Quantifying Uncertainty in Physics-Informed Neural Networks

Jun Kai Yip [1] **Eduardo de Conto** [1 2] **Arvind Easwaran** [1]
[1]*College of Computing and Data Science, Nanyang Technological University, Singapore, Singapore* [2]*CNRS@CREATE, Singapore, Singapore*. Correspondence to: Jun Kai Yip yipj0014@e.ntu.edu.sg, Eduardo de Conto eduardo002@ntu.edu.sg, Arvind Easwaran arvinde@ntu.edu.sg.

1. Introduction

Physics-informed neural networks (PINNs) [1] have emerged as a transformative approach to scientific machine learning, integrating physical laws and constraints into the modeling process. This innovative framework has found applications in various fields, including fluid dynamics, electromagnetism, and material science [2, 3, 4], where it has been shown to improve learning efficiency, improve interpretability, and achieve superior generalization [5].

Despite their advantages, PINNs remain fragile. Most studies evaluate PINN performance using error-based metrics [6], which often fail to provide a complete picture of the model's performance. Indeed, in real-world applications, data can be noisy or incomplete, and models may encounter scenarios that differ from their training environments.

Uncertainty. Therefore, beyond assessing accuracy, it is crucial to assess the *uncertainty* of the model predictions. We can distinguish two types of uncertainty: aleatoric and epistemic uncertainty [7]. Aleatoric uncertainty is the inherent randomness in the data due to sensor noise, measurement errors, or underlying stochasticity in the data and is thus irreducible. Epistemic uncertainty, on the other hand, is the uncertainty in the model's parameters. This can be due to the model's inability to capture the underlying data distribution or insufficient training data (either on the entire input space or in specific regions). Epistemic uncertainty can, in principle, be reduced.

Uncertainty quantification (UQ) [8, 9] provides confidence estimates in conjunction with the predictions, providing us insight into the reliability of model prediction. This is particularly crucial for PINNs as they are often deployed in high-stakes scenarios, such as predicting the behavior of complex physical systems or optimizing critical processes, where the consequences of inaccurate predictions can be severe. A PINN prediction with a high degree of uncertainty should be carefully verified and replaced, e.g., by a fallback method or any number of other appropriate actions to mitigate risks.

Research gap. Despite its significance, traditional PINNs lack robust mechanisms to capture and use uncertainty [10]. Here, we explore state-of-the-art UQ techniques in PINNs, comparing their strengths and weaknesses. We also introduce an approach to quantify uncertainty in PINNs via deep evidential regression (DER) [11] and evaluate it on two canonical PINN benchmarks: Burgers and Laplace equations.

2. Related works

Uncertainty quantification in PINNs remains an open challenge. Several UQ-enhanced extensions of PINNs have been proposed using techniques such as deep ensembles [12], Bayesian learning [13], or spectral expansions into a polynomial basis [14, 10]. Although these approaches have shown promising results, they each have their limitations [8].

Specifically, deep ensembles require training (and inferring) multiple models to infer uncertainty, which can be computationally expensive. Bayesian learning approaches incorporate uncertainty directly into the model's weight during training, thus learning a single model. However, to obtain uncertainty estimates, multiple forward passes are often required during model inference, leading to a similarly high computational cost. Lastly, spectral expansions allow uncertainty to be evaluated during inference in a single forward pass but often do not scale to high-dimensional systems due to rapid growth in the number of polynomial terms required.

In contrast, here we consider the evidential learning approach [11]. Evidential learning holds the promise of an efficient and scalable UQ, but it has not yet been explored in the context of PINNs. Specifically, evidential learning learns a *single model* that can be evaluated in a *single forward pass*, like spectral expansions. In contrast to spectral expansions, it only requires learning the hyperparameters of an evidential distribution, giving it better scaling properties. The closest to our approach is Tan et al. [15], which also considers evidential learning in PINNs. Both methods were developed independently, with neither party aware of the other's research initially. Our work distinguishes itself by comparing standard PINNs with evidential PINNs and also by considering different metrics and benchmarks.

