GPGreen: Learning Linear Operators with Gaussian Processes

Thomas Cowperthwaite
Department of Applied Mathematics and Theoretical Physics, University of Cambridge
Cambridge, UK
tc656@cam.ac.uk

Henry Moss
School Of Mathematical Sciences,
Lancaster University
Lancaster, UK
henry.moss@lancaster.ac.uk

Abstract

Operator learning has emerged as a promising data-driven approach to emulating solutions of partial differential equations. Existing deep learning-based models lack principled uncertainty quantification, rely on access to large numbers of training examples, and remain largely uninterpretable. Here, we use Gaussian process regression to make uncertainty-aware estimates of PDE solutions. We show our method is competitively accurate compared to existing approaches, while additionally providing uncertainty quantification and improving sample efficiency. The framework exploits Kronecker structures and Fast Fourier Transforms to achieve resolution-invariant prediction cost scaling.

1 Introduction

Machine learning models have been shown to be capable emulators (or surrogate models) of complex functions describing properties of natural and technological systems, such as those found in weather forecasting [1], groundwater flow [2], and cardiology [3]. More recently, the field of operator learning has extended these ideas to develop emulators of systems described by partial differential equations (PDEs) [4–10]. These approaches attempt to learn an *operator* mapping between function spaces, such as that which maps a PDE forcing term to its solution. These problems typically exist on some (usually physical) domain, under the constraint of some boundary conditions.

Operator learning has the potential to revolutionise scientific machine learning, and neural network-based approaches have already found utility in a number of domains, including medium-range weather forecasting [11, 12] and magnetohydrodynamics [13]. Current approaches do not provide principled estimates of uncertainty, which would allow for Bayesian active learning and experimental design [14–16]. In this work, we introduce a novel Bayesian operator learning framework and demonstrate its well-calibrated predictive power and understandability via easy access to the learnt Green's function.

2 Related Work

Operator Learning. PDE emulators such as DeepONet [17], PCA-Net [18, 19], Neural Operators [20], and the Fourier Neural Operator [5] decouple discretisation from operator estimation, enabling mesh-invariant predictions. These methods excel on nonlinear tasks but lack principled uncertainty estimates and interpretable features. Kernel learning has also been applied to operator estimation, again without uncertainty estimates [6].

Green's Function Discovery. Another line of work estimates Green's functions directly with neural networks. DeepGreen is effective for solution prediction, however it fails to recover accurate Green's functions, limiting interpretability [21]. GreenLearning restricts to linearised boundary value

1st Workshop on Differentiable Systems and Scientific Machine Learning @ EurIPS 2025.

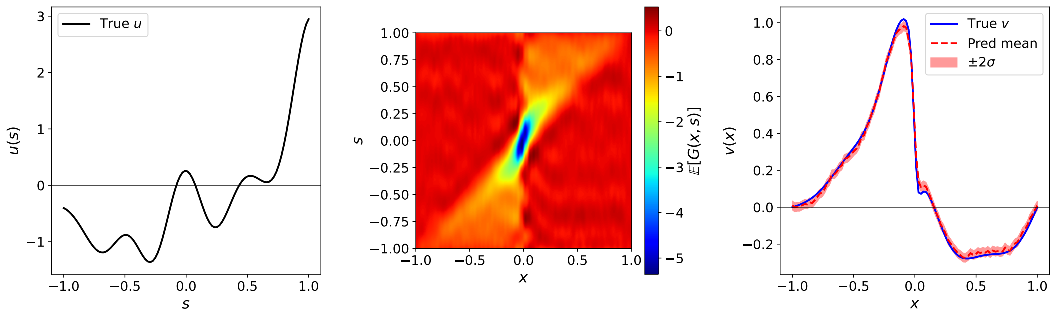

Figure 1: Demonstration of our proposed GPGreen framework. We build a probabilistic emulation of the PDE describing a viscous shock in one spatial dimension, mapping a forcing $u(\boldsymbol{s})$ (*left*) to a response $v(\boldsymbol{x})$ (*right*) by using a Gaussian process to learn the underlying Green's function $G(\boldsymbol{x}, \boldsymbol{s})$ governing the PDE's integral operator (*centre*).

problems (BVPs) but more successfully recovers Green's functions, yielding accurate deterministic predictions [10]. Our work builds on this by additionally providing uncertainty quantification.

