节点的全连接层的输出。请注意，$p > n+2$ 而且在缺少输入函数的网格信息时，输入维度将从 $(s\times r\times c\times n+2)$ 变为 $(s\times r\times c\times n)$。全连接层的输出会传递给小波块，并且经过小波块的最终运算后，其维度保持不变。

2.1.1 小波块

在 WNO 架构中，输入经过升维后，会被传递至小波块（WB）。小波块的结构设计使其能够从数据中提取相关信息，同时降低神经网络的参数需求。这是通过使用小波变换和卷积神经网络实现的。小波块有两个主要组件：（i）小波映射组件（WMC）和（ii）卷积映射组件（CMC）。这两个组件接收相同的输入，经过处理后，返回与输入维度相同的输出。

WMC 中使用的小波变换（WT）将数据分解为高频和低频小波系数，并对数据进行下采样，从而提取相关信息。WMC 首先对输入执行多层小波变换。通过这一过程，我们可以（i）在降低总体维度的同时，仅提取相关信息；（ii）减少从数据中学习特征所需的参数数量。在维度得到充分降低后，可使用简单的稠密线性层学习小波变换最后一层得到的小波系数。随后，使用线性层的输出执行逆小波变换。需要注意的是，在小波变换的最后一层之后，可以使用稠密线性层学习全部小波系数，也可以只学习选定的小波系数。未被学习的系数在执行逆小波变换时保持原样。

WMC 中执行的小波变换可根据所学习数据集的要求采用连续形式或离散形式。连续小波变换比离散小波变换更精确，并且能够产生更多可供学习的小波系数，从而获得更好的性能；然而，与离散小波变换相比，其计算成本更高。对于前述 WMC 内二维数据集的示例，考虑离散小波变换（DWT）时，数据集大小的变化可表示如下：

$$\mathbf{I}_{WB} \in \mathbb{R}^{s\times r\times c\times p} \xrightarrow{\text{WT}} \mathbf{O}_{WT} \in \mathbb{R}^{s\times r'\times c'\times p\times 4} \xrightarrow{\text{Dense linear layer}}$$
$$\mathbf{O}_{DLL} \in \mathbb{R}^{s\times r'\times c'\times p\times 4} \xrightarrow{\text{Inverse WT}} \mathbf{O}_{WMC} \in \mathbb{R}^{s\times r\times c\times p}, \tag{3}$$

其中，I_{WB}是小波块的输入，$I_{WB} = \mathbf{O}_D$是 WNO 架构中第一个小波块的输入。$\mathbf{O}_{WT}$是执行多次小波变换后的输出。$r' < r$ 和 $s' < s$ 是经过多级离散小波变换后的数据降维结果，并且由于离散小波变换会为二维区域生成四个小波系数，因此数据集大小也相应体现了这一点。$\mathbf{O}_{DLL}$ 是稠密线性层的输出，该层保持其输入大小不变；$\mathbf{O}_{WMC}$ 是经过逆小波变换后得到的最终输出。需要注意的是，$\mathbf{O}_{WMC}$ 的维度与$\mathbf{I}_{WMC}$的维度相同。

接下来，在 CMC 内部，卷积层的使用方式可确保最终输出的大小保持不变。对于二维空间，CMC 内部数据集大小的变化可追踪如下，

$$\mathbf{I}_{WB} \in \mathbb{R}^{s\times r\times c\times p} \xrightarrow{\text{Convolution layers}} \mathbf{O}_{CMC} \in \mathbb{R}^{s\times r\times c\times p}, \tag{4}$$

其中 $\mathbf{O}_{CMC}$ 是 CMC 的最终输出。将小波块的两个组成部分的输出相加，以获得最终输出 $\mathbf{O}_{WB} = \mathbf{O}_{WMC} + \mathbf{O}_{CMC}$。对于 WNO 架构中的第二个及后续小波块，$\mathbf{I}_{WB}$ 等于前一个小波块的 $\mathbf{O}_{WB}$。图 1 展示了小波块内部信息流的示意图。从最终小波块获得输出 $\mathbf{O}_{WB}$后，使用密集连接层降低提升后的数据的维度，使其与最终输出的大小相匹配。对于二维空间，如果映射针对单个输出函数进行，则最终小波块之后数据集大小的变化可追踪如下，

$$\mathbf{O}_{FWB} \in \mathbb{R}^{s\times r\times c\times p} \xrightarrow{\text{Forward neural net}} \mathbf{O}_{WNO} \in \mathbb{R}^{s\times r\times c\times 1}, \tag{5}$$

其中，$\mathbf{O}_{FWB}$是最终小波块的输出，$\mathbf{O}_{WNO}$是 WNO 的最终输出。图 2 展示了 WNO 架构的示意图。有关 WNO 架构的更多详情，请参阅 [37]。

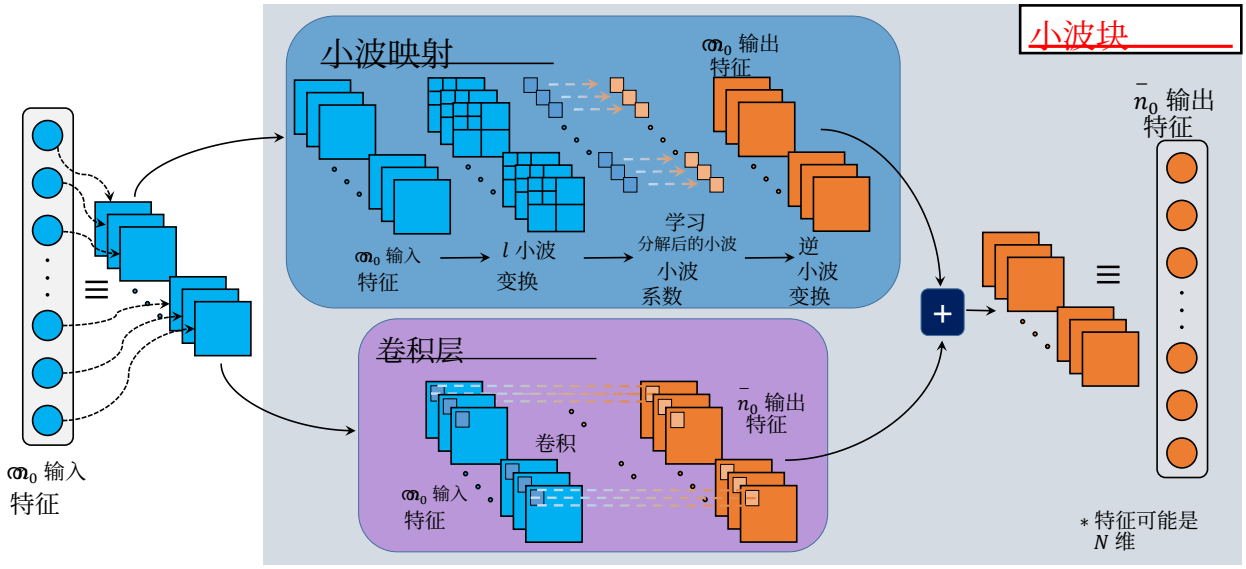


图 1：小波块内部信息流示意图。

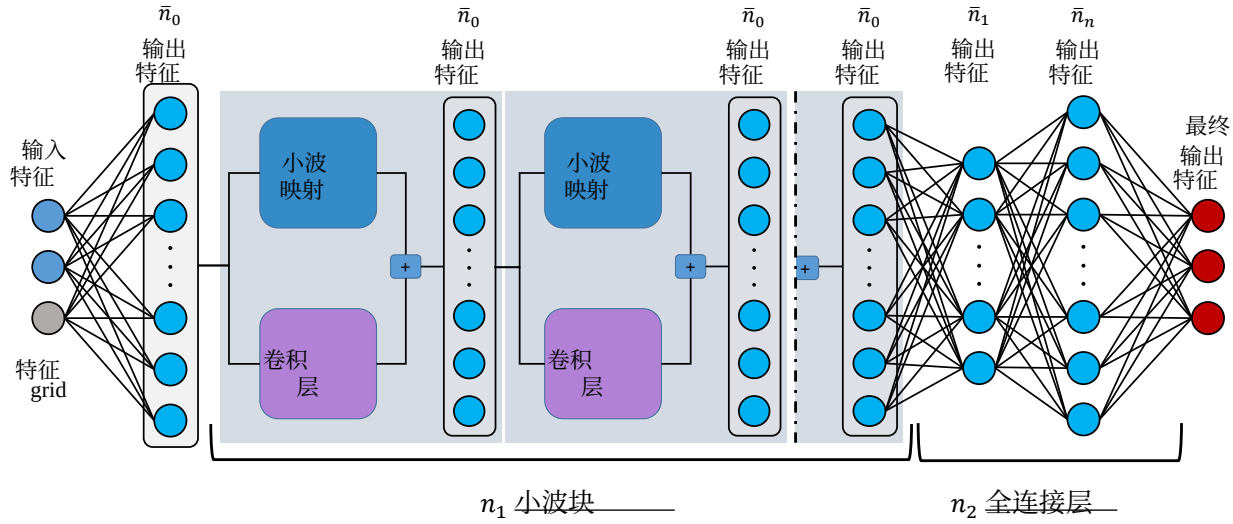


图 2：小波神经算子示意图。

2.2 随机先验小波神经算子（RP-WNO）

所提出的 RP-WNO 采用了 [49]中介绍的随机先验（RP）网络，用于不确定性量化。RP网络采用两个架构相同但随机初始化不同的神经网络，以实现深度学习模型中的不确定性量化。在这两个网络中，一个具有可训练权重和偏置（可训练网络），另一个具有固定权重和偏置（不可训练网络）。与集成训练类似，会训练多个随机先验网络副本，以进行不确定性量化。其优势在于，具有固定权重和偏置的网络使得引入先验信息成为可能，而这在传统的集成训练中是无法实现的。现在，在训练过程中，在每次前向传播中，先将可训练网络和不可训练网络的输出相加，然后再对可训练网络进行反向传播。RP网络的输出$\mathbf{O}_{RP}$可描述为：

$$\mathbf{O}_{RP} = \mathbf{O}_T + \beta \mathbf{O}_{NT}, \tag{6}$$

其中， O_T 是可训练网络的输出，$\mathbf{O}_{NT}$ 是不可训练网络的输出。需要注意的是，RP 网络的损失函数基于式 (6) 得到的输出计算；反向