3. The PINN-DER method

PINNs fundamentally work by training a neural network on a modified loss function. Instead of relying solely on data-driven loss functions, PINNs also use the residual obtained from the partial differential equation (PDE) describing the system's physics. This added prior knowledge can improve the training process, but traditional PINNs still provide only a point estimate prediction of the system's output.

PINN-DER. To address this, we integrate evidential learning [11] via DER with PINNs. In a nutshell, DER works by assuming that the target distribution follows a Gaussian with unknown mean μ and variance σ^2, a normal-inverse-gamma (NIG) distribution

Table 1: Error metrics for Burgers and Laplace equations under PINN and PINN-DER models.

Metric	**PINN**		**PINN-DER**	
	Burgers	**Laplace**	**Burgers**	**Laplace**
Max Error	1.6624	0.7147	**1.2026**	**0.6690**
MSE	0.0179	0.0027	0.0385	0.0108
NRMSE	0.0655	0.0339	0.0927	0.0845
RMSE	0.1310	0.0399	0.1854	0.0996

can act as a conjugate prior [16, 17, 18]. DER aims to learn the parameters of the NIG distribution by minimizing the negative log-likelihood (NLL) of the true output given the predicted distribution. The learned NIG distribution allows us to directly obtain the mean $\mathbb{E}[\mu]$, aleatoric $\mathbb{E}[\sigma^2]$, and epistemic $\mathrm{Var}[\mu]$ uncertainty during inference. In contrast, a traditional PINN would predict only $\mathbb{E}[\mu]$. Refer to Appendix A for details on DER.

To implement PINN-DER, we modify the traditional PINN loss function and learn the NIG parameters. Instead of minimizing the NLL of the true output given the predicted distribution, we minimize the PDE residual NLL. Thus, we combine the benefits of both PINNs and DER into PINN-DER. The network learns a distribution modeling the underlying physical equations rather than just fitting observed data points. This distribution can then be used to quantify the uncertainties in the predictions. PINN-DER is available as open-source software (`https://github.com/yipjunkai/pinn-der-ai4x`).

4. Experiments

We evaluate PINN-DER against traditional PINNs on Burgers' and Laplace benchmarks for PINNs. These benchmarks were chosen as they are public benchmarks widely used in PINN research. The *Burgers' equation* is a fundamental PDE used in fluid dynamics, non-linear wave theory, and turbulence modeling. It is used in the testing of PINNs as it contains non-linear and diffusive terms. It is an excellent showcase of the models' ability to learn smooth and discontinuous (shock wave) solutions. Similarly, the *Laplace equation* is used in physics and engineering to describe steady-state heat conduction, electrostatics, and fluid flow. The PDE defining the Laplace equation is relatively simple and often serves as an initial validation.

We observe that PINN-DER without hyperparameter changes reduces the *maximum error* across both benchmarks with *marginal difference* to other metrics, as seen in Table 1. The maximum error measures the worst-case error and can indicate how well the network learned discontinuities and challenging regions of input space [19]. For the Burgers equation, the absolute error (Fig. 1) indicates that PINN-DER can learn a smoother prediction on sharp regions. This can likely be attributed to the model minimizing the epistemic uncertainty in that region. Lastly, PINN-DER can also split the aleatoric and epistemic uncertainties (Fig. 2).

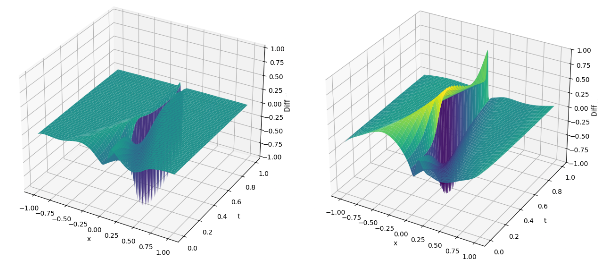

Fig. 1: Absolute error for Burgers equation by PINN (left), PINN-DER (right) after 8000 epochs.