3 Problem Statement

Without loss of generality, we consider a rectangular domain $\Omega = \left[-\frac{1}{2}, \frac{1}{2}\right]^D$, a linear differential operator $\mathcal{L}$ (the solution operator to which we aim to learn), alongside a known linear operator $\mathcal{B}$ enforcing constraints on the BVP solution. Together, we have an inhomogeneous BVP of the form

$$\mathcal{L}v(\boldsymbol{x}) = u(\boldsymbol{x}), \quad \mathcal{B}\left(v, \Omega\right) = g(\boldsymbol{x}), \quad \boldsymbol{x} \in \Omega \tag{1}$$

Such systems can generally be solved as the sum of a particular integral and a homogeneous solution, where the particular integral is found through integration against a Green's function G

$$v(\boldsymbol{x}) = v_p(\boldsymbol{x}) + v_h(\boldsymbol{x}) = \int_{\Omega} G(\boldsymbol{x}, \boldsymbol{s}) u(\boldsymbol{s}) \, \mathrm{d}\boldsymbol{s} + v_h(\boldsymbol{x}).$$

Finding the Green's Function of a system is often impossible analytically, so our work is concerned with learning the particular integral from data. In the following we denote v_p as v for brevity. Future work will extend our approach to simultaneously learn the homogeneous solution v_h.

Data. The aim of our model is to rapidly and robustly solve inhomogeneous BVPs by learning the integral operator encoded by the Green's function. We have training data $\{\mathcal{D}_u, \mathcal{D}_v\} = \{u_i(\boldsymbol{x}_m), v_i(\boldsymbol{x}_m)\}_{i=1}^{N_{\text{train}}}$, where the functions u_i and v_i solve the BVP and are sampled at the same M locations in a Cartesian grid, $m = 1, \ldots M$. The domain can be sampled at different resolutions in each dimension, so $M = \prod_d M_d$, where M_d is the resolution in dimension d.

4 GPGreen

Our primary contribution is to leverage the Green's function solution of a linear BVP to define a prior over operators, leading to accurate predictions of PDE solutions. In particular, we place a GP prior on the Green's function with a separable, stationary kernel $k = k_{\boldsymbol{x}}(\boldsymbol{x} - \boldsymbol{x}')k_{\boldsymbol{s}}(\boldsymbol{s} - \boldsymbol{s}')$, *i.e.*

$$G(\boldsymbol{x}, \boldsymbol{s}) \sim \mathcal{GP}\left(0, k((\boldsymbol{x}, \boldsymbol{s}), (\boldsymbol{x}', \boldsymbol{s}'))\right).$$

As GPs are closed under the action of linear operators and ϵ and G are independent, v is also distributed as a GP with a prior derived from our prior on G [22, 23]. We could directly regress an unseen real-space function v_* against the forcing u_*, calculating $p\left(v_* | \mathcal{D}_u, \mathcal{D}_v, u_*\right)$. However this paradigm is computationally intractable for reasonable resolutions M_d, with time and space complexities of $\mathcal{O}(N_{\text{train}}^3 M^3)$ and $\mathcal{O}(N_{\text{train}}^2 M^2)$, respectively.

The Fourier basis has been successfully used for operator learning in previous work due to its spectral sparsity for smooth fields [5]. We therefore choose to condition our GP on Fourier modes of the output function $\{\mathcal{D}_{\hat{u}}, \mathcal{D}_{\hat{v}}\} = \{\hat{u}_i(\boldsymbol{\omega}_k), \hat{v}_i(\boldsymbol{\omega}_k)\}_{i=1}^{N_{\text{train}}}$, where $\hat{(\cdot)} = \mathcal{F}[\cdot]$ represents a function in frequency space, defined by the *linear* periodic Fourier operator

$$\mathcal{F}[f](\boldsymbol{\omega}_k) = \int_\Omega e^{-i\boldsymbol{\omega}_k \boldsymbol{x}} f(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x}. \tag{2}$$

