DATA-EFFICIENT NEURAL OPERATOR TRAINING VIA PHYSICS-BASED ACTIVE LEARNING

Alicja Polanska
University College London
United Kingdom Atomic Energy Authority
alicja.polanska.22@ucl.ac.uk

Lorenzo Zanisi
United Kingdom Atomic Energy Authority
lorenzo.zanisi@ukaea.uk

Vignesh Gopakumar
United Kingdom Atomic Energy Authority
University College London
vignesh.gopakumar@ukaea.uk

Stanislas Pamela
United Kingdom Atomic Energy Authority
stanislas.pamela@ukaea.uk

ABSTRACT

Solving partial differential equations with neural operators significantly reduces computational costs but remains bottlenecked by high training data requirements. Active learning offers a natural framework to mitigate this by selectively acquiring the most informative samples in an iterative manner. We introduce physics-based acquisition – a novel physics-informed active learning algorithm that leverages the partial differential equation residual to guide data selection. We validate the method by presenting numerical experiments for the 1D Burgers equation and the 2D compressible Navier-Stokes equations. We show that, in our experiments, physics-based acquisition consistently outperforms random acquisition and matches the state of the art in data efficiency. At the same time, it has the unique advantage of injecting a physics inductive bias into the training process, ensuring that simulation cost is spent where the model's physical understanding is weakest.

1 INTRODUCTION

In the physical sciences, partial differential equations (PDEs) are the language used to describe complex dynamical systems. From modelling plasma dynamics in fusion reactors to simulating galaxy formation, we rely on numerical solvers to predict the evolution of complex and highly parameterised PDE systems. However, numerical solvers can be extremely computationally expensive, which prevents their use in highly iterative control and optimisation workflows. For example, high-fidelity plasma simulations can take several weeks if not months on modern supercomputers (Smith et al., 2020), galaxy formation simulations require millions of CPU hours (Pillepich et al., 2017). To alleviate this computational burden, many machine learning-based methods for approximating PDE solution operators have been proposed, most notably neural operator surrogate models (Li et al., 2020b; Kovachki et al., 2023; Lu et al., 2021; Takamoto et al., 2022).

Neural operators require large training sets to obtain surrogate models that perform well for a desired range of PDE parameters, initial and boundary conditions. As the training data comes from the very simulators the surrogate models are supposed to replace, we are posed with the proverbial 'chicken-and-egg' problem in surrogate modelling (Brandstetter et al., 2022): for very expensive simulators, it is essential to reduce the size of the training sets to make surrogate modelling feasible in the first place. Active learning (AL) offers a natural solution to this problem. As a paradigm designed to minimise labelling costs, AL iteratively selects the most informative unlabelled data points to add to the training set (MacKay, 1992; Ren et al., 2021; Li et al., 2024). This approach is especially useful in settings where obtaining data labels carries a high cost, making it ideal for improving the data efficiency of neural operators. AL for PDEs has been utilised in previous studies for simplified tasks such as the prediction of univariate outputs, or fixed time point solutions (Pestourie et al., 2020; Pickering et al., 2022; Wu et al., 2023a; Bajracharya et al., 2024; Wu et al., 2023b). Recently, AL

has also been explored for the task of full spatio-temporal trajectory prediction with neural operators (Musekamp et al., 2024; Kim et al.).

Most of these works attempt to adapt standard AL tools to PDE learning by, e.g., prioritising acquisition of simulations where an ensemble of surrogates produces the highest variance (Pestourie et al., 2020; Pickering et al., 2022), information-theoretic arguments (Wu et al., 2023b;a), variance reduction methods (Kim et al.) or clustering (Musekamp et al., 2024). Instead, we notice that the residual of the PDE provides a viable metric over the solution accuracy of the neural surrogate. Solutions with high residual values often tend to deviate further from the physical solution. PDE residuals are utilised widely for machine learning frameworks in the context of physics-informed neural networks (PINNs) (Raissi et al., 2019) and have been the main workhorse of PDE analysis and optimisation within numerical and computational physics (Quarteroni & Valli, 2008). In recent work, the physics residual error (PRE) was introduced as an uncertainty measure for neural operators (Gopakumar et al., 2025). The PRE provides a natural and physically-grounded acquisition function for AL: we can improve the surrogate by querying simulations where its current predictions are the most unphysical. Concurrently, PDE residuals have been utilised for single-shot coreset selection for neural operator training (Satheesh et al., 2025). However, such static approaches cannot adapt as the surrogate model improves during training.

