Roussel Desmond Nzoyem

Research Statement

rd.nzoyemngueguin@bristol.ac.uk | +44 7878430616 | October 24, 2025

*"First you **U**nderstand, second you **A**pply, and only then you **I**mprove."*
— Anonymous Author

1. Overview

My long-term research **goal** is to establish a unified framework for rapid and robust model adaptation in Scientific Machine Learning (SciML). The central challenge in deploying machine learning models lies in their ability to generalize and adapt to new, unseen scenarios—a problem known as out-of-distribution (OoD) adaptation. I believe the key to unlocking this capability lies not just in scaling models, but in fundamentally understanding how models acquire and refine new skills.

My **vision** is to develop models that can rapidly acquire new skills on-the-fly, paving the way for truly autonomous general intelligence. I will realize this vision through three interconnected meta-learning paradigms: adaptive conditioning [11], where a model's behaviour is modulated by an explicit context; test-time training [31], where a model refines its parameters during inference; and in-context learning (ICL) [32], where a model learns a new task from prompt examples without explicit gradient updates. While each has its merits, I argue that the future of adaptable SciML lies in a powerful synthesis, i.e., a combination of in-context learning with the explicit refinement of a model's capabilities through *weight-space learning* [25].

My research **philosophy** follows Pasteur's Quadrant [27] as I strive to conduct "use-inspired research". That is, my work seeks a fundamental understanding of nature, and at the same time is motivated by the need to solve immediate problems in both academic and industry settings.

2. Summary of Past Research

My research has laid the groundwork for this vision by powering SciML research with custom differentiable programming tools [13]. I developed novel context-aware meta-learning frameworks for OoD adaptation in physical systems and time series. This body of work, which progressed from conventional meta-learning to weight-space learning, is built upon a strong foundation in scientific computing and high-performance computing (HPC).

2.1 From Meta-Learning to Weight-Space Adaptation

Initially, I focused on methods for efficient, gradient-based adaptation. My work on **N**eural **C**ontext **F**lows (NCF) [15] introduced a meta-learning framework that learns to modulate a shared Neural ODE predictor, f_θ. By conditioning f_θ on a low-dimensional context vector ξ^e for each task (or environment) e, information is shared within the meta-training set. Specifically, we approximate the vector field for a target context ξ^e using information from a source context ξ^j via a Taylor expansion:

$$f_\theta(x, \xi^e) = f_\theta(x, \xi^j) + \nabla_\xi f_\theta(x, \xi^j)(\xi^e - \xi^j) + o(\|\xi^e - \xi^j\|_2^2).$$

This formulation, termed Contextual Self-Modulation, allows for efficient one-shot adaptation and provides an interpretable, uncertainty-aware framework for generalizing across physical systems.

NCFs were built on cutting-edge differentiable programming techniques [14, 13], and subsequently extended to cover a wide variety of data modalities and tasks, including infinite-dimensional settings

[16]. NCFs have also been leveraged as the backbone for mixture of experts models, in an attempt to build the first foundational model for physical sciences [21].

Building on this, I sought to overcome the limitations of test-time training and gradient-based fine-tuning. This led to the development of **W**eight-space **A**daptive **R**ecurrent **P**rediction (WARP) [18], a model that unifies weight-space learning with linear recurrence. WARP reinterprets the hidden state of an RNN as the full set of weights $\theta_t \in \mathbb{R}^{D_\theta}$ of an auxiliary neural network, MLP_{θ_t}. A linear recurrence updates these weights at each time step based on input differences $\Delta x_t \in \mathbb{R}^{D_x}$:

$$\theta_t = A\theta_{t-1} + B\Delta x_t, \quad y_t = \text{MLP}_{\theta_t}(\tau),$$

where $A \in \mathbb{R}^{D_\theta \times D_\theta}$ and $B \in \mathbb{R}^{D_\theta \times D_x}$ are learned transition matrices, and τ is a coordinate input. This represents a significant leap beyond traditional meta-learning, as adaptation becomes an intrinsic part of the model's forward pass, eliminating the need for explicit gradient descent at test time. WARP demonstrated remarkable in-context learning abilities, proving that weight-space learning is a powerful paradigm for creating self-adapting models.