Here, k is shortened notation for the indices $\{k_d\}_{d=1}^{D}$, where each index is such that $-k_{\max} \leq k_d \leq k_{\max}$, i.e. the same truncation in each dimension. As the Fourier transform is a linear operator, $\hat{v}$ is distributed as a complex multivariate Gaussian distribution:

$$\hat{v}\,(\boldsymbol{\omega}_k) \sim \mathcal{CN}(\mu, \hat{K}, \hat{R}).$$

The prior on $\hat{\boldsymbol{v}}$ has mean $\mu = 0$ and covariance given by $\hat{K} = k_u(u, u')k_{\boldsymbol{\omega}}(\boldsymbol{\omega}_k, \boldsymbol{\omega}_{k'}) + k_n(\boldsymbol{\omega}_k, \boldsymbol{\omega}_{k'})$ where we define a function similarity kernel functional k_u, a frequency kernel $k_{\boldsymbol{\omega}}$ derived from the real-space kernel $k_{\boldsymbol{x}}$, and a kernel k_n encoding the covariance structure of our real-space i.i.d. noise in frequency space. Due to conjugate symmetry between $\hat{K}$ and $\hat{R}$, the expression derived above is equally helpful for calculating the pseudo-covariance: $\hat{R}(u, u', \boldsymbol{\omega}_k, \boldsymbol{\omega}_{k'}) = \hat{K}(u, u', \boldsymbol{\omega}_k, -\boldsymbol{\omega}_{k'})$.

We express k_u in terms of Fourier transforms of our forcings u and the kernel $k_{\boldsymbol{s}}$, which we compute efficiently through fast Fourier transform algorithms. It is possible to express $k_{\boldsymbol{\omega}}$ in closed form, which we calculate for the Matérn-$\frac{1}{2}$ kernel for this work. The noise covariance structure k_n may also be computed using a closed form expression.

Using standard complex GP regression on our discretely-sampled data, we condition on the measured training set of Fourier modes to obtain the parameters of the posterior distribution over the Fourier modes of our test data, $p(\hat{\boldsymbol{v}}_* | \mathcal{D}_{\hat{u}}, \mathcal{D}_{\hat{v}}, \hat{\boldsymbol{u}}_*)$. This distribution can now be transformed with the *linear* inverse Fourier transform to obtain a posterior over possible real-space PDE solutions. As the transformation is linear, this real-space posterior is also a Gaussian process.

Kronecker Structure Allows Scalability. If implemented naïvely, the predictive distribution over the Fourier modes would be computationally prohibitive, due to the growth of the Gram matrices with increasing N_{train} or $k_{\max}$. As such, we exploit the Kronecker structure of our kernels and reduce the regression time complexity from $\mathcal{O}\left(N_{\text{train}}^3 k_{\max}^{3D}\right)$ to $\mathcal{O}\left(N_{\text{train}} k_{\max}^{D}\right)$ and peak memory requirements from $\mathcal{O}\left(N_{\text{train}}^2 k_{\max}^{2D}\right)$ to $\mathcal{O}\left(N_{\text{train}} k_{\max}^{D}\right)$, translating to a 50× acceleration for $k_{\max} = 50$, $N_{\text{train}} = 75$ [24].

Hyperparameter Tuning. To find optimal kernel hyperparameters, we use a gradient-based maximum-likelihood estimate from the negative marginal log-likelihood loss, calculated on predicted Fourier modes, i.e. we minimise $\mathcal{L}(\theta) = -\log p\,(\hat{\boldsymbol{v}} | \boldsymbol{u}, \theta)$ over the hyperparameters for the Matérn-$\frac{1}{2}$ kernel. We attain time and space complexities of $\mathcal{O}\left(N_{\text{train}} k_{\max}^{D}\right)$ for each loss evaluation [24].

5 Experimental Results

We demonstrate the model's performance across 20 linear inhomogeneous BVPs, first introduced by [10]. Each example consists of a randomly-generated forcing $u(\boldsymbol{x})$ and its corresponding response $v(\boldsymbol{x})$. We choose to present the relative L^2 error for deterministic comparisons, and the full negative marginal log-likelihood (MLL) as a probabilistic loss. The MLL quantifies the likelihood of the entire test set, given the joint posterior distribution over all real-space points in all test examples.