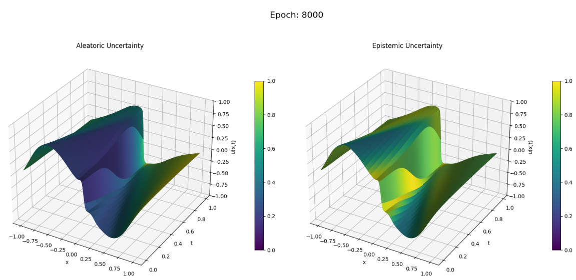

Fig. 2: PINN-DER relative uncertainties for Burgers equation: aleatoric (left), epistemic (right)

5. Conclusion

Uncertainty quantification is critical to ensure the reliability of predictions in real-world applications, but it remains an open challenge in PINNs. Indeed, traditional PINNs lack mechanisms to provide confidence estimates. In this work, we introduce PINN-DER, a framework that enhances PINNs with evidential learning to provide uncertainty estimates alongside PINN prediction. Our results on the Burgers and Laplace equations indicate that PINN-DER effectively captures uncertainty while maintaining predictive accuracy. Furthermore, our approach reduces the maximum error in regions with high variability, reinforcing the role of uncertainty quantification in improving the reliability of PINN. Thus, PINN-DER represents a step toward making PINNs more reliable and practical for real-world use.

Acknowledgments

This research is part of the program DesCartes and is supported by the National Research Foundation, Prime Minister's Office, Singapore under its Campus for Research Excellence and Technological Enterprise (CREATE) program.

References

[1] Maziar Raissi, Paris Perdikaris, and George E Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational physics*, 378:686–707, 2019.

[2] Weiqi Ji, Weilun Qiu, Zhiyu Shi, Shaowu Pan, and Sili Deng. Stiff-PINN: Physics-Informed Neural Network for Stiff Chemical Kinetics. *The Journal of Physical Chemistry A*, 125(36):8098–8106, August 2021.

[3] David Anton and Henning Wessels. Physics-Informed Neural Networks for Material Model Calibration from Full-Field Displacement Data, 2023.

[4] Shengze Cai, Zhiping Mao, Zhicheng Wang, Minglang Yin, and George Em Karniadakis. Physics-informed neural networks (PINNs) for fluid mechanics: A review, 2021.

[5] Siddhartha Mishra and Roberto Molinaro. Estimates on the generalization error of physics-informed neural networks for approximating PDEs. *IMA Journal of Numerical Analysis*, 43(1):1–43, 2023.

[6] Juan Diego Toscano, Vivek Oommen, Alan John Varghese, Zongren Zou, Nazanin Ahmadi Daryakenari, Chenxi Wu, and George Em Karniadakis. From PINNs to PIKANs: Recent Advances in Physics-Informed Machine Learning, 2024.

[7] Eyke Hüllermeier and Willem Waegeman. Aleatoric and epistemic uncertainty in machine learning: An introduction to concepts and methods. *Machine learning*, 110(3):457–506, 2021.

[8] Moloud Abdar, Farhad Pourpanah, Sadiq Hussain, Dana Rezazadegan, Li Liu, Mohammad Ghavamzadeh, Paul Fieguth, Xiaochun Cao, Abbas Khosravi, U. Rajendra Acharya, Vladimir Makarenkov, and Saeid Nahavandi. A review of uncertainty quantification in deep learning: Techniques, applications and challenges. *Information Fusion*, 76:243–297, December 2021.

[9] Apostolos F. Psaros, Xuhui Meng, Zongren Zou, Ling Guo, and George Em Karniadakis. Uncertainty quantification in scientific machine learning: Methods, metrics, and comparisons. *Journal of Computational Physics*, 477:111902, March 2023.

[10] Dongkun Zhang, Lu Lu, Ling Guo, and George Em Karniadakis. Quantifying total uncertainty in physics-informed neural networks for solving forward and inverse stochastic problems. *Journal of Computational Physics*, 397:108850, November 2019.

[11] Alexander Amini, Wilko Schwarting, Ava Soleimany, and Daniela Rus. Deep Evidential Regression, 2020.

[12] Stanislav Fort, Huiyi Hu, and Balaji Lakshminarayanan. Deep Ensembles: A Loss Landscape Perspective, 2020.

[13] 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.

[14] Lukáš Novák, Himanshu Sharma, and Michael D. Shields. Physics-Informed Polynomial Chaos Expansions, 2023.