2.2 Applying Foundational Principles in Scientific Computing and HPC

Some domains that have inspired and largely benefited from my research in applied mathematics and computational sciences are listed below:

- **Inverse Problems in Radiative Transfer [17]:** I developed a framework to solve a medical imaging inverse problem. This involved first building a C++ solver for the 2D radiative transfer Partial Differential Equation (PDE) to generate simulation data, and then training a U-Net [23] to reconstruct the medium's density from boundary signals.

- **Fracture Mechanics of Sea Ice [19]:** I modeled the percussion and fracture of sea ice floes using a granular, mass-spring-damper system. This work required applying Griffith's theory of fracture within a variational framework to predict crack nucleation and propagation.

- **Accelerating Numerical Solvers with GNNs [20]:** I tackled a core challenge in HPC by accelerating the Algebraic Multigrid (AMG) method, a state-of-the-art solver for large sparse linear systems. I designed a Graph Neural Network (GNN) to learn the computationally expensive prolongation operator in AMG, via an unsupervised learning objective. The large-scale model was deployed on the AWS cloud-based HPC platform.

- **Proxy Modeling for Reservoir Simulation:** I tackled the development of a GNN-based proxy model for carbon capture and storage simulations. This work involved designing and implementing novel GNN layers in JAX/Jraph, achieving a 2x speedup over standard PyTorch implementations [24]. A key outcome was demonstrating zero-shot super-resolution, enabling the model to generalize from small to large-scale reservoir graphs.

3. Research Agenda: Accelerated In-Context Skill Acquisition

My future research program is centred on a single, critical question: **How can we accelerate in-context learning skill acquisition?** While large models demonstrate impressive in-context learning (ICL) capabilities, their efficiency in the long-context, data-scarce, high-precision domain of SciML remains an open question [4]. At the same time, ICL is increasingly analyzed through *simplified* neural network training dynamics with great success [32, 30], and little is done to apply these findings to complex models ready for real-world deployment. ICL acceleration would mean achieving greater accuracy with shorter prompts, in real-world resource-constrained scenarios. My agenda addresses these problems by developing the foundational and algorithmic innovations necessary to make ICL a cornerstone of next-generation applied scientific modeling.

3.1 Foundational Understanding of In-Context Learning

Two major problems currently limit the application of ICL. First, findings that demonstrate that ICL abilities naturally emerge in most popular large-scale architectures remain *empirical* [12]. Second, recent work has shown that ICL abilities can be *fragile*, vanishing under covariate shifts [30]. My first research thrust aims to address these fundamental questions by theoretically investigating how ICL emerges in deep sequence models such as Transformers [29], recurrent neural networks [22], state-space models [7], and weight-space models [18].

Initial analyses with weight-space models suggest that simple operations like *cumulative sums* can enable powerful ICL adaptation in function space [18]. My goal is to develop the theoretical under-pinnings of this phenomenon by analyzing, for instance, the trajectory of the learned WARP weights $\{\theta_t\}_{t=1}^{T}$. This involves studying the eigenspectrum of the transition matrix A to characterize the stability and convergence of the weight-space dynamics, thereby providing a theoretical foundation for designing more efficient and robust ICL-driven models.

3.2 Hybridizing In-Context Learning and Differentiable Simulation

Differentiable Simulation[1] (DS) presents a powerful method for enhancing ICL [5]. By integrating DS, we can explicitly encode physical laws and domain knowledge as strong inductive biases directly into the learning framework. This approach yields significant benefits, making models more sample-efficient to adapt and substantially improving their accuracy and robustness, especially in OoD scenarios where purely data-driven models often falter [1]. My prior work has contributed to this area by comparing these DS methods [14] and utilizing cutting-edge developments in Taylor-mode Automatic Differentiation to significantly improve how models of physical systems generalize [16]. I now aim to apply these advanced DS techniques to accelerate ICL.

3.3 Applications in Scientific Machine Learning

My theoretical and algorithmic work on accelerated ICL will be grounded in two high-impact application areas central to the future of scientific discovery.