传播时，仅更新可训练网络的权重和偏置。式 (6) 中的 β 是一个超参数，其值由用户定义。现有文献表明，$\beta = 1$ 可产生最佳结果。

在所提出的 RP-WNO 中，按照 RP 网络的理论，采用两个完全相同的小波神经算子网络，其中一个可训练，另一个不可训练。将二者的输出相加，RP-WNO 的输出 $\mathbf{O}_{RP-WNO}$描述如下，

$$\mathbf{O}_{RP-WNO} = \mathbf{O}_{WNOT} + \mathbf{O}_{WNONT}, \tag{7}$$

其中，$\mathbf{O}_{WNOT}$是可训练小波神经算子模型的输出，$\mathbf{O}_{WNONT}$是不可训练小波神经算子模型的输出。遵循现有文献以及 A 中开展的 β研究，选取 A β 值为 1。图 3 展示了所提出 RP-WNO 的示意图。为实现不确定性量化，我们需要训练多个副本的

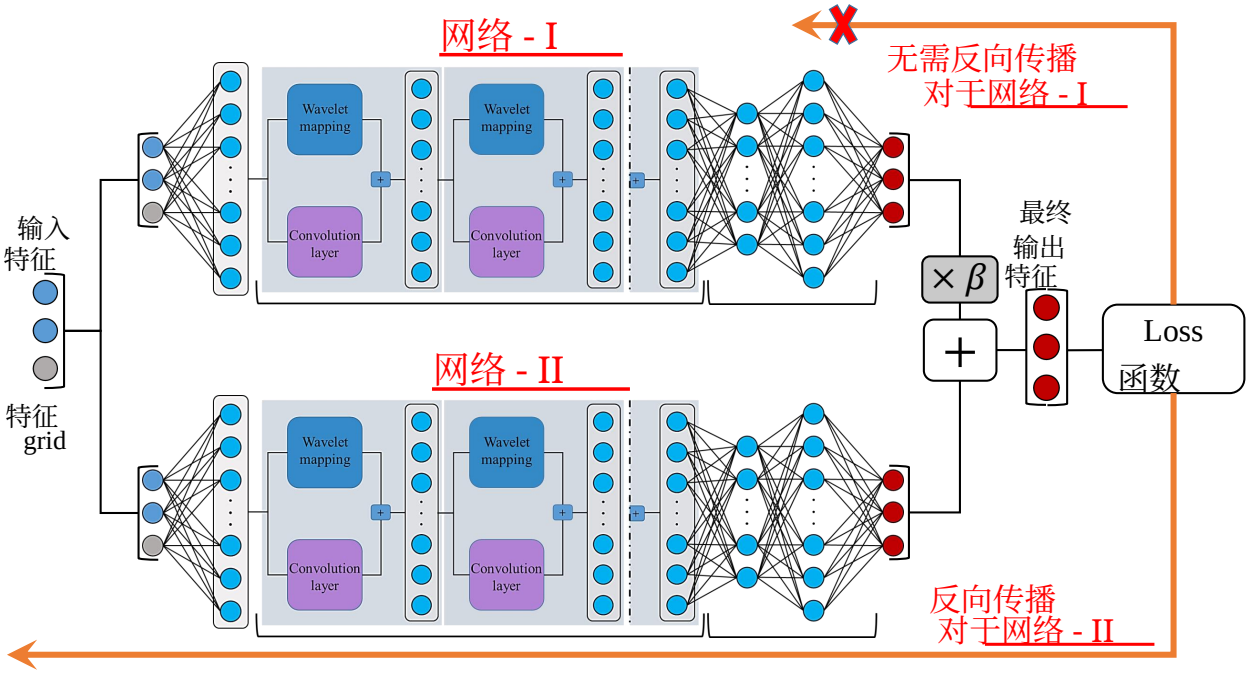


图 3：所提出 RP-WNO 的示意图。

RP-WNO，每个副本采用不同的初始化，但具有相同的架构和训练数据。由于训练任意两个 RP-WNO 副本互不影响，因此可以并行训练，从而节省总计算时间。假设 $\mathcal{P}_i$ 是第 i^{th} 个 RP-WNO 副本的预测值，则平均预测值 $\mathcal{P}_M$ 可表示为，

$$\mathcal{P}_m = \frac{1}{n_c}\sum_{i=1}^{n_c}\mathcal{P}_i, \tag{8}$$

其中，n_c 表示针对相同输入数据训练的 RP-WNO 副本数量。预测的标准差 $\mathcal{P}_s$ 可按如下方式计算，

$$\mathcal{P}_s = \sqrt{\frac{\sum_{i=1}^{n_c}(\mathcal{P}_i - \mathcal{P}_m)^2}{n_c}}. \tag{9}$$

请注意，如果 $\mathcal{P}_i$ 是向量或多维数组，则方程（8）和（9）中规定的运算将按元素逐一执行。现在可以将 95% 置信区间计算为 $\mathcal{P}_m \pm 1.96\mathcal{P}_s$。算法1详细说明了训练所提出的 RP-WNO 集成所涉及的步骤。需要注意的是，与训练数据对应的不可训练网络的输出可在初始化步骤后计算并存储，以便在后续训练中使用，从而节省计算时间。此外，RP-WNO 集成中的多个 RP-WNO模型可以同时并行训练，因为一个模型的训练不会影响其他模型。

算法1RP-WNO的训练算法。

对于i = 1： 集成中RP-WNO网络的数量 **执行**
    使用以下架构初始化可训练WNO和不可训练WNO的权重与偏置
    相同的架构 .
    对于j = 1： 训练轮次 **执行**
        遍历k = 1： 批次数量 **do**
            将训练数据输入可训练网络，并获得 $\mathcal{O}_{WNOT}$。
            将训练数据输入不可训练网络，并获得 $\mathcal{O}_{WNONT}$。
            计算组合输出 $\mathcal{O}_{RP-WNO} = \mathcal{O}_{WNOT} + \mathcal{O}_{WNONT}$。
            使用 $\mathcal{O}_{RP-WNO}$ 计算相关损失函数。
            执行反向传播，并更新可训练网络的权重和偏置。
        结束循环
    结束循环
    保存训练好的模型，以便稍后进行预测。
结束循环
输出： 训练好的 RP-WNO 网络集成。

3 数值示例

本节通过四个案例研究来展示所提出的 RP-WNO 的有效性。第一个案例研究（CS-I）研究一维、含时的伯格斯方程；第二个案例研究（CS-II）研究二维达西流方程。第三个案例研究（CS-III）旨在求解速度模型可变的地震波方程；第四个案例研究（CS-IV）旨在建立晶体固体中原子结构缺陷与介观尺度性质之间的联系。前两个案例研究的数据集使用论文 [37,41] 中提供的代码生成。第三个案例研究的速度场取自 [50]。第四个案例研究的数据集取自 [51]。第一和第二个案例研究中的 RP-WNO 集成采用离散小波变换，而第三和第四个案例研究中的 RP-WNO 集成则在 RP-WNO架构中采用连续小波变换。各案例研究的更多详情见相应的小节。

3.1 CS-I：伯格斯方程

所研究的一维伯格斯方程具有周期边界条件，定义如下：

$$\begin{gathered}\frac{\partial u(x,t)}{\partial t} + u\frac{\partial u(x,t)}{\partial x} = \nu\frac{\partial^2 u(x,t)}{\partial x^2}, \\ x \in [0,1], \quad t \in (0,1] \\ u(x,0) \sim \mathcal{N}(0, 625(-\Delta + 25\,\mathbb{I})^{-2})\end{gathered} \tag{10}$$

其中， ν 是所建模流动的黏度。在当前案例研究中， ν 取值为 0.01。伯格斯方程的初始条件从高斯随机场中采样，如式（10）所示，其中 Δ 是拉普拉斯算子。利用所提出的 RP-WNO，学习初始条件 $u(x,0)$ 与在 $t = 1$, $u(x,1)$ 处获得的速度场之间的映射。图4展示了使用 RP-WNO 得到的平均预测值和置信区间