Our contributions are as follows: We introduce a novel physics-based acquisition strategy that utilises the PDE residual as a principled physics-informed measure of model epistemic uncertainty; We integrate this strategy into the AL4PDE framework (Musekamp et al., 2024), providing a robust benchmark against established methods; We demonstrate that our physics-based approach achieves competitive data efficiency with state-of-the-art methods, while injecting a physics inductive bias.

2 METHODOLOGY

2.1 ACTIVE LEARNING FOR NEURAL OPERATORS

Neural operators learn mappings between infinite-dimensional function spaces. For a given family of PDEs, the model learns a discretisation-independent solution operator. Although many neural operator architectures have been devised in recent years (Cao et al., 2023; Alkin et al., 2025; Serrano et al., 2024), in this short proof-of-concept work, we focus on Fourier neural operators (FNOs) (Li et al., 2020a) due to their efficiency and cost-accuracy trade-off (de Hoop et al., 2022). The FNO architecture is based on spectral convolutions, where the input to each layer is first converted to the frequency domain via a fast Fourier transform (FFT), then multiplied by a weight matrix in Fourier space, and finally transformed back to the original domain using an inverse FFT (Li et al., 2020a). We condition the model on the PDE parameters by appending the raw parameter values as additional constant channels to the model input (Takamoto et al., 2023), following the AL4PDE implementation (Musekamp et al., 2024).

AL algorithms utilise acquisition functions to sample new training data iteratively from a data pool and have demonstrated data-efficiency gains across classification (Ren et al., 2021) and regression (Holzmüller et al., 2023) compared to the random sampling baseline. AL4PDE is a recently introduced extensible framework for the development and evaluation of AL algorithms for neural PDE solvers (Musekamp et al., 2024). It is implemented as an open-source Python package[1], providing solvers for a variety of parametric PDEs, neural surrogate models and state-of-the-art AL algorithms borrowed from the literature on AL for standard regression problems (Holzmüller et al., 2023). Specifically, acquisition functions pick sets of initial conditions (ICs) and PDE parameters by exploiting surrogate solutions obtained autoregressively from trained surrogate models. Two notable acquisition schemes that achieve state-of-the-art results are largest cluster maximum distance (LCMD) (Holzmüller et al., 2023) and stochastic batch active learning (SBAL) (Kirsch et al., 2023), clustering-based and uncertainty-based AL methods, respectively.

2.2 PHYSICS RESIDUAL ERROR

We extend the AL4PDE framework with a new acquisition function based on the physics residual error (Gopakumar et al., 2025). We can define a PDE governing the dynamics of n field variables

[1] https://github.com/dmusekamp/al4pde

$u \in \mathbb{R}^n$ through the equations:

$$D = D_t(u) + D_X(u; \delta) = 0, \quad X \in \Omega,\ t \in [0, T], \tag{1}$$

where X is the spatial domain bounded by Ω, $[0, T]$ the temporal domain, D_X and D_t the composite operators of the associated spatial and temporal derivatives and δ indicates the parameters that appear in the PDE. The solution is further determined by the boundary condition g and initial condition a, parameterised by λ:

$$u(X, t) = g, \quad X \in \partial\Omega, \tag{2}$$

$$u(X, 0) = a(\lambda, X). \tag{3}$$

The PDE residual is the evaluation of the composite differential operator D over an approximate solution $\hat{u}$ (Morton & Mayers, 2005):

$$D(\hat{u}) - R = 0. \tag{4}$$

For an exact solution, $R = 0$ everywhere in the domain. For approximate solutions from a numerical solver or surrogate model, it gives a quantitative measure of how physical the solution is. The PRE is defined as the residual R estimated over the discretised solution to the PDE. PRE has been used as a measure of convergence and stability in PDE analysis across numerical methods (Quarteroni & Valli, 2008). Recently, it has been adopted as an uncertainty measure for neural operators deployed as surrogate models for a well-defined family of PDEs (Gopakumar et al., 2025).