[15] Hai Siong Tan, Kuancheng Wang, and Rafe McBeth. Evidential Physics-Informed Neural Networks, 2025.

[16] Nis Meinert, Jakob Gawlikowski, and Alexander Lavin. The Unreasonable Effectiveness of Deep Evidential Regression. *Proceedings of the AAAI Conference on Artificial Intelligence*, 37(8):9134–9142, June 2023.

[17] Junyu Gao, Mengyuan Chen, Liangyu Xiang, and Changsheng Xu. A Comprehensive Survey on Evidential Deep Learning and Its Applications, 2024.

[18] Kai Ye, Tiejin Chen, Hua Wei, and Liang Zhan. Uncertainty Regularized Evidential Regression, 2024.

[19] Zhongkai Hao, Jiachen Yao, Chang Su, Hang Su, Ziao Wang, Fanzhi Lu, Zeyu Xia, Yichi Zhang, Songming Liu, Lu Lu, and Jun Zhu. PINNacle: A Comprehensive Benchmark of Physics-Informed Neural Networks for Solving PDEs, 2023.

Appendix A. Deep evidential regression (DER)

DER [11] aims to estimates the mean μ and evidential uncertainty for a regression task. This technique assumes that a target y is drawn i.i.d. from the Gaussian distribution, with unknown mean μ and variance σ^2. Bayesian statistics tell us that a Gaussian prior can act as a conjugate prior for the unknown mean μ, an inverse-Gamma prior for the unknown variance σ^2, and consequently a normal inverse-Gamma (NIG) prior for both mean μ and variance σ^2 [16, 17, 18].

$$\begin{gathered} y \sim \mathcal{N}(\mu, \sigma^2) \\ \mu \sim \mathcal{N}(\gamma, \sigma^2 v^{-1}) \quad \sigma^2 \sim \Gamma^{-1}(\alpha, \beta) \\ (\mu, \sigma^2) \sim \text{NIG}(\gamma, v, \alpha, \beta) \end{gathered} \tag{A1}$$

where $\mathrm{m} = (\gamma, v, \alpha, \beta)$ are the parameters of the NIG.

DER models the mean μ and variance σ^2 of the target y by minimizing the negative logarithm of the marginal likelihood (NLL) as an objective function, with a regularization term (REG) to minimize evidence on incorrect regions.

$$m^* = \arg\min_m \left(\mathcal{L}^{\text{NLL}}(m) + \lambda \mathcal{L}^{\text{REG}}(m)\right) \tag{A2}$$

where m^* represents the optimal parameters, $\mathcal{L}^{\text{NLL}}(m)$ is the NLL loss, $\mathcal{L}^{\text{REG}}(m)$ is the REG term, and λ is the regularization weight controlling the trade-off between the two terms.

The associated mean and uncertainties can then be calculated:

$$\underbrace{\mathbb{E}[\mu] = \gamma,}_{\text{prediction}} \quad \underbrace{\mathbb{E}[\sigma^2] = \frac{\beta}{\alpha - 1},}_{\text{aleatoric}} \quad \underbrace{\text{Var}[\mu] = \frac{\beta}{\nu(\alpha - 1)}}_{\text{epistemic}}.$$

In the case of NIG distribution, an analytical solution exists:

$$p(y \mid \mathbf{m}) = \text{St}\left(y; \gamma, \frac{\beta(1+v)}{v\alpha}, 2\alpha\right) \tag{A3}$$

where St is the Student-t distribution. The corresponding losses are as follows:

$$\begin{aligned} \mathcal{L}_{\text{NLL}} &= \frac{1}{2}\log\left(\frac{\pi}{v}\right) - \alpha \log(\Omega) \\ &+ (\alpha + \frac{1}{2})\log\left((y-\gamma)^2 v + \Omega\right) \\ &+ \Gamma(\alpha) - \Gamma(\alpha + \frac{1}{2}) \end{aligned} \tag{A4}$$

where $\Omega = 2\beta(1+v)$

$$\mathcal{L}_{\text{REG}} = |y - \gamma| \cdot (2v + \alpha) \tag{A5}$$

Implementation. PINN-DER replaces the difference term $(y-\gamma)$ inside Eq. A4 and A5 with the residual of the PDE of the PINN. We also apply a scaling for increased learning rate and L2 normalization.