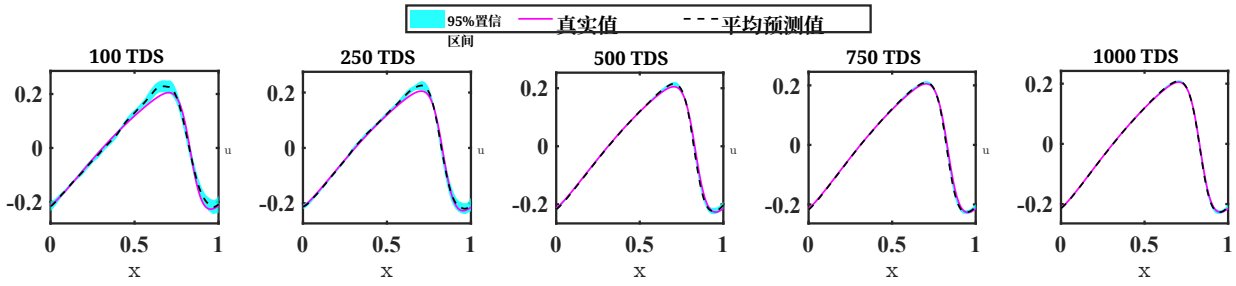


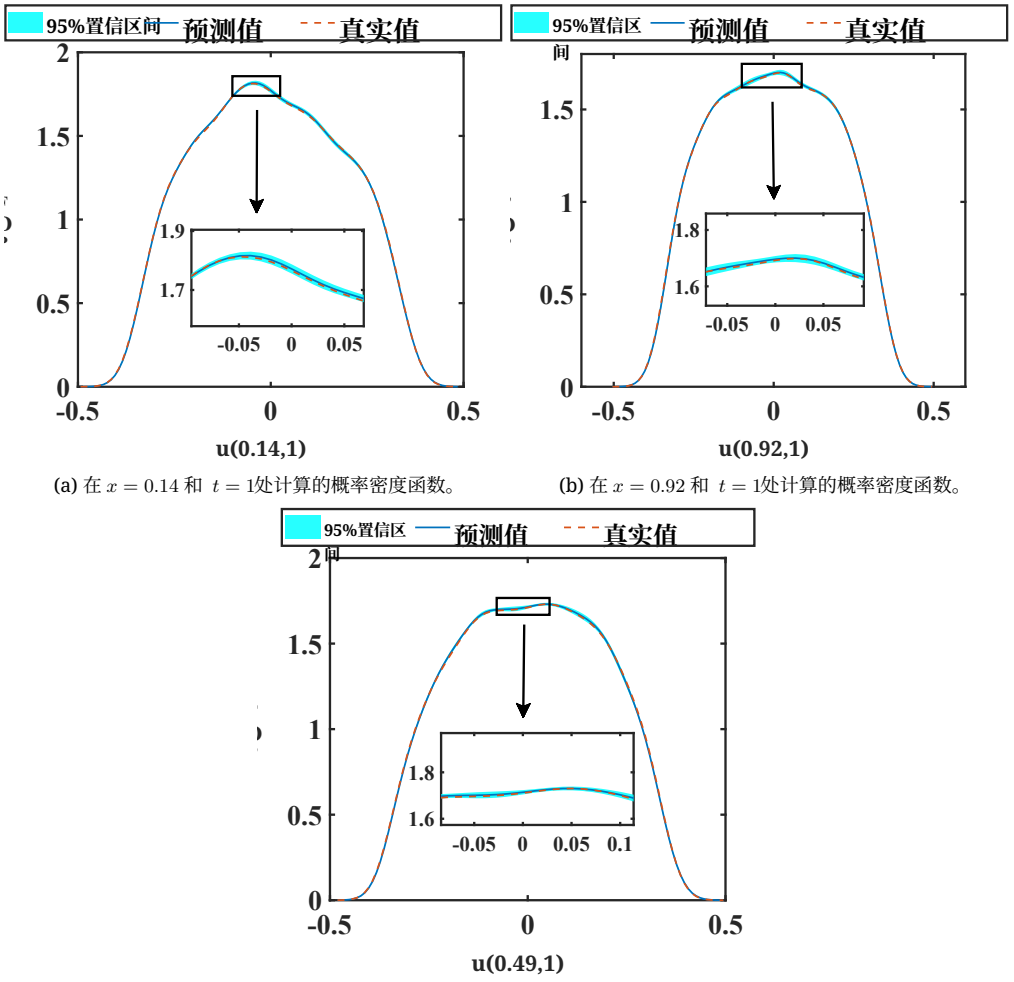
图 4： CS-I 中使用不同数量的训练数据样本（TDS）进行训练时 RP-WNO 的预测。

对应不同数量的训练数据样本（TDS）。可以观察到，随着可用于训练的数据增加，预测的不确定性降低，这符合预期。此外，

表1：CS-I 中 RP-WNO 预测的平均绝对误差和平均标准差。

训练样本	100	250	500	750	1000
MAE	0.0192	0.0098	0.0046	0.0033	0.0025
平均标准差	0.0107	0.0072	0.0039	0.0031	0.0027

随着 TDS 增加，平均预测值逐渐趋近于真实值。表1量化了图4所示的结果。为计算平均绝对误差（MAE）和平均标准差（Std.），使用训练好的 RP-WNO集成生成了包含 8000 个样本的预测集成。计算 MAE 时，将 RP-WNO 的平均预测值与真实值进行比较；计算平均标准差时，则考虑 RP-WNO 集成内部的变化。表1中观察到的趋势进一步印证了图4所示的结果。



(a) 在 $x = 0.14$ 和 $t = 1$处计算的概率密度函数。

(b) 在 $x = 0.92$ 和 $t = 1$处计算的概率密度函数。

（c）在 $x = 0.49$ 和 $t = 1$处计算的概率密度函数。

图5：CS-I 的偶然不确定性估计。

图5展示了 RP-WNO模型在使用1000个训练样本进行训练时，偶然不确定性估计与真实值的对比。RP-WNO 的预测紧密贴合真实值，并提供了能够有效涵盖真实值的额外置信区间。

3.2 CS-II：达西流方程

与时间无关的二维达西流方程可表示为：

$$\begin{aligned} -\nabla(a(x,y)\nabla u(x,y)) &= 1, \\ x \in (0,1), \quad y &\in (0,1), \\ a(x,y) &\sim \psi\mathcal{N}(0,(-\Delta+9\mathbb{I})^{-2}), \end{aligned} \tag{11}$$

其中，$a(x,\ y)$ 为渗透率场，采样自高斯随机场 $\mathcal{N}(0,\ (-\Delta+9\mathbb{I})^{-2})$。方程（11）中的 Δ 为拉普拉斯算子，满足零诺伊曼边界条件；ψ 是一种映射，它将函数变换为正值设为 12、负值设为 3。该案例研究中的映射是在渗透率场 $a(x,\ y)$ 与相应的压力 $u(x,\ y)$ 之间进行的。与前一个案例研究类似，图6展示了

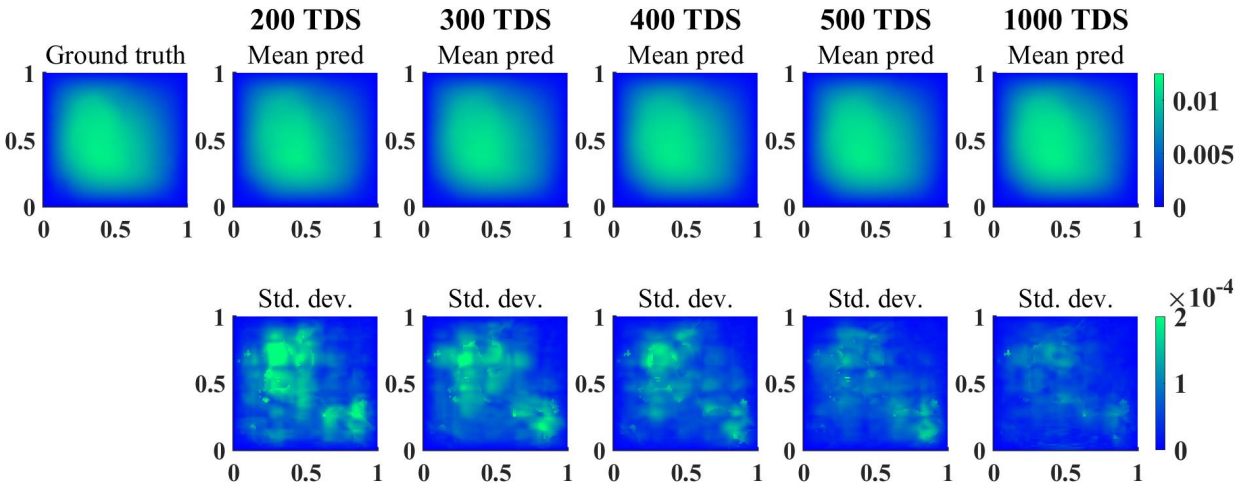


图6：CS-II中针对不同训练数据样本（TDS）数量训练时的RP-WNO预测结果。

当RP-WNO集成使用不同数量的训练样本进行训练时，二维达西流案例研究的预测结果。结果呈现出预期的趋势：随着训练数据增加，预测结果中的变化减小。随着训练

表2：CS-II 中 RP-WNO 预测的平均绝对误差和平均标准差。

训练样本	200	300	400	500	1000
平均绝对误差（$\times 10^{-3}$）	0.1280	0.0962	0.0770	0.0631	0.0408
平均标准差（$\times 10^{-4}$）	0.7224	0.6025	0.5323	0.4813	0.3601

数据如表2所示。采用了八千个测试样本来得出所示结果。平均绝对误差和平均标准差均单调递减，这符合预期。

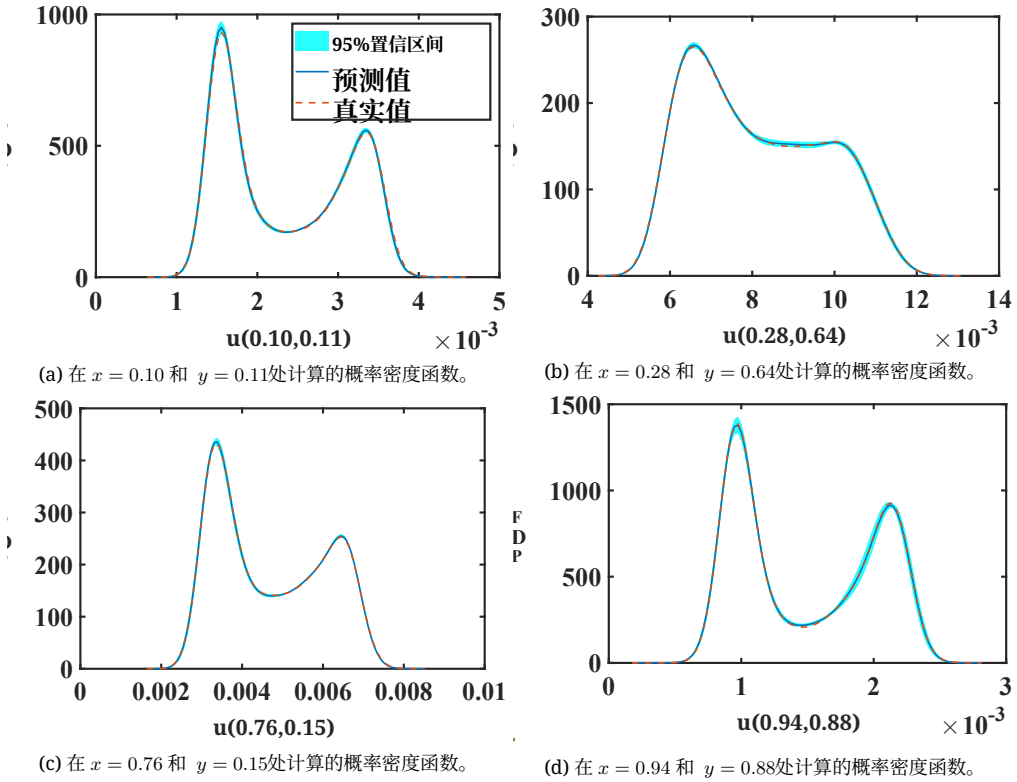
图7展示了 CS-II 中二维 Darcy 流动方程的偶然不确定性估计。如图所示，平均预测值与真实值高度吻合，同时我们还获得了与所提出的 RP-WNO 预测相关的额外置信区间。这清楚地表明，所提出的 RP-WNO 能够捕捉此类多模态分布。RP-WNO 集成使用 1000 个训练样本进行训练，以获得这些结果。