Foundation Models for Dynamical Systems. The vision of "foundation models" for dynamical systems is compelling, but adapting a massive model to a specific PDE—for instance with an unseen physical parameter or a new differential operator—is a major hurdle [28, 10]. To date, fine-tuning remains computationally expensive and risks catastrophic forgetting [21], which has sparked intense research into true zero-shot adaptation [26, 9, 33, 18]. I propose that *accelerated* ICL, through differentiable simulation [14], is the key to unlocking the potential of these foundation models. A pre-trained model with rapid ICL could be prompted with a few data points from a new system (e.g., boundary conditions, sensor measurements) and instantly adapt its internal solver.

Generative Video Models as World Simulators. The promise of foundation models for dynamical systems extends to videos, which capture the real world as governed by interacting physical systems. A generative video model that predicts future frames is, in essence, a predictive world model [2, 8]. A critical component of such models is the *compression* network, which encodes video patches into succinct latent representations. Note that compression is the lifeblood of meta-learning and amortized inference—which sadly have remained limited to single image (or frame) encoding [6]. Given such a critical research gap, I wish to understand the fundamental information-theoretic mechanisms that not only underpin (meta-)learning as a way to compress sequences, but also compression as a way to learn [3].

---

[1]Differentiable Simulation, sometimes referred to as Differentiable Programming, is the practice of building computational models in a way that derivatives of outputs with respect to inputs and parameters can be computed efficiently, often via automatic differentiation.

References

[1] Peter W Battaglia, Jessica B Hamrick, Victor Bapst, Alvaro Sanchez-Gonzalez, Vinicius Zambaldi, Mateusz Malinowski, Andrea Tacchetti, David Raposo, Adam Santoro, Ryan Faulkner, et al. Relational inductive biases, deep learning, and graph networks. *arXiv*, 2018.

[2] Tim Brooks, Bill Peebles, Connor Holmes, Will DePue, Yufei Guo, Li Jing, David Schnurr, Joe Taylor, Troy Luhman, Eric Luhman, Clarence Ng, Ricky Wang, and Aditya Ramesh. Video generation models as world simulators. 2024.

[3] Sam Buchanan, Druv Pai, Peng Wang, and Yi Ma. *Learning Deep Representations of Data Distributions*. Online, August 2025. `https://ma-lab-berkeley.github.io/deep-representation-learning-book/`.

[4] Salvatore Cuomo, Vincenzo Schiano Di Cola, Fabio Giampaolo, Gianluigi Rozza, Maziar Raissi, and Francesco Piccialli. Scientific machine learning through physics–informed neural networks: Where we are and what's next. *Journal of Scientific Computing*, 92(3):88, 2022.

[5] Nima Hosseini Dashtbayaz, Hesam Salehipour, Adrian Butscher, and Nigel Morris. Physics-informed reduced order modeling of time-dependent pdes via differentiable solvers. *arXiv*, 2025.

[6] Jonathan Gordon, Wessel P. Bruinsma, Andrew Y. K. Foong, James Requeima, Yann Dubois, and Richard E. Turner. Convolutional conditional neural processes. In *International Conference on Learning Representations*, 2020.

[7] Albert Gu, Karan Goel, and Christopher Ré. Efficiently modeling long sequences with structured state spaces. *arXiv*, 2021.

[8] Yuchao Gu, Weijia Mao, and Mike Zheng Shou. Long-context autoregressive video modeling with next-frame prediction. *arXiv*, 2025.

[9] Christoph Jürgen Hemmer and Daniel Durstewitz. True zero-shot inference of dynamical systems preserving long-term statistics. *arXiv*, 2025.

[10] Maximilian Herde, Bogdsn Raonić, Tobias Rohner, Roger Käppeli, Roberto Molinaro, Emmanuel de Bézenac, and Siddhartha Mishra. Poseidon: Efsicient foundation models for pdes. *arXiv*, 2024.

[11] Armand Kassaï Koupaï, Jorge Mifsut Benet, Jean-Noël Vittaut, and Patrick Gallinari. Geps: Boosting generalization in parametric pde neural solvers through adaptive conditioning. *NeurIPS*, 2024.

[12] Ivan Lee, Nan Jiang, and Taylor Berg-Kirkpatrick. Is attention required for icl? exploring the relationship between model architecture and in-context learning ability. *arXiv*, 2023.