Appendix B. Experimental Setup

We evaluated the proposed method for UQ on the Burgers and Laplace benchmarks for PINNs.

2.1 PINN

For the Burgers equation, the baseline model consists of a multi-layer perceptron (MLP) with two hidden layers of size 32, using the Tanh activation function. For the Laplace equation, the model is also an MLP but with two hidden layers of size 30, employing the SiLU activation function. The objective function was the PDE of the respective equation.

2.2 PINN-DER

The DER extension only modifies the MLP to output 4 variables (μ, v, α, β). To enforce the non-negative property of (v, α, β), a Softplus function is used, and an additional +1 is added to α. All other hyperparameters remain consistent. The objective function was replaced by the modified Eq. A4 and A5.

2.3 Training Setup

All Burgers' benchmarks were run for 8000 epochs, and all Laplace benchmarks were run for 5000 epochs. PINN and PINN-DER were compared to test the severity impact on performance metrics.

Appendix C. Experimental results

The following figures are the graphical output of the PINN-DER method for both the Burgers and Laplace equations, for their respective domains.

3.1 Burgers results using PINN-DER

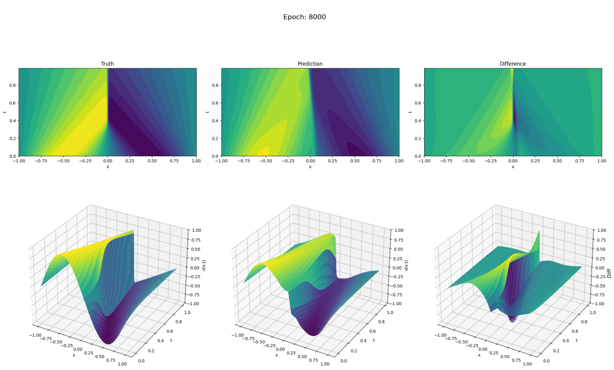

Fig. A1: Comparison of true (left), predicted (center), and absolute difference (right) of Burgers equation using PINN-DER after 8000 epochs.

3.2 *Laplace results using PINN-DER*

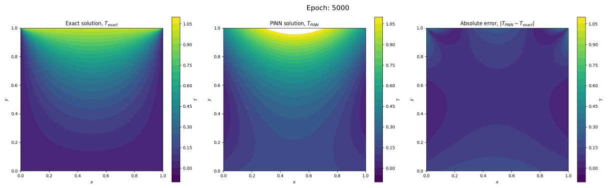

Fig. A2: Comparison of true (left), predicted (center), and absolute difference (right) of Laplace equation using PINN-DER after 5000 epochs.

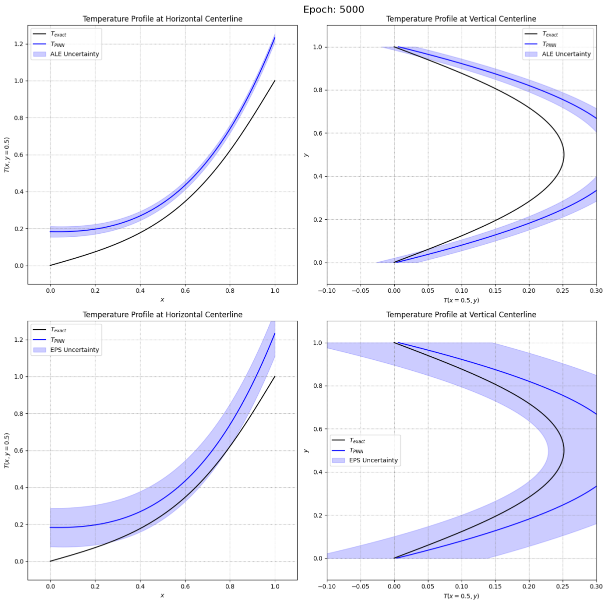

Fig. A3: Temperature Profiles at Key Center-lines with aleatoric (left) and epistemic (right) uncertainties of Laplace equation using PINN-DER after 5000 epochs.