3.3 CS-III：亥姆霍兹方程

在第三个案例研究中，我们将讨论亥姆霍兹方程，其解给出了频域中的压力波场。该方程适用于需要研究行波效应的场景，例如电磁辐射研究、地震勘探等。二维行波的亥姆霍兹方程定义如下：

$$\nabla^2 u(x,z;w) + k^2 u(x,z;w) = f(x_s,z_s;w), \tag{12}$$

其中，u 是 频域中的压力波场，w 是角频率。k 是波数
，点源 ce 位于位置 $(x_s,\ z_s)$。在当前案例研究中，为了生成数据集，Helmholt z



(a) 在 $x = 0.10$ 和 $y = 0.11$处计算的概率密度函数。

(b) 在 $x = 0.28$ 和 $y = 0.64$处计算的概率密度函数。

(c) 在 $x = 0.76$ 和 $y = 0.15$处计算的概率密度函数。

(d) 在 $x = 0.94$ 和 $y = 0.88$处计算的概率密度函数。

图 7：CS-II 中二维达西流的偶然不确定性估计。

方程并非直接求解，而是先在时域中求解二维声波方程，然后将所得结果转换到频域，以获得压力波场$u(x, z)$。二维声波方程定义如下：

$$\nabla^2 p(x,z,t) - \frac{1}{v(x,z)^2}\frac{\partial^2 p(x,z,t)}{\partial t^2} = s(x_s, z_s, t), \tag{13}$$

其中， $p(x, z, t)$ 是时域中的压力波场， $v(x, z)$ 是位置 (x, z) 处的波速。点源 s 位于位置 (x_s, z_s)。式（13）中的空间域 $(x, z) \in [0\,m, 690\,m]^2$ 使用 $\Delta x = \Delta z = 10\,m$ 进行离散化，并选取幅值为 1 的恒定震源进行建模。通过改变速度模型引入不同样本间的变异性，这些速度模型取自 [50]（FlatVel-A 数据集，Vel Family）。震源位置固定为 $x_s = 680\,m$ 和 $z_s = 10\,m$。现在，可以计算宽频率范围内的频域压力波场，但初始的低频分量将贡献大部分能量。因此，为了检验所提出的 RP-WNO 的有效性，所创建的数据集包含对应于 1赫兹、3赫兹、5赫兹、7赫兹和 9赫兹频率的压力波场。

现在，尽管我们可以使用不同的 RP-WNO 集成，将速度场映射为不同频率的压力波场，但为展示所提出的算子学习框架的有效性，本案例研究中针对全部五个频率的映射均使用单个 RP-WNO集成。为此，将 RP-WNO 的输入（原本应为速度场及其离散化所在的网格）与一个额外矩阵拼接，该矩阵包含用于获得输出的相应频率信息。拼接的额外矩阵与速度波场具有相同的维度，其元素均为常数，且等于当前考察的频率。例如，若速度场 $v(x,\ z) = \mathbf{V}_1$, 对应的压力波场为 $u(x,\ z; w) = \mathbf{U}_1^{(w)}$，则我们将得到五个样本，用于训练所提出的 RP-WNO，这些样本可以表示为

如下，

$$\begin{array}{ll} \textbf{Inputs} & \textbf{Outputs} \\ \mathbf{V}_1,\, 1\mathcal{U},\, \mathcal{G}_x,\, \mathcal{G}_z & \mathbf{U}_1^{(1)} \\ \mathbf{V}_1,\, 3\mathcal{U},\, \mathcal{G}_x,\, \mathcal{G}_z & \mathbf{U}_1^{(3)} \\ \mathbf{V}_1,\, 5\mathcal{U},\, \mathcal{G}_x,\, \mathcal{G}_z & \mathbf{U}_1^{(5)} \\ \mathbf{V}_1,\, 7\mathcal{U},\, \mathcal{G}_x,\, \mathcal{G}_z & \mathbf{U}_1^{(7)} \\ \mathbf{V}_1,\, 9\mathcal{U},\, \mathcal{G}_x,\, \mathcal{G}_z & \mathbf{U}_1^{(9)}, \end{array} \tag{14}$$

其中 $\mathcal{U}$ 是一个与矩阵 $\mathbf{V}_1$维度相同的单位矩阵。 $\mathcal{G}_x$ 和 $\mathcal{G}_z$ 是包含$\mathbf{V}_1$ 离散化所在网格信息的矩阵。训练 RP-WNO 集成时使用了 500 个速度模型。此外，频域中的压力波场将用复数表示。因此，压力波场 $u(x, z; w)$ 的实部和虚部分别使用两个不同的 RP-WNO 集成进行训练。图 8 展示了 RP-WNO 对.....实部的平均预测，并与真实值进行了比较

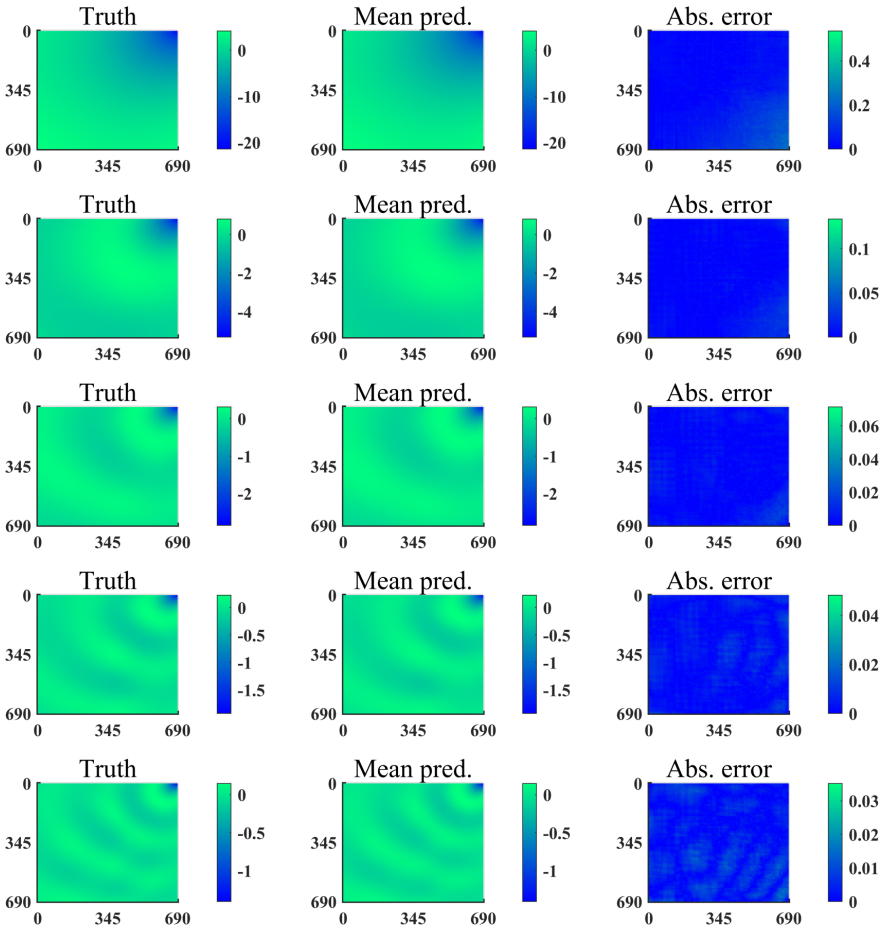
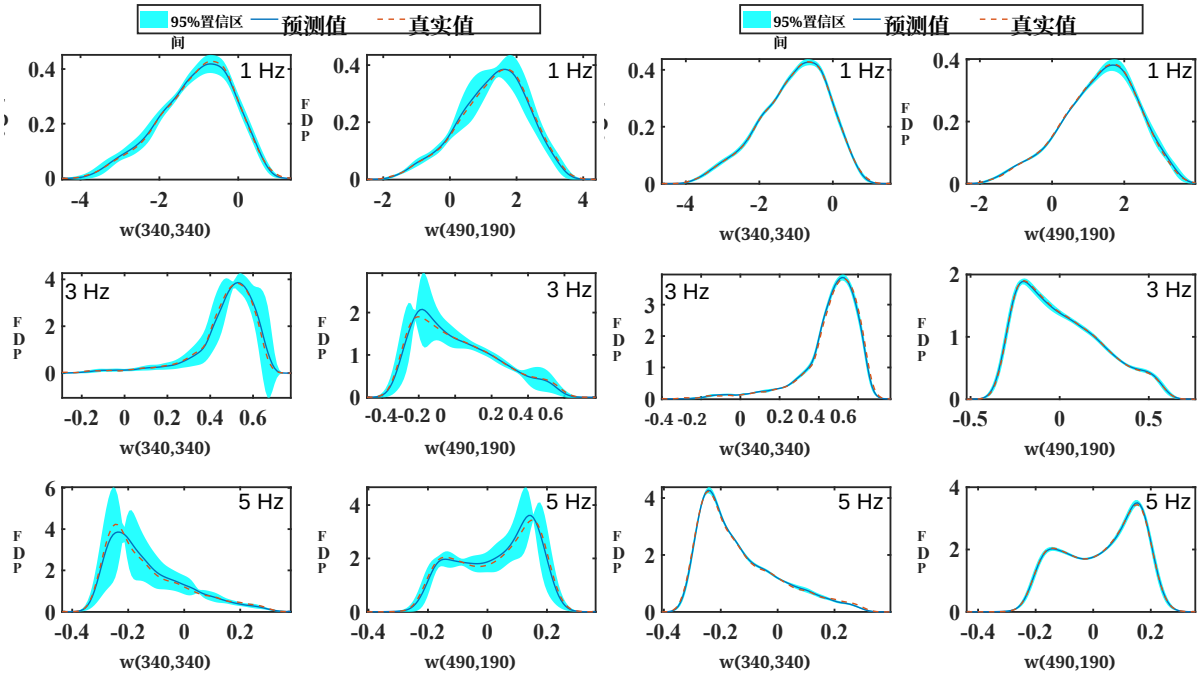