To estimate the PRE, we followed the method deployed in (Gopakumar et al., 2025)[2]. Finite difference stencils were deployed as convolutional kernels, rendering the PRE estimation as an additive convolutional operation. This method allows for gradient estimation without having access to the computational graph or model parameters and can be accelerated on GPUs.

2.3 Physics-Based Acquisition

Our physics-based acquisition strategy leverages the PRE as a direct measure of the physical accuracy of the surrogate model. For each candidate pair of initial conditions and PDE parameters in the pool (δ, λ), we first rolled out the surrogate model to generate a trajectory. We then calculated an acquisition score $s(\delta, \lambda)$, by taking the mean absolute PRE of the trajectory:

$$s(\delta, \lambda) = \frac{1}{N_s N_t} \sum_{i=1}^{N_s} \sum_{t=1}^{N_t} |\text{PRE}(\mathbf{x}_i, t)|, \tag{5}$$

where $\mathbf{x}_i$ represents the spatial grid points, t the temporal steps, and N_s and N_t the total number of spatial and temporal points, respectively.

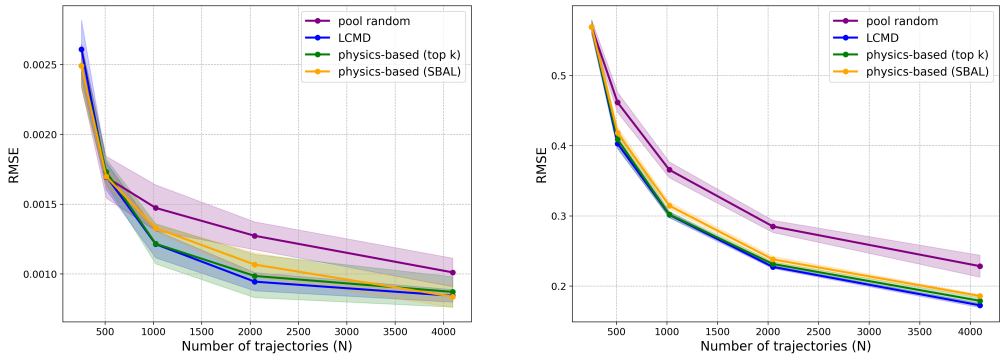
We employed two popular uncertainty-based acquisition strategies: *top k* and stochastic batch active learning (SBAL) (Kirsch et al., 2023). For *top k* acquisition, we ranked the members of the pool by their score, and added the k highest ranked to the training set. For SBAL, pool data were acquired into the training set in a stochastic manner, by perturbing the scores in equation 5 with a power-law noise distribution. This method has been shown to help diversify the acquisition without the need for bespoke batch AL algorithms (Kirsch et al., 2023).

The PDE residuals cannot be directly compared across different parameter choices because changes in coefficients alter the natural scaling and conditioning of the equations, making the same residual magnitude represent different levels of error. In order to account for varying dynamical regimes, we normalised each candidate trajectory's acquisition score by that of the ground-truth trajectory corresponding to the closest training-set member in PDE parameter space, measured by Euclidean distance (Ferziger et al., 2020).