Accuracy and Uncertainty. A subset of median relative L^2 error and negative MLL is provided in Table 1, alongside corresponding results from GreenLearning [10]. The functions are sampled at 100 equally-spaced locations across the domain, and training examples of v contain a noise content of 1%. Each model is fit 5 times with a randomly-shuffled training set of 75 examples, and tested on 25 unseen examples. Our method is competitive with GreenLearning across the whole suite of equations, obtaining a median relative L^2 error of $< 1\%$ on 17 out of 20.

Sample Efficiency. We fit both models with a range of training set sizes to measure sample efficiency in each case, and in the case of Helmholtz (Figure 2a), GPGreen shows superior performance for very small ($N_{\text{train}} < 13$) training sets, and also obtains a lower error for the maximum size measured. The negative MLL decreases steadily with increasing N_{train}, indicating the model not only gets more accurate pointwise, but its confidence intervals scale accordingly to maintain a good estimate of uncertainty.

Table 1: Achieved accuracy and uncertainty quantification for a wide range of linear operator learning problems. We provide median and inter-quartile range across repeated experiments. Lower is better in each case, and the (often tied) best algorithm for each metric is in **bold**. GreenLearning does not provide uncertainty estimates so GPGreen is stronger on this metric by default.

	GreenLearning [10]		GPGreen (ours)	
Equation	Rel L^2 error / %	$-$MLL	Rel L^2 error / %	$-$MLL
`advection_diffusion`	**0.611** $\pm$ **0.021**	—	0.756 $\pm$ 0.012	-10241 $\pm$ 4
`advection_diffusion_jump`	**0.588** $\pm$ **0.022**	—	**0.608** $\pm$ **0.011**	-11592 $\pm$ 11
`airy_equation`	**0.735** $\pm$ **0.046**	—	**0.739** $\pm$ **0.003**	-10487 $\pm$ 11
`biharmonic`	0.609 $\pm$ 0.010	—	**0.279** $\pm$ **0.021**	-11714 $\pm$ 23
`boundary_layer`	**0.588** $\pm$ **0.016**	—	0.880 $\pm$ 0.005	-10345 $\pm$ 19
`cubic_helmholtz`	**1.956** $\pm$ **0.026**	—	3.078 $\pm$ 0.108	-5503 $\pm$ 210
`cusp`	0.565 $\pm$ 0.017	—	**0.437** $\pm$ **0.004**	-10110 $\pm$ 26
`dawson`	**0.624** $\pm$ **0.014**	—	0.917 $\pm$ 0.001	-9876 $\pm$ 16
`helmholtz`	1.074 $\pm$ 0.028	—	**0.702** $\pm$ **0.008**	-10553 $\pm$ 7