图 8：使用训练好的 RP-WNO集成，针对一个速度场样本得到的压力波场预测（实部）。顶部显示频率为 1 赫兹时的结果，底部显示频率为 9 赫兹时的结果。中间结果从上到下分别对应频率为 3 赫兹、5 赫兹和 7 赫兹的情况。

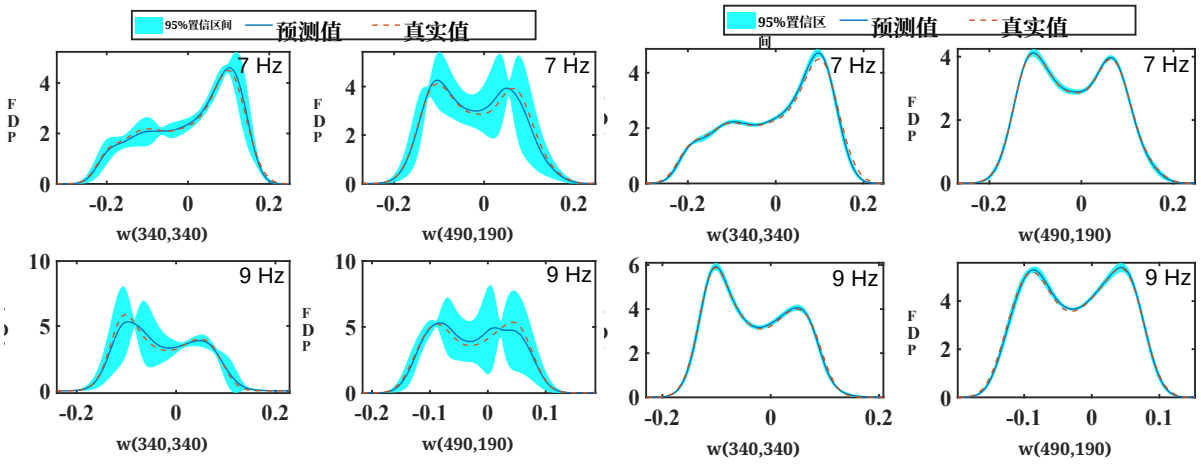
压力波场。从上到下，所示结果分别对应频率为 1 赫兹、3 赫兹、5 赫兹、7 赫兹和 9 赫兹的情况；可以看出，RP-WNO 的平均预测值与真实值高度吻合。

图 9 和图 10 展示了不同频率下压力波场实部的偶然不确定性估计。为获得所述结果，使用了五千种不同的速度模型。图 9 和图 10 的 (a) 部分展示了采用离散小波变换的 RP-WNO集成所对应的结果。在该方法中，离散小波变换的四个小波系数中，训练时仅学习近似分量和水平分量。尽管使用离散小波变换训练的 RP-WNO模型的平均预测值



(a) 离散小波变换 (b) 连续小波变换

图9：压力波场（实部）的偶然不确定性估计，分别针对不同频率（1赫兹、3赫兹和5赫兹）计算。概率密度函数在两个位置处计算：(i)$x = 340, z = 340$ 和(ii)$x = 490, z = 190$。



(a) 离散小波变换 (b) 连续小波变换

图10：压力波场（实部）的偶然不确定性估计，对应于不同频率（7赫兹和9赫兹）。在两个位置计算概率密度函数：(i)$x = 340, z = 340$和(ii)$x = 490, z = 190$。

能够合理近似真实值，但由于预测结果存在细微偏差，预测中的不确定性很大。

另一方面，图 9 和图 10 的 (b) 部分展示了采用连续小波变换的 RP-WNO集成的预测结果。在这种情况下，模型学习了连续小波变换的所有小波系数。可以看出，平均预测值与真实值非常接近，因此预测结果的不确定性范围也较窄。

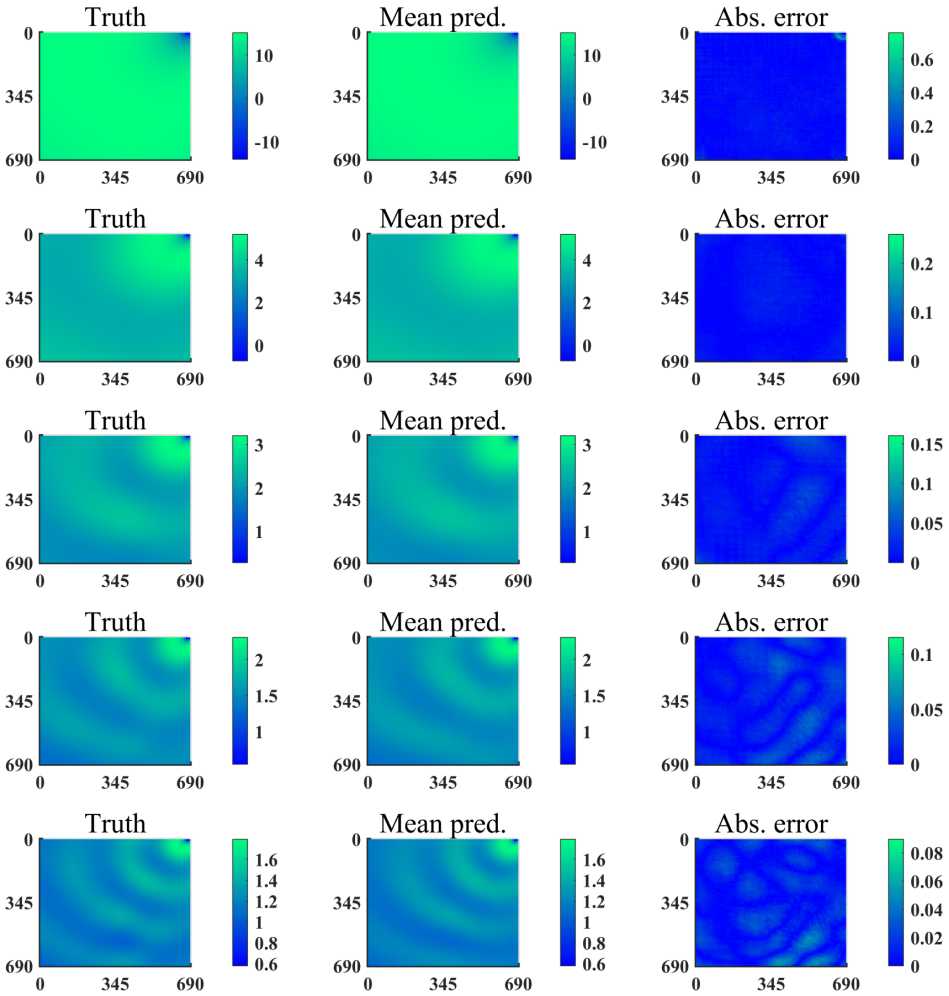


图11：使用训练好的RP-WNO集成获得的、对应于一个速度场样本的压力波场预测（虚部）。顶部给出了频率等于1赫兹时的结果，底部给出了频率等于9赫兹时的结果。中间结果从上到下分别对应于3赫兹、5赫兹和7赫兹的频率。

图 11 展示了压力波场虚部的 RP-WNO 平均预测值与真实值的对比。从上到下，结果分别对应频率为 1 赫兹、3 赫兹、5 赫兹、7 赫兹和 9 赫兹的情况。可以看出，训练所得 RP-WNO集成的平均预测值与真实值非常接近。

图12和图13展示了对应于不同频率的压力波场实部的偶然不确定性估计。为获得所述结果，使用了五千个不同的速度模型。预测结果是使用采用连续小波变换的RP-WNO模型获得的。连续小波变换的所有小波系数均经过学习；与压力波场实部的偶然不确定性估计类似，平均预测值紧密跟随真实值，从而产生一个紧密的相关不确定性界限。

3.4 CS-IV：介观尺度性质

在讨论了一些颇为传统的算子学习问题后，第四个案例研究将探讨如何将晶体固体中的原子结构缺陷映射到介观尺度性质。材料合成过程中结构缺陷不可避免，而这些缺陷也会影响电导率、弹性模量等材料性质。直接依据原子结构估算材料性质并非易事，需要大量模拟和实验；因此，探究所提出的 RP-WNO 在学习二者之间映射方面的有效性是值得的。