[13] Roussel Desmond Nzoyem. Updes. Third Updes pre-release version.

[14] Roussel Desmond Nzoyem, David AW Barton, and Tom Deakin. A comparison of mesh-free differentiable programming and data-driven strategies for optimal control under pde constraints. In *SC'23 Workshops*, pages 21–28, 2023.

[15] Roussel Desmond Nzoyem, David A.W. Barton, and Tom Deakin. Neural context flows for meta-learning of dynamical systems. In *ICLR*, 2025.

[16] Roussel Desmond Nzoyem, David AW Barton, and Tom Deakin. Reevaluating meta-learning optimization algorithms through contextual self-modulation. In *CoLLAs*, 2025.

[17] Roussel Desmond Nzoyem, Emmanuel Franck, Laurent Navoret, Vincent Vigon, and Christophe PRUD'HOMME. Simulation 2d de l'équation du transfert radiatif et reconstruction de la densité par un réseau de neurones. *University of Strasbourg*, 2020.

[18] Roussel Desmond Nzoyem, Nawid Keshtmand, Idriss Tsayem, David AW Barton, and Tom Deakin. Weight-space linear recurrent neural networks. *arXiv*, 2025.

[19] Roussel Desmond Nzoyem and Stephane Labbe. Fracturation de floes de glace par percussion dans un modèle granulaire. *GitHub*, 2021.

[20] Roussel Desmond Nzoyem, Thorben Louw, Simon McIntosh-Smith, and Tom Deakin. Accelerating algebraic multigrid using machine learning. *University of Bristol*, 2022.

[21] Roussel Desmond Nzoyem, Grant Stevens, Amarpal Sahota, David A.W. Barton, and Tom Deakin. Towards foundational models for dynamical system reconstruction: Hierarchical meta-learning via mixture of experts. In *SCOPE @ ICLR*, 2025.

[22] Antonio Orvieto, Samuel L Smith, Albert Gu, Anushan Fernando, Caglar Gulcehre, Razvan Pascanu, and Soham De. Resurrecting recurrent neural networks for long sequences. In *ICML*, pages 26670–26698. PMLR, 2023.

[23] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.

[24] Surya Teja Sathujoda and Soham Sheth. Coupled graph neural network and fourier neural operator architecture for ensemble workflows in 3d reservoir simulation. *SPE Res. Sim. Conf.*, 2025.

[25] Konstantin Schürholt, Michael W Mahoney, and Damian Borth. Towards scalable and versatile weight space learning. *arXiv preprint arXiv:2406.09997*, 2024.

[26] Louis Serrano, Armand Kassaï Koupaï, Thomas X Wang, Pierre Erbacher, and Patrick Gallinari. Zebra: In-context and generative pretraining for solving parametric pdes. *arXiv*, 2024.

[27] Donald E Stokes. *Pasteur's quadrant: Basic science and technological innovation*. Brookings Institution Press, 2011.

[28] Shashank Subramanian, Peter Harrington, Kurt Keutzer, Wahid Bhimji, Dmitriy Morozov, Michael W Mahoney, and Amir Gholami. Towards foundation models for scientific machine learning: Characterizing scaling and transfer behavior. *NeurIPS*, 36, 2024.

[29] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *NeurIPS*, 30, 2017.

[30] Ruiqui Zhang, Yuan Yao, Mo Li, Yaodong Yu, H. Gui, Joseph E. Gonzalez, Kannan Ramchandran, Peter L. Bartlett, and Quanquan Gu. Why do in-context learning abilities vanish under covariate shift? *arXiv*, 2023.

[31] Tianyuan Zhang, Sai Bi, Yicong Hong, Kai Zhang, Fujun Luan, Songlin Yang, Kalyan Sunkavalli, William T Freeman, and Hao Tan. Test-time training done right. *arXiv*, 2025.

[32] Yedi Zhang, Aaditya K Singh, Peter E Latham, and Andrew Saxe. Training dynamics of in-context learning in linear attention. *ICML*, 2025.

[33] Yuanzhao Zhang and William Gilpin. Zero-shot forecasting of chaotic systems. *arXiv*, 2024.