3 Results

To validate its efficiency, we implemented our novel physics-based acquisition algorithm in the AL4PDE benchmark (Musekamp et al., 2024). We made use of their AL experiment set-up, focusing

[2] https://github.com/gitvicky/CP-PRE



(a) Burgers

(b) 2D Compressible Navier Stokes

Figure 1: RMSE as a function of the number of training trajectories (N) for (a) the 1D Burgers equation and (b) 2D compressible Navier-Stokes equation. Solid lines show mean RMSE averaged over multiple seeds, with shaded regions indicating 95% confidence intervals. Our method significantly outperforms random acquisition and achieves similar data efficiency to LCMD.

on the Burgers and 2D compressible Navier-Stokes systems. For details of the architecture and experimental setup, as well as the relevant equations, we refer the reader to (Musekamp et al., 2024). All experiments were executed on a single NVIDIA H100 SXM 80GB GPU. For both examples, we evaluated the model performance measured by root mean square error (RMSE) on the test set as a function of the number of trajectories in the training set N (Figure 1). Our physics-based acquisition strategy was benchmarked against random sampling (pool random) and LCMD. We repeat the experiment for five different random seeds and plot the average over these runs.

Figure 1a shows the comparison between random, LCMD and our physics-based acquisitions for the 1D Burgers equation for different acquisition schemes. The pool range for the kinematic viscosity was set to $\nu \in [0.1, 1)$. Figure 1b shows an analogous plot for the 2D compressible Navier-Stokes equation for different acquisition schemes for the shear and bulk viscosities $\eta, \zeta \in [10^{-2}, 10^{-1})$. It can be seen that our acquisition strategy consistently outperforms random sampling, achieving similar performance for a significantly lower number of training trajectories. The performance is on par with LCMD, which was the best performing method in the benchmark of (Musekamp et al., 2024).

Our results establish that physics-based acquisition is a competitive alternative to purely data-driven strategies on parameter ranges that correspond to moderately turbulent (for compressible Navier-Stokes) and diffusion-dominated (for Burgers) cases. We identify two primary factors that may influence performance for wide parameter ranges: limitations in the FNO's conditioning scheme and availability of sufficient data for accurate PRE normalisation. In future work we plan to explore more parameter-aware normalisation schemes, for instance based on theoretical scaling laws for gradients. While geometric methods like LCMD provide broad coverage of the PDE parameter space, our acquisition offers a unique physics-informed refinement of the model.

4 Conclusions

In this work, we introduced physics-based acquisition, a novel physics-informed AL strategy for generating training datasets from costly PDE simulators efficiently, leveraging the PRE as an uncertainty measure. By preferentially acquiring training data for which the surrogate model produces the most unphysical solutions, our method actively guides the model towards adhering to the governing PDE. Our experiments on the Burgers and 2D compressible Navier-Stokes equations demonstrate that our acquisition strategy significantly improves data efficiency compared to random sampling and achieves performance competitive with established, purely data-driven AL methods like LCMD. This capability is crucial for mitigating the high computational cost associated with generating large training datasets for neural operators. Our findings demonstrate the significant potential of injecting a physics inductive bias to steer data acquisition in complex, compute-bound physics domains.

Our current normalisation strategy utilises local ground-truth information to calibrate the PRE across PDE parameters. Our results demonstrate, that using the PRE provides a strong signal for physics inconsistency, but robust normalisation is needed to ensure this signal is directly comparable across the parameter space for broad parameter range acquisition. While this makes acquisition for solutions across distinct dynamical regimes a challenge, it also makes the method particularly well-suited for applications where dynamics vary continuously with respect to PDE parameters or where selecting initial conditions is the primary goal. In future work, we plan to further investigate and refine the normalisation scheme, as well as apply this methodology to more complex, real-world examples, for instance simulations of plasma dynamics.

REFERENCES

Benedikt Alkin, Andreas Fürst, Simon Schmid, Lukas Gruber, Markus Holzleitner, and Johannes Brandstetter. Universal physics transformers: A framework for efficiently scaling neural operators, 2025. URL https://arxiv.org/abs/2402.12365.