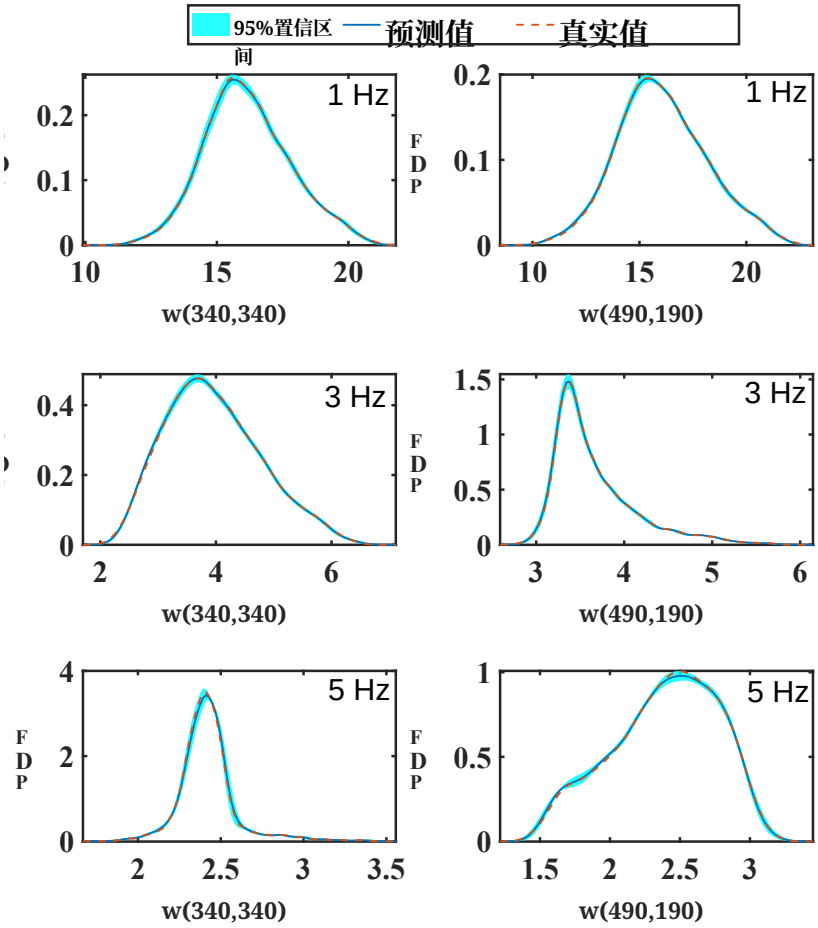


图12：不同频率（1赫兹、3赫兹和5赫兹）下压力波场（虚部）的偶然不确定性估计。

在本案例研究中，映射的两端分别是尺寸为 $127.9\,\mathring{\mathrm{A}} \times 127.8\,\mathring{\mathrm{A}}$ 的二维多孔石墨烯片的初始原子结构，以及对应于 5% 拉伸应变的原子应力。通过随机改变多孔石墨烯片中空位的位置，生成不同样本。本案例研究所用数据集取自 [51], ，建议读者参阅该来源，了解数据集生成的更多详情。

图 14 展示了 RP-WNO 的预测结果与真实值的对比。RP-WNO集成使用 1500 个训练样本进行训练，平均预测值与真实值高度吻合。图15展示了预测集成中一个样本对应的置信区间和平均预测值。可以看出，平均预测值总体上很好地近似了真实值；对于出现偏差的情况，真实值几乎总是落在置信区间内。

CS-I	CS-II	CS-III						CS-IV
0.080	0.095	频率	1Hz	3Hz	5Hz	7Hz	9Hz	2.527
		实部	0.082	0.100	0.250	0.674	1.918	
		虚部	0.002	0.003	0.005	0.008	0.013	

表 3：各案例研究的归一化均方误差百分比。

各种案例研究的百分比归一化均方误差值见表 3。

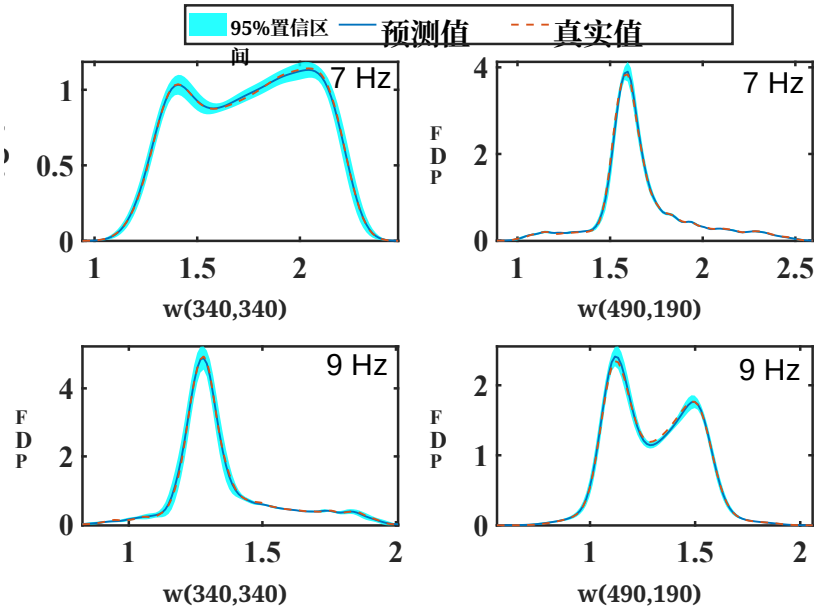


图13：遍历 压力波场（虚部）的不确定性估计，对应于不同频率 （7赫兹和9赫兹）。

4 结论

本文提出了一种新颖的随机先验小波神经算子，这是一种能够估计其预测结果相关不确定性的算子学习框架。尽管在谈到不确定性量化时，人们常常会提及基于贝叶斯推断的技术，但这类技术的复杂性和高昂的资源需求，可能会阻碍其与复杂深度学习架构的结合。所提出的算法在这方面具有优势，因为它采用确定性框架，因此可以轻松地应用于复杂的深度神经网络架构。所提出的 RP-WNO 基于 WNO，利用小波变换和卷积层学习最相关的信息，同时将整个网络的参数需求保持在相对较低的水平。

所提出的 RP-WNO 采用集成训练方案进行训练，从而能够在预测阶段进行不确定性量化。与传统集成训练相比，其优势在于借助随机先验网络，可在训练过程中编码先验信息。在涉及决策的应用中，预测阶段的不确定性量化具有重要作用，而高效的算子学习框架本身也可用作众多应用中的代理模型。为检验所提出的框架，我们开展了四项不同的案例研究，并将预测结果与真实值进行了比较。根据所得结果，我们得出以下观察结论：

- 所有案例研究中的平均预测值都与真实值高度吻合，并对其进行了良好近似。
- 在 CS-I 和 CS-II 中开展的训练数据案例研究表明，随着训练数据增加，且平均预测值逐渐收敛于真实值，使用所提出的 RP-WNO 得到的置信区间宽度会减小。这是预期的表现，也展现了所提出的 RP-WNO 估计其预测结果所包含真实不确定性的能力。
- 在 CS-III 中，对比采用离散小波变换和连续小波变换的 RP-WNO集成所获得的预测结果可以看出，尽管连续小波变换需要更多资源，但其结果更接近真实值。
- 此外，在训练过程中观察到，在 RP-WNO 架构中使用连续小波变换时，隐藏层所需的节点数量会减少，从而在一定程度上抵消了执行连续小波变换及学习其各种小波系数所增加的成本。

总而言之，基 于这些结果，可以认为所提出的 RP-WNO 能够高效地完成其既定任务 s
并且 以有效的方式。尽管如此，作者指出，与其他研究一样， s

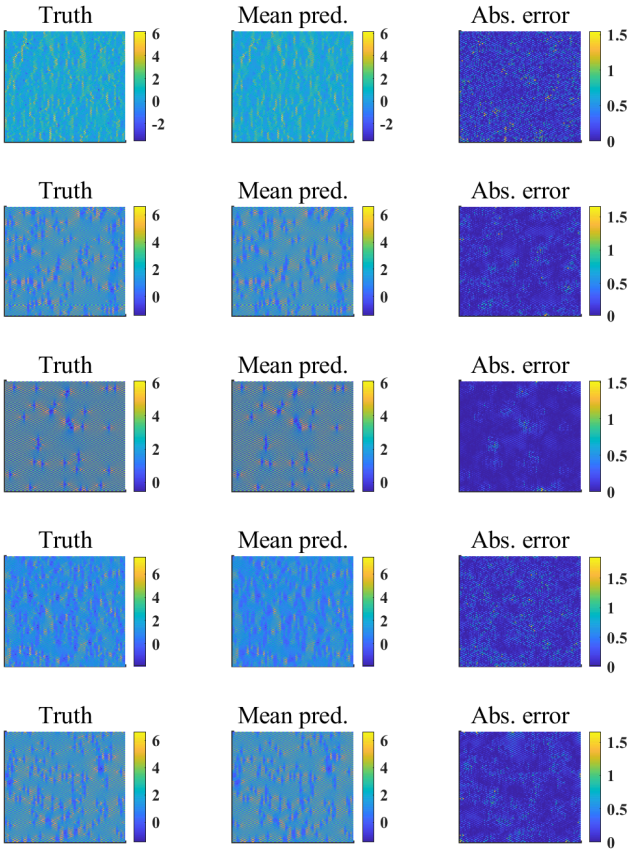


图14：案例研究 IV 中 RP-WNO 的预测与真实值对比。

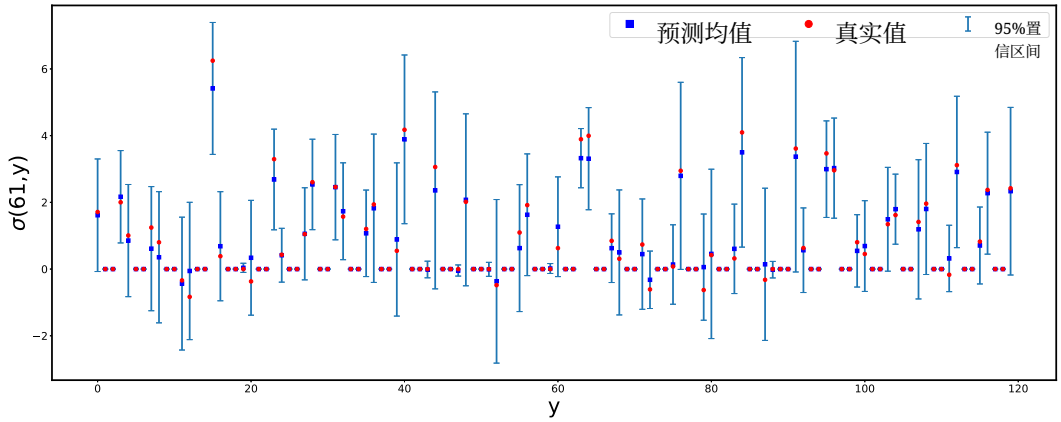
提出的 RP-WNO 框架仍有改进空间。例如，可以开展研究，以降低训练 RP-WNO集成所需的计算成本。此外，目前引入的先验不够模块化，因为固定权重网络与可训练网络采用相同的架构。因此，未来的研究可以探索修改固定权重网络所产生的影响。

