Roussel Desmond Nzoyem

Research Statement

roussel.desmond.nzoyem@outlook.com | +1 368 380 2808 | November 28, 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 [7], where a model's behaviour is modulated by an explicit context; test-time training [22], where a model refines its parameters during inference; and in-context learning (ICL) [24], 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* [16].

My research **philosophy** follows Pasteur's Quadrant [18] 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

Initially, I focused on methods for efficient, gradient-based adaptation. My work on **N**eural **C**ontext **F**lows (NCF) [12] 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 [10, 11], and subsequently extended to cover a wide variety of data modalities and tasks, including infinite-dimensional settings [13]. 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 [15].

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) [14], 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.

3. Research Agenda

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 [2]. At the same time, ICL is increasingly analyzed through *simplified* neural network training dynamics with great success [21, 24], 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* [8]. Second, recent work has shown that ICL abilities can be *fragile*, vanishing under covariate shifts [21]. My first research thrust aims to address these fundamental questions by theoretically investigating how ICL emerges in deep sequence models such as Transformers [20], recurrent neural networks [9], state-space models [4], and weight-space models [14].

Initial analyses with weight-space models suggest that simple operations like *cumulative sums* can enable powerful ICL adaptation in function space [14]. My goal is to develop the theoretical underpinnings 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 [3]. 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 [11] and utilizing cutting-edge developments in Taylor-mode Automatic Differentiation to significantly improve how models of physical systems generalize [13]. I now aim to apply these advanced DS techniques to accelerate ICL.

[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.

3.3 Applications for Foundation Models in Scientific Machine Learning

The promise 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 [6, 19]. To date, fine-tuning remains computationally expensive and risks catastrophic forgetting [15], which has sparked intense research into true zero-shot adaptation [5, 14, 17, 23]. I propose that *accelerated* ICL, through differentiable simulation [11], 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.

References

[1] P. W. Battaglia, J. B. Hamrick, V. Bapst, A. Sanchez-Gonzalez, V. Zambaldi, M. Malinowski, A. Tacchetti, D. Raposo, A. Santoro, R. Faulkner, et al. Relational inductive biases, deep learning, and graph networks. *arXiv*, 2018.

[2] S. Cuomo, V. S. Di Cola, F. Giampaolo, G. Rozza, M. Raissi, and F. Piccialli. Scientific machine learning through physics–informed neural networks: Where we are and what's next. *Journal of Scientific Computing*, 92(3):88, 2022.

[3] N. H. Dashtbayaz, H. Salehipour, A. Butscher, and N. Morris. Physics-informed reduced order modeling of time-dependent pdes via differentiable solvers. *arXiv*, 2025.

[4] A. Gu, K. Goel, and C. Ré. Efficiently modeling long sequences with structured state spaces. *arXiv*, 2021.

[5] C. J. Hemmer and D. Durstewitz. True zero-shot inference of dynamical systems preserving long-term statistics. *arXiv*, 2025.

[6] M. Herde, B. Raonić, T. Rohner, R. Käppeli, R. Molinaro, E. de Bézenac, and S. Mishra. Poseidon: Efsicient foundation models for pdes. *arXiv*, 2024.

[7] A. Kassaï Koupaï, J. Mifsut Benet, J.-N. Vittaut, and P. Gallinari. Geps: Boosting generalization in parametric pde neural solvers through adaptive conditioning. *NeurIPS*, 2024.

[8] I. Lee, N. Jiang, and T. Berg-Kirkpatrick. Is attention required for icl? exploring the relationship between model architecture and in-context learning ability. *arXiv*, 2023.

[9] A. Orvieto, S. L. Smith, A. Gu, A. Fernando, C. Gulcehre, R. Pascanu, and S. De. Resurrecting recurrent neural networks for long sequences. In *ICML*, pages 26670–26698. PMLR, 2023.

[10] **R. D. Nzoyem**. Updes, 2023. URL `https://github.com/ddrous/Updes/`. Third Updes pre-release version.

[11] **R. D. Nzoyem**, D. A. Barton, and T. 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.

[12] **R. D. Nzoyem**, D. A. Barton, and T. Deakin. Neural context flows for meta-learning of dynamical systems. In *ICLR*, 2025. URL `https://openreview.net/forum?id=8vzMLo8LDN`.

[13] **R. D. Nzoyem**, D. A. Barton, and T. Deakin. Reevaluating meta-learning optimization algorithms through contextual self-modulation. In *CoLLAs*, 2025.

[14] **R. D. Nzoyem**, N. Keshtmand, I. Tsayem, D. A. Barton, and T. Deakin. Weight-space linear recurrent neural networks. *arXiv*, 2025.

[15] **R. D. Nzoyem**, G. Stevens, A. Sahota, D. A. Barton, and T. Deakin. Towards foundational models for dynamical system reconstruction: Hierarchical meta-learning via mixture of experts. In *SCOPE @ ICLR*, 2025. URL `https://openreview.net/forum?id=NteuHm0UXw`.

[16] K. Schürholt, M. W. Mahoney, and D. Borth. Towards scalable and versatile weight space learning. *arXiv preprint arXiv:2406.09997*, 2024.

[17] L. Serrano, A. K. Koupaï, T. X. Wang, P. Erbacher, and P. Gallinari. Zebra: In-context and generative pretraining for solving parametric pdes. *arXiv*, 2024.

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

[19] S. Subramanian, P. Harrington, K. Keutzer, W. Bhimji, D. Morozov, M. W. Mahoney, and A. Gholami. Towards foundation models for scientific machine learning: Characterizing scaling and transfer behavior. *NeurIPS*, 36, 2024.

[20] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin. Attention is all you need. *NeurIPS*, 30, 2017.

[21] R. Zhang, Y. Yao, M. Li, Y. Yu, H. Gui, J. E. Gonzalez, K. Ramchandran, P. L. Bartlett, and Q. Gu. Why do in-context learning abilities vanish under covariate shift? *arXiv*, 2023.

[22] T. Zhang, S. Bi, Y. Hong, K. Zhang, F. Luan, S. Yang, K. Sunkavalli, W. T. Freeman, and H. Tan. Test-time training done right. *arXiv*, 2025.

[23] Y. Zhang and W. Gilpin. Zero-shot forecasting of chaotic systems. *arXiv*, 2024.

[24] Y. Zhang, A. K. Singh, P. E. Latham, and A. Saxe. Training dynamics of in-context learning in linear attention. *ICML*, 2025.