Pradeep Bajracharya, Javier Quetzalcoatl Toledo-Marín, Geoffrey Fox, Shantenu Jha, and Linwei Wang. Feasibility study on active learning of smart surrogates for scientific simulations. *arXiv preprint arXiv:2407.07674*, 2024.

Johannes Brandstetter, Max Welling, and Daniel E. Worrall. Lie point symmetry data augmentation for neural PDE solvers, 2022. URL https://arxiv.org/abs/2202.07643.

Qianying Cao, Somdatta Goswami, and George Em Karniadakis. Lno: Laplace neural operator for solving differential equations, 2023. URL https://arxiv.org/abs/2303.10528.

Maarten V. de Hoop, Daniel Zhengyu Huang, Elizabeth Qian, and Andrew M. Stuart. The cost-accuracy trade-off in operator learning with neural networks, 2022. URL https://arxiv.org/abs/2203.13181.

Joel H. Ferziger, Milovan Perić, and Robert L. Street. *Computational Methods for Fluid Dynamics*. Springer Cham, Switzerland, 4th edition, 2020. ISBN 978-3-319-99693-6. doi: 10.1007/978-3-319-99693-6.

Vignesh Gopakumar, Ander Gray, Lorenzo Zanisi, Timothy Nunn, Daniel Giles, Matt J Kusner, Stanislas Pamela, and Marc Peter Deisenroth. Calibrated physics-informed uncertainty quantification. *arXiv preprint arXiv:2502.04406*, 2025.

David Holzmüller, Viktor Zaverkin, Johannes Kästner, and Ingo Steinwart. A framework and benchmark for deep batch active learning for regression. *Journal of Machine Learning Research*, 24 (164):1–81, 2023.

Yegon Kim, Hyunsu Kim, Gyeonghoon Ko, and Juho Lee. Active Learning with Selective Time-Step Acquisition for PDEs. In *Forty-second International Conference on Machine Learning*.

Andreas Kirsch, Sebastian Farquhar, Parmida Atighehchian, Andrew Jesson, Frédéric Branchaud-Charron, and Yarin Gal. Stochastic batch acquisition: A simple baseline for deep active learning. *Transactions on Machine Learning Research*, 2023.

Nikola Kovachki, Zongyi Li, Burigede Liu, Kamyar Azizzadenesheli, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Neural operator: Learning maps between function spaces with applications to PDEs. *Journal of Machine Learning Research*, 24(89):1–97, 2023. URL http://jmlr.org/papers/v24/21-1524.html.

Dongyuan Li, Zhen Wang, Yankai Chen, Renhe Jiang, Weiping Ding, and Manabu Okumura. A survey on deep active learning: Recent advances and new frontiers, 2024. URL https://arxiv.org/abs/2405.00334.

Zongyi Li, Nikola Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Fourier neural operator for parametric partial differential equations. Oct 2020a. doi: 10.48550/arXiv.2010.08895. Funding by Kortschak Scholars Program.

Zongyi Li, Nikola B. Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Kaushik Bhattacharya, Andrew M. Stuart, and Anima Anandkumar. Neural operator: Graph kernel network for partial differential equations. *CoRR*, abs/2003.03485, 2020b. URL https://arxiv.org/abs/2003.03485.

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, March 2021. ISSN 2522-5839. doi: 10.1038/s42256-021-00302-5. URL http://dx.doi.org/10.1038/s42256-021-00302-5.

David J. C. MacKay. Information-based objective functions for active data selection. *Neural Computation*, 4(4):590–604, Jul 1992. ISSN 0899-7667. doi: 10.1162/neco.1992.4.4.590. Funding by Caltech Fellowship.

Keith W Morton and David Francis Mayers. *Numerical solution of partial differential equations: an introduction*. Cambridge university press, 2005.

Daniel Musekamp, Marimuthu Kalimuthu, David Holzmüller, Makoto Takamoto, and Mathias Niepert. Active learning for neural PDE solvers. *CoRR*, abs/2408.01536, 2024. doi: 10.48550/ARXIV.2408.01536. URL https://doi.org/10.48550/arXiv.2408.01536.