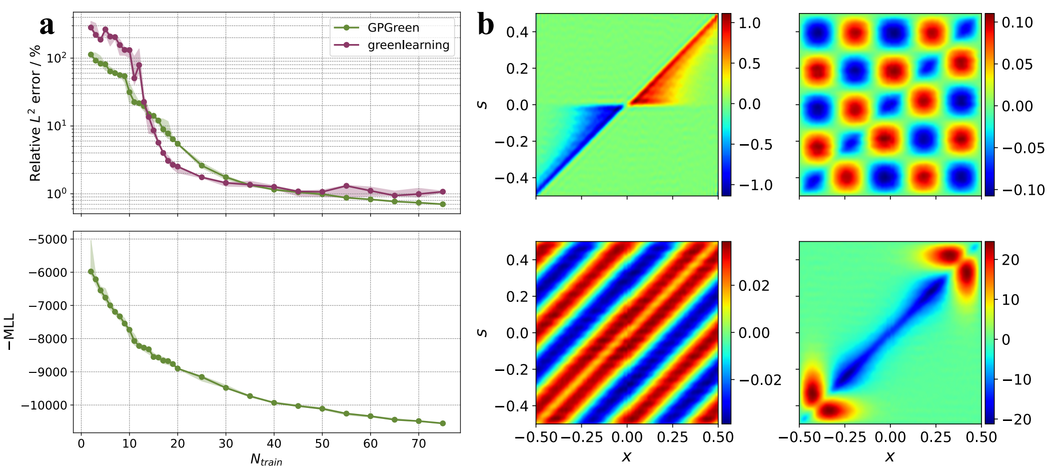

Figure 2: **a**. Median and IQR error metrics on the Helmholtz equation problem, measured for GPGreen and GreenLearning over a range of N_{train}. **b**. Plots of the learned Green's functions for the Dawson (*top left*), Helmholtz (*top right*), periodic Helmholtz (*bottom left*), and potential barrier (*bottom right*) equations.

Interpretability. A significant advantage of our approach over prevalent operator learning frameworks is interpretability. By providing a distribution over the Green's functions learned by our model, we can infer properties of the system, *e.g.* the symmetry of the learnt Helmholtz operator (Figure 2b) indicates that the learned operator is self-adjoint [10].

6 Discussion

Conclusion. We have proposed and demonstrated a Bayesian operator learning framework which allows the specification of prior and posterior distributions over the space of linear operators. By placing a GP prior on the Green's function for an inhomogeneous BVP and conditioning on data, we have made accurate and interpretable predictions of PDE solutions.

Future Work. This work is a first step in scalable and reliable probabilistic operator learning. Key extensions include extending GPGreen to also learn the homogeneous solution of BVPs and to support more expressive kernels. We will explore the many applications enabled by probabilistic operator learning including active learning [14–16] and generating ensembles of solutions for uncertain or fundamentally stochastic differential equations [25]. Finally, future work will use methods for PDE linearisation to apply our framework to broader classes of non-linear PDEs [26–29].

Acknowledgments and Disclosure of Funding

T. Cowperthwaite was supported by the UKRI Centre for Doctoral Training in Application of Artificial Intelligence to the study of Environmental Risks [EP/S022961/1]. This work used JASMIN, the UK's collaborative data analysis environment (https://www.jasmin.ac.uk).

References

[1] Remi Lam, Alvaro Sanchez-Gonzalez, Matthew Willson, Peter Wirnsberger, Meire Fortunato, Ferran Alet, Suman Ravuri, Timo Ewalds, Zach Eaton-Rosen, Weihua Hu, Alexander Merose, Stephan Hoyer, George Holland, Oriol Vinyals, Jacklynn Stott, Alexander Pritzel, Shakir Mohamed, and Peter Battaglia. Learning skillful medium-range global weather forecasting. *Science*, 382:1416–1421, November 2023. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.adi2336. URL https://www.science.org/doi/10.1126/science.adi2336.

[2] Oliver G. Ernst, Björn Sprungk, and Chao Zhang. Uncertainty modeling and propagation for groundwater flow: a comparative study of surrogates. *GEM - International Journal on Geomathematics*, 15(1):11, July 2024. ISSN 1869-2680. doi: 10.1007/s13137-024-00250-3. URL https://doi.org/10.1007/s13137-024-00250-3.

[3] Shuang Qian, Devran Ugurlu, Elliot Fairweather, Laura Dal Toso, Yu Deng, Marina Strocchi, Ludovica Cicci, Richard E. Jones, Hassan Zaidi, Sanjay Prasad, Brian P. Halliday, Daniel Hammersley, Xingchi Liu, Gernot Plank, Edward Vigmond, Reza Razavi, Alistair Young, Pablo Lamata, Martin Bishop, and Steven Niederer. Developing cardiac digital twin populations powered by machine learning provides electrophysiological insights in conduction and repolarization. *Nature Cardiovascular Research*, 4(5):624–636, May 2025. ISSN 2731-0590. doi: 10.1038/s44161-025-00650-0. URL https://www.nature.com/articles/s44161-025-00650-0. Publisher: Nature Publishing Group.

[4] Zongyi Li, Nikola Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Neural Operator: Graph Kernel Network for Partial Differential Equations, March 2020. URL http://arxiv.org/abs/2003.03485. arXiv:2003.03485 [cs].

[5] Zongyi Li, Nikola Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Fourier Neural Operator for Parametric Partial Differential Equations, May 2021. URL http://arxiv.org/abs/2010.08895. arXiv:2010.08895 [cs, math].