致谢

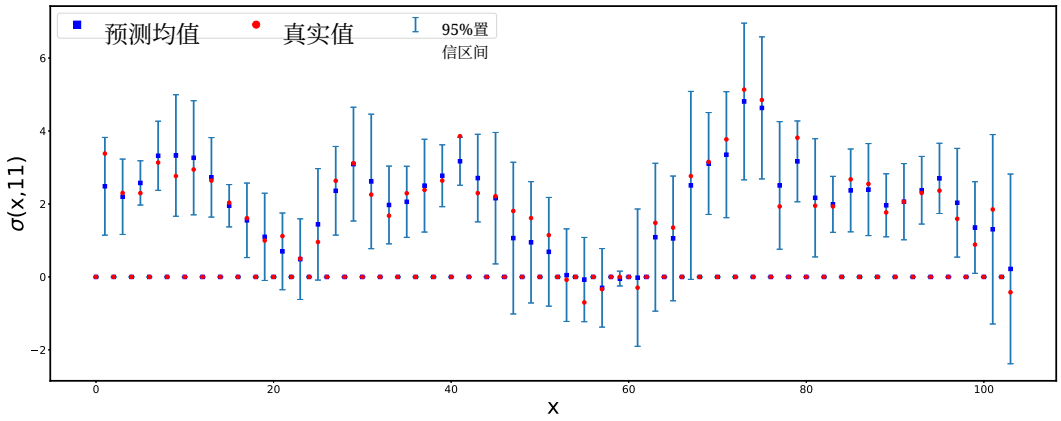
SG 感谢印度教育部以总理研究员（PMRF）奖学金的形式提供的资金支持。SC 感谢科学与工程研究委员会（SERB）通过拨款编号 SRG/2021/000467 提供的资金支持。作者还感谢印度理工学院德里分校应用力学系的博士生 Tapas Tripura，他为小波神经算子中的连续小波变换提供了意见。

参考文献

[1] Julia Slingo 和 Tim Palmer. 天气与气候预测中的不确定性. 《皇家学会哲学汇刊*A*辑：数学、物理与工程科学》，369(1956):4751–4767，2011。[2] TN Palmer、GJ Shutts、R Hagedorn、FJ Doblas-Reyes、Thomas Jung 和 M Leutbecher. 天气与气候预测中模型不确定性的表示. 《地球与行星科学年度评论》，33(1):163–193,



(a)



(b)

图15：展示某一特定初始原子结构的平均预测值和置信区间的图。

2005。[3] FL Moon 和 AE Aktan. 认识不确定性（偏置）对已建（民用）系统结构识别的影响. *Shock and Vibration Digest*，38(5):399–422，2006。[4] Brian R Mace、Keith Worden 和 Graeme Manson. 结构动力学中的不确定性. *Journal of Sound andVibration*，288(3):423–429，2005。[5] Behnam Vakhshouri 和 Shami Nejadi. 混凝土结构长期挠度计算的局限性与不确定性. 载于脆弱性、不确定性与风险：量化、缓解与管理，第535–546页。2014年。[6] Armen Der Kiureghian 和 Ove Ditlevsen. 偶然不确定性还是认识不确定性？这重要吗？*Structural safety*，31(2):105–112，2009。[7] William D Rowe. 理解不确定性. *Risk analysis*，14(5):743–750，1994。[8] R¨udiger Rackwitz. 可靠性分析——综述与若干展望. *Structural safety*，23(4):365–395，2001。[9] Shailesh Garg、Harshit Gupta 和 Souvik Chakraborty. 评估 DeepONet 在随机荷载作用下动力系统时变可靠性分析中的应用. *Engineering Structures*，270:114811，2022。[10] Souvik Chakraborty 和 Rajib Chowdhury. 结构可靠性分析的半解析框架. 应用力学与工程中的计算机方法，289:475–497，2015。[11] R Venkata Rao 和 Vimal J Savsani. 运用先进优化技术进行机械设计优化. 2012年。[12] Tanmoy Chatterjee、Souvik Chakraborty 和 Rajib Chowdhury. 替代模型辅助鲁棒设计优化的批判性综述. *Archives of Computational Methods in Engineering*，26(1):245–274，2019。

[13] Souvik Chakraborty、Tanmoy Chatterjee、Rajib Chowdhury 和 Sondipon Adhikari。基于代理模型的多保真鲁棒设计优化方法。应用数学建模，47:726–744，2017。[14] Pierric Kersaudy、Bruno Sudret、Nad`ege Varsier、Odile Picon 和 Joe Wiart。一种结合克里金法与多项式混沌展开的新型代理建模技术——在计算剂量学不确定性分析中的应用。Journal of Computational Physics，286:103–117，2015。[15] Lukas Novak 和 Drahomir Novak。用于代理建模的多项式混沌展开：理论与软件。Beton-und Stahlbetonbau，113:27–32，2018。[16] Bruno Sudret。多项式混沌展开与随机有限元方法。岩土工程中的风险与可靠性，第265–300页，2014。[17] Dequan Zhang、Xu Han、Chao Jiang、Jie Liu 和 Qing Li。通过响应面法进行时变可靠性分析。机械设计杂志，139(4):041404，2017。[18] Cheng Lu、Yun-Wen Feng、Rhea P Liem 和 Cheng-Wei Fei。采用极值响应面法改进克里金法，用于结构动态可靠性与灵敏度分析。航空航天科学与技术，76:164–175，2018。[19] Christian Gogu 和 Jean-Charles Passieux。结合响应面方法与降阶建模，高效构建代理模型。结构与多学科优化，47(6):821–837，2013。[20] Subrata Chakraborty 和 Arunabh Sen。基于自适应响应面的高效有限元模型更新。有限元分析与设计，80:33–40，2014。[21] Umberto Alibrandi。一种用于随机动态分析的响应面法。可靠性工程与系统安全，126:44–53，2014。[22] Guoshao Su、Lifeng Peng 和 Lihua Hu。一种基于高斯过程的复杂工程结构可靠性分析动态代理模型。Structural Safety，68:97–109，2017。[23] Bin Zhu、Tetsuya Hiraishi、Huafu Pei 和 Qing Yang。结合随机场方法与基于高斯过程回归的代理模型，实现高效边坡可靠性分析。国际岩土力学数值与分析方法杂志，45(4):478–501，2021。[24] Souvik Chakraborty、Sondipon Adhikari 和 Ranjan Ganguli。代理模型在动态系统数字孪生开发中的作用。应用数学建模，90:662–681，2021。[25] Vincent Dubourg、Bruno Sudret 和 Jean-Marc Bourinet。利用克里金代理模型和子集模拟进行基于可靠性的设计优化。结构与多学科优化，44(5):673–690，2011。[26] Yann LeCun、Yoshua Bengio 和 Geoffrey Hinton。深度学习。Nature，521(7553):436–444，2015。[27] Li Deng、Dong Yu 等。深度学习：方法与应用。信号处理领域的基础与趋势®，7(3–4):197–387，2014。[28] Pramila P Shinde 和 Seema Shah。机器学习与深度学习应用综述。载于：2018年第四届计算、通信、控制与自动化国际会议（ICCUBEA），第1–6页。IEEE，2018。[29] Maryam M Najafabadi、Flavio Villanustre、Taghi M Khoshgoftaar、Naeem Seliya、Randall Wald 和 Edin Muharemagic。大数据分析中的深度学习应用与挑战。大数据杂志，2(1):1–21，2015。[30] Justin Ker、Lipo Wang、Jai Rao 和 Tchoyoson Lim。深度学习在医学图像分析中的应用。IEEE Access，6:9375–9389，2017。[31] Amir Mosavi、Mohsen Salimi、Sina Faizollahzadeh Ardabili、Timon Rabczuk、Shahaboddin Shamshirband 和 Annamaria R Varkonyi-Koczy。能源系统中机器学习模型的最新进展：系统综述。Energies，12(7):1301，2019。[32] Sina F Ardabili、Amir Mosavi、Pedram Ghamisi、Filip Ferdinand、Annamaria R Varkonyi-Koczy、Uwe Reuter、Timon Rabczuk 和 Peter M Atkinson。利用机器学习预测新冠肺炎疫情。算法，13(10):249，2020。[33] James Daniell、Kazuma Kobayashi、Dinesh Kumar、Souvik Chakraborty、Ayodeji Alajo、Ethan Taber、Joseph Graham 和 Syed Alam。面向数字孪生的基于物理信息的多阶段深度学习框架开发，用于基于状态的反应堆功率预测。arXiv预印本 arXiv:2211.13157，2022。[34] M Rahman、Abid Khan、Sayeed Anowar、Md Al-Imran、Richa Verma、Dinesh Kumar、Kazuma Kobayashi 和 Syed Alam。将工业4.0、深度学习、代理模型和迁移学习与不确定性量化相结合，应用于核系统数字孪生。arXiv预印本 arXiv:2210.00074，2022。