Raphaël Pestourie, Youssef Mroueh, Thanh V Nguyen, et al. Active learning of deep surrogates for pdes: application to metasurface design. *Computational Materials*, 6(1):164, 2020.

Ethan Pickering, Stephen Guth, George Em Karniadakis, and Themistoklis P Sapsis. Discovering and forecasting extreme events via active learning in neural operators. *Nature Computational Science*, 2(12):823–833, 2022.

Annalisa Pillepich, Dylan Nelson, Lars Hernquist, Volker Springel, Rüdiger Pakmor, Paul Torrey, Rainer Weinberger, Shy Genel, Jill P Naiman, Federico Marinacci, and Mark Vogelsberger. First results from the illustristng simulations: the stellar mass content of groups and clusters of galaxies. *Monthly Notices of the Royal Astronomical Society*, 475(1):648–675, 12 2017. ISSN 0035-8711. doi: 10.1093/mnras/stx3112. URL https://doi.org/10.1093/mnras/stx3112.

Alfio M. Quarteroni and Alberto Valli. *Numerical Approximation of Partial Differential Equations*. Springer Publishing Company, Incorporated, 1st ed. 1994. 2nd printing edition, 2008. ISBN 3540852670.

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.

Pengzhen Ren, Yun Xiao, Xiaojun Chang, Po-Yao Huang, Zhihui Li, Brij B. Gupta, Xiaojiang Chen, and Xin Wang. A survey of deep active learning, 2021. URL https://arxiv.org/abs/2009.00236.

Anirudh Satheesh, Anant Khandelwal, Mucong Ding, and Radu Balan. PICore: Physics-informed unsupervised coreset selection for data efficient neural operator training. *Transactions on Machine Learning Research*, 2025. ISSN 2835-8856. URL https://openreview.net/forum?id=l0VSewTJCI.

Louis Serrano, Thomas X Wang, Etienne Le Naour, Jean-Noël Vittaut, and Patrick Gallinari. Aroma: Preserving spatial structure for latent pde modeling with local neural fields, 2024. URL https://arxiv.org/abs/2406.02176.

S.F. Smith, S.J.P. Pamela, A. Fil, M. Hölzl, G.T.A. Huijsmans, A. Kirk, D. Moulton, O. Myatra, A.J. Thornton, H.R. Wilson, and the JOREK team. Simulations of edge localised mode instabilities in mast-u super-x tokamak plasmas. *Nuclear Fusion*, 60(6):066021, may 2020. doi: 10.1088/1741-4326/ab826a. URL https://dx.doi.org/10.1088/1741-4326/ab826a.

Makoto Takamoto, Timothy Praditia, Raphael Leiteritz, Daniel MacKinlay, Francesco Alesiani, Dirk Pflüger, and Mathias Niepert. PDEBench: An extensive benchmark for scientific machine learning. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh (eds.), *Advances in Neural Information Processing Systems*, volume 35, pp. 1596–1611. Curran Associates, Inc., 2022. URL https://proceedings.neurips.cc/paper_files/paper/2022/file/0a9747136d411fb83f0cf81820d44afb-Paper-Datasets_and_Benchmarks.pdf.

Makoto Takamoto, Francesco Alesiani, and Mathias Niepert. Learning neural pde solvers with parameter-guided channel attention, 2023. URL https://arxiv.org/abs/2304.14118.

Dongxia Wu, Ruijia Niu, Matteo Chinazzi, Yian Ma, and Rose Yu. Disentangled multi-fidelity deep bayesian active learning. In *International Conference on Machine Learning*, pp. 37624–37634. PMLR, 2023a.

Dongxia Wu, Ruijia Niu, Matteo Chinazzi, Alessandro Vespignani, Yi-An Ma, and Rose Yu. Deep bayesian active learning for accelerating stochastic simulation. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 2559–2569, 2023b.