[6] Pau Batlle, Matthieu Darcy, Bamdad Hosseini, and Houman Owhadi. Kernel methods are competitive for operator learning. *Journal of Computational Physics*, 496:112549, January 2024. ISSN 0021-9991. doi: 10.1016/j.jcp.2023.112549. URL https://www.sciencedirect.com/science/article/pii/S0021999123006447.

[7] Levi Lingsch, Mike Y. Michelis, Emmanuel de Bezenac, Sirani M. Perera, Robert K. Katzschmann, and Siddhartha Mishra. Beyond Regular Grids: Fourier-Based Neural Operators on Arbitrary Domains, May 2024. URL http://arxiv.org/abs/2305.19663. arXiv:2305.19663 [cs] version: 4.

[8] Zongyi Li, Hongkai Zheng, Nikola Kovachki, David Jin, Haoxuan Chen, Burigede Liu, Kamyar Azizzadenesheli, and Anima Anandkumar. Physics-Informed Neural Operator for Learning Partial Differential Equations. *ACM / IMS J. Data Sci.*, 1(3):9:1–9:27, May 2024. doi: 10.1145/3648506. URL https://dl.acm.org/doi/10.1145/3648506.

[9] Nicolas Boullé and Alex Townsend. Chapter 3 - A mathematical guide to operator learning. In Siddhartha Mishra and Alex Townsend, editors, *Handbook of Numerical Analysis*, volume 25 of *Numerical Analysis Meets Machine Learning*, pages 83–125. Elsevier, January 2024. doi: 10.1016/bs.hna.2024.05.003. URL https://www.sciencedirect.com/science/article/pii/S1570865924000036.

[10] Nicolas Boullé, Christopher J. Earls, and Alex Townsend. Data-driven discovery of Green's functions with human-understandable deep learning. *Scientific Reports*, 12(1):4824, March 2022. ISSN 2045-2322. doi: 10.1038/s41598-022-08745-5. URL https://www.nature.com/articles/s41598-022-08745-5. Number: 1 Publisher: Nature Publishing Group.

[11] Thorsten Kurth, Shashank Subramanian, Peter Harrington, Jaideep Pathak, Morteza Mardani, David Hall, Andrea Miele, Karthik Kashinath, and Anima Anandkumar. FourCastNet: Accelerating Global High-Resolution Weather Forecasting Using Adaptive Fourier Neural Operators. In *Proceedings of the Platform for Advanced Scientific Computing Conference*, PASC '23, pages 1–11, New York, NY, USA, June 2023. Association for Computing Machinery. ISBN 979-8-4007-0190-0. doi: 10.1145/3592979.3593412. URL https://dl.acm.org/doi/10.1145/3592979.3593412.

[12] Boris Bonev, Thorsten Kurth, Ankur Mahesh, Mauro Bisson, Jean Kossaifi, Karthik Kashinath, Anima Anandkumar, William D. Collins, Michael S. Pritchard, and Alexander Keller. FourCastNet 3: A geometric approach to probabilistic machine-learning weather forecasting at scale, July 2025. URL http://arxiv.org/abs/2507.12144. arXiv:2507.12144 [cs].

[13] Shawn G Rosofsky and E A Huerta. Magnetohydrodynamics with physics informed neural operators. *Machine Learning: Science and Technology*, 4(3):035002, July 2023. ISSN 2632-2153. doi: 10.1088/2632-2153/ace30a. URL https://doi.org/10.1088/2632-2153/ace30a. Publisher: IOP Publishing.

[14] Burr Settles. Active Learning Literature Survey. Technical Report, University of Wisconsin-Madison Department of Computer Sciences, 2009. URL https://minds.wisconsin.edu/handle/1793/60660. Accepted: 2012-03-15T17:23:56Z.

[15] Kevin Maik Jablonka, Giriprasad Melpatti Jothiappan, Shefang Wang, Berend Smit, and Brian Yoo. Bias free multiobjective active learning for materials design and discovery. *Nature Communications*, 12(1):2312, April 2021. ISSN 2041-1723. doi: 10.1038/s41467-021-22437-0. URL https://www.nature.com/articles/s41467-021-22437-0. Publisher: Nature Publishing Group.