[35] Romit Maulik、Kai Fukami、Nesar Ramachandra、Koji Fukagata 和 Kunihiko Taira。用于流体流动代理建模与数据恢复的概率神经网络。*Physical ReviewFluids*，5(10):104401，2020。[36] Romit Maulik 和 Omer San。用于湍流盲反卷积的神经网络方法。*JournalofFluidMechanics*，831:151–181，2017。[37] Tapas Tripura 和 Souvik Chakraborty。用于求解计算力学问题中参数化偏微分方程的小波神经算子。*Computer Methods in Applied Mechanics and Engineering*，404:115783，2023。[38] Akshay Thakur、Tapas Tripura 和 Souvik Chakraborty。用于不确定性量化的多保真小波神经算子。*arXiv*预印本 *arXiv:2208.05606*，2022。[39] Lu Lu、Pengzhan Jin、Guofei Pang、Zhongqiang Zhang 和 George Em Karniadakis。基于 DeepONet 和算子通用逼近定理学习非线性算子。*NatureMachineIntelligence*，3(3):218–229，2021。[40] Yibo Yang、Georgios Kissas 和 Paris Perdikaris。利用随机先验实现深度算子网络的可扩展不确定性量化。*arXiv*预印本 *arXiv:2203.03048*，2022。[41] Zongyi Li、Nikola Kovachki、Kamyar Azizzadenesheli、Burigede Liu、Kaushik Bhattacharya、Andrew Stuart 和 Anima Anandkumar。用于参数化偏微分方程的傅里叶神经算子。*arXiv*预印本 *arXiv:2010.08895*，2020。[42] Maziar Raissi、Paris Perdikaris 和 George E Karniadakis。物理信息神经网络：用于求解涉及非线性偏微分方程的正问题和逆问题的深度学习框架。*JournalofComputational physics*，378:686–707，2019。[43] George EP Box 和 George C Tiao。统计分析中的贝叶斯推断。John Wiley & Sons，2011。[44] Dong-xiao Niu、Hui-feng Shi 和 Desheng Dash Wu。利用混合蒙特卡罗算法训练的贝叶斯神经网络进行短期负荷预测。*Applied Soft Computing*，12(6):1822–1827，2012。[45] Theodore Papamarkou、Jacob Hinkle、M Todd Young 和 David Womble。贝叶斯神经网络中马尔可夫链蒙特卡罗面临的挑战。*Statistical Science*，37(3):425–442，2022。[46] Alex Graves。神经网络的实用变分推断。*Advances in Neural Information Processing Systems*，24，2011。[47] Charles Blundell、Julien Cornebise、Koray Kavukcuoglu 和 Daan Wierstra。神经网络中的权重不确定性。载于 *International Conference on Machine Learning*，第1613–1622页。PMLR，2015。[48] Rahul Rahaman 等。深度集成与不确定性量化。*Advances in Neural Information Processing Systems*，34:20063–20075，2021。[49] Ian Osband、John Aslanides 和 Albin Cassirer。用于深度强化学习的随机先验函数。*Advances in Neural Information ProcessingSystems*，31，2018。[50] Chengyuan Deng、Shihang Feng、Hanchen Wang、Xitong Zhang、Peng Jin、Yinan Feng、Qili Zeng、Yinpeng C hen 和 Youzuo Lin。OpenFWI：用于地震全波形反演的大规模多结构基准数据集。*arXiv*预印本 *arXiv:2111.02926*，2021。[51] Zhenze Yang 和 Markus J Buehler。利用图神经网络将原子结构缺陷与晶体固体的介观尺度性质联系起来。*npj Computational Materials*，8(1):1–13，2022。[52] Dan Hendrycks 和 Kevin Gimpel。高斯误差线性单元（GeLU）。*arXiv*预印本 *arXiv:1606.08415*，2016。[53] Ingrid Daubechies。小波十讲。SIAM，1992。

A β项研究

随机先验网络需要将可训练网络和不可训练网络的输出相加，以获得最终输出，并以此对可训练网络进行反向传播。然而，在加入不可训练网络的输出之前，会先将其乘以一个系数 β。该系数由用户定义，并会影响最终结果。在现有文献中，[49], 建议将 β 值设为 1，以获得最佳效果。为评估 β 的影响，本文在将随机先验网络与小波神经算子结合使用时，开展了 β 项研究。

针对 CS-II 中的二维达西流方程训练 RP-WNO 集成，并抽取了 100 个样本用于训练。针对不同的 β 参数训练了多个 RP-WNO 集成，并计算了各集成对应的测试损失。

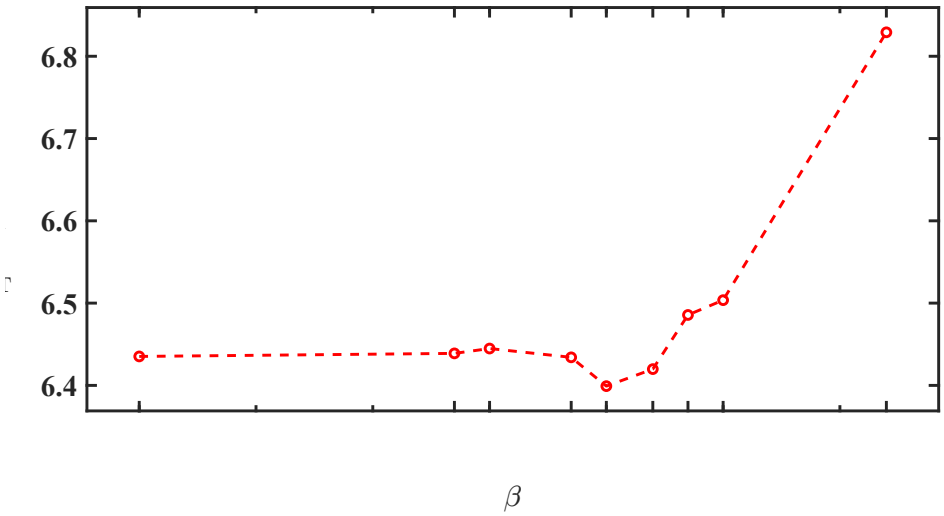


图16：RP-WNO 集成中 β 对测试损失的影响。

图 16 所示结果表明， β 的值过高会对 RP-WNO 集成的性能产生不利影响。此外， β 的值过小会削弱随机先验的作用，从而违背使用随机先验网络的初衷。从图 16 可以看出，当 $\beta = 1$时，损失保持在合理范围内，这进一步印证了现有文献中的发现。作者指出，在 $\beta = 1$ 附近观察到的测试误差额外下降可归因于统计异常；更改编程时使用的随机种子可能会使这一现象略有减弱。

B RP-WNO架构详情

本节将讨论 RP-WNO 集成和 RP-WNO模型的架构详情。在各项案例研究中，使用10个 RP-WNO模型副本训练 RP-WNO 集成。采用高斯误差线性单元 [52] 作为激活函数，并使用 [37] 中采用的相对 $\mathcal{L}_2$ 误差来训练 RP-WNO网络。各项案例研究中所用 RP-WNO架构的详情如下。

B.1 CS-I架构详情

下文讨论的架构适用于可训练网络，不可训练网络也采用相同的架构。CS-I可训练 RP-WNO模型中小波神经算子模块的 WMC 使用离散小波变换（Daubechies小波[53]），并执行八次小波变换以提取相关信息。通过线性全连接层学习最终一次小波变换后得到的小波系数。逆小波变换时，前几次小波变换得到的系数保持不变。此外，在小波神经算子模块的 CMC 中，使用一个卷积核大小为1的卷积层来学习 CMC 的输入。该可训练 RP-WNO模型中使用了四个此类小波神经算子模块。
现在，用于提升可训练 RP-WNO 输入的第一个全连接层有 64 个节点。第四个 WNO 模块之后的最终前向神经网络包含两个全连接层：第一层有 128 个节点，最后的输出层有一个节点。第一个、第二个和第三个小波块之后，以及末尾前向神经网络的两个全连接层之间，均使用 GeLU 激活函数。

B.2 CS-II 架构详情

与 CS-I 类似，CS-II 中的 WMC 使用离散小波变换（Daubechies小波），并进行四次小波变换以提取相关信息。最后一次小波变换后得到的近似小波系数和水平小波系数通过线性全连接层进行学习。先前小波变换得到的系数以及未参与学习的系数，在逆小波变换时保持原样。与 CS-I 的架构类似，WNO 模块中的 CMC 使用一个卷积核大小为 1 的卷积层来学习其输入。在可训练的 RP-WNO模型中使用了四个这样的 WNO 模块。

用于提升可训练 RP-WNO 输入的第一个全连接层有 64 个节点，最后的前向神经网络包含两个全连接层，第一层有 192 个节点，最后的输出层有 1 个节点。在第一个、第二个和第三个小波块之后，以及末尾前向神经网络的两个全连接层之间，均使用 GeLU 激活函数。

B.3 CS-III 架构详情

CS-III 中的 WMC 使用连续小波变换，并执行四次小波变换以提取相关信息。最后一次小波变换后得到的所有小波系数均通过线性全连接层进行学习。此前小波变换得到的系数则在逆小波变换时保持原样。与先前的架构类似，WNO模块中的 CMC 使用一个卷积核大小为 1 的卷积层来学习其输入。可训练 RP-WNO模型中使用了四个这样的 WNO模块。

用于提升 CS-III 中可训练 RP-WNO 输入的第一个全连接层有 64 个节点，而最终的前向神经网络包含两个全连接层：第一层有 64 个节点，最后的输出层只有一个节点。第一个、第二个和第三个小波块之后，以及末尾前向神经网络的两个全连接层之间，均使用 GeLU 激活函数。用于训练 RP-WNO 集成以学习压力波场实部的架构，与用于学习压力波场虚部的架构相同。

B.4 CS-IV 架构详情

CS-IV 中的 WMC 与 CS-III 一样采用连续 WT，并执行三次 WT 以提取相关信息。最终 WT 后获得的所有小波系数均通过线性全连接层进行学习。逆 WT 时，前几次 WT 获得的系数保持不变。与之前的架构类似，WNO 模块的 CMC 使用单个卷积层，卷积核大小为 1，以学习其输入。可训练的 RP-WNO 模型中使用了五个这样的 WNO 模块。

用于提升可训练 RP-WNO 输入的第一个全连接层包含 128 个节点，而最后的前向神经网络包含两个全连接层：第一层有 192 个节点，最后的输出层有一个节点。在第一个、第二个、第三个和第四个小波块之后，以及末尾前向神经网络的两个全连接层之间，均使用 GeLU 激活函数。