[16] Parmida Atighehchian, Frédéric Branchaud-Charron, and Alexandre Lacoste. Bayesian active learning for production, a systematic study and a reusable library, June 2020. URL http://arxiv.org/abs/2006.09916. arXiv:2006.09916 [cs].

[17] 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 https://www.nature.com/articles/s42256-021-00302-5. Publisher: Nature Publishing Group.

[18] J. S. Hesthaven and S. Ubbiali. Non-intrusive reduced order modeling of nonlinear problems using neural networks. *Journal of Computational Physics*, 363:55–78, June 2018. ISSN 0021-9991. doi: 10.1016/j.jcp.2018.02.037. URL https://ui.adsabs.harvard.edu/abs/2018JCoPh.363...55H. ADS Bibcode: 2018JCoPh.363...55H.

[19] 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. ISSN 2426-8399. doi: 10.5802/smai-jcm.74. URL https://smai-jcm.centre-mersenne.org/articles/10.5802/smai-jcm.74/.

[20] 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. ISSN 1533-7928. URL http://jmlr.org/papers/v24/21-1524.html.

[21] Craig R. Gin, Daniel E. Shea, Steven L. Brunton, and J. Nathan Kutz. DeepGreen: deep learning of Green's functions for nonlinear boundary value problems. *Scientific Reports*, 11(1):21614, November 2021. ISSN 2045-2322. doi: 10.1038/s41598-021-00773-x. URL https://www.nature.com/articles/s41598-021-00773-x. Publisher: Nature Publishing Group.

[22] Roman Garnett. Lecture 11: Conditioning on Outputs of Linear Operators, 2017. URL https://www.cse.wustl.edu/~garnett/cse515t/spring_2015/files/lecture_notes/11.pdf.

[23] Carl Edward Rasmussen and Christopher K. I. Williams. *Gaussian processes for machine learning*. Adaptive computation and machine learning. MIT Press, Cambridge, Mass, 2006. ISBN 978-0-262-18253-9. OCLC: ocm61285753.

[24] Yunus Saatçi. *Scalable Inference for Structured Gaussian Process Models*. PhD, University of Cambridge, Cambridge, December 2011. URL https://mlg.eng.cam.ac.uk/pub/pdf/Saa11.pdf.

[25] Pavel Perezhogin, Laure Zanna, and Carlos Fernandez-Granda. Generative Data-Driven Approaches for Stochastic Subgrid Parameterizations in an Idealized Ocean Model. *Journal of Advances in Modeling Earth Systems*, 15(10):e2023MS003681, 2023. ISSN 1942-2466. doi: 10.1029/2023MS003681. URL https://onlinelibrary.wiley.com/doi/abs/10.1029/2023MS003681.

[26] R. H. Flake. Volterra series representation of nonlinear systems. *Transactions of the American Institute of Electrical Engineers, Part II: Applications and Industry*, 81(6):330–335, January 1963. ISSN 2379-6774. doi: 10.1109/TAI.1963.6371765. URL https://ieeexplore.ieee.org/document/6371765.

[27] Gabriele Immordino, Andrea Da Ronch, and Marcello Righi. Parametric Nonlinear Volterra Series via Machine Learning: Transonic Aerodynamics, October 2024. URL http://arxiv.org/abs/2410.19514. arXiv:2410.19514 [cs] version: 1.

[28] B. O. Koopman. Hamiltonian Systems and Transformation in Hilbert Space. *Proceedings of the National Academy of Sciences*, 17(5):315–318, May 1931. doi: 10.1073/pnas.17.5.315. URL https://www.pnas.org/doi/abs/10.1073/pnas.17.5.315. Publisher: Proceedings of the National Academy of Sciences.

[29] Igor Mezić. Spectral Properties of Dynamical Systems, Model Reduction and Decompositions. *Nonlinear Dynamics*, 41(1):309–325, August 2005. ISSN 1573-269X. doi: 10.1007/s11071-005-2824-x. URL https://doi.org/10.1007/s11071-005-2824-x.