DATA EFFICIENCY AND LONG-TERM PREDICTION CAPABILITIES FOR NEURAL OPERATOR SURROGATE MODELS OF EDGE PLASMA CODES

[1]N. CAREY, L. [1]ZANISI, [1]S. PAMELA, [1,4]V. GOPAKUMAR, [1]J. OMOTANI, [1]J. BUCHANAN, [2,3]J. BRANDSTETTER, [2]F. PAISCHER, [2]GIANLUCA GALLETTI, [2]PAUL SETINEK

[1]UKAEA, Culham Centre for Fusion Energy, Abingdon, UK
[2]ELLIS Unit, Linz, LIT AI Lab, Institute for Machine Learning, Johannes Kepler University, Linz, Austria
[3]NXAI GmbH, Austria
[4]Centre for Artificial Intelligence, UCL, London, UK

Email: Naomi.Carey@ukaea.uk

Modelling of plasma dynamics is fundamental to ensure appropriate diverter and core performance, and is desirable for both interpreting the current generation of experiments and informing the next generation devices like ITER [1, 2]. Yet the computational expense of many plasma simulations makes them unsuitable for real-time applications or iterative design workflows. Neural operator surrogate models of JOREK [3] and STORM [4] are evaluated, investigating their capability to replicate plasma dynamics accurately whilst reducing computational cost. It is found that the accuracy of the surrogate models will degrade for long term predictions, and that physics considerations are important in assessing the performance of the surrogates. Surrogates trained on one dataset can be effectively fine tuned with only a few simulations from a target domain. This is particularly effective where the source domain is a low fidelity physics model and the target domain is a high fidelity model, with an order of magnitude improvement in performance for a small dataset and a short rollout.

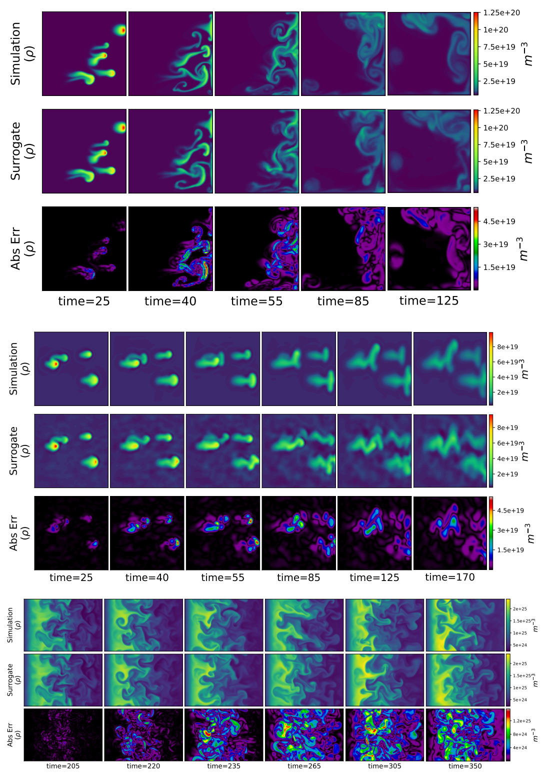

Figure 1: An example trajectory for the density field. (TOP) electrostatic JOREK, (MIDDLE) reduced-MHD JOREK, (BOTTOM) STORM

Neural operators: [5, 6] represent a powerful class of models for learning mappings between function spaces, making them particularly well-suited for approximating the solutions of partial differential equations (PDEs). Unlike traditional neural networks, which are limited to mappings between finite-dimensional spaces, neural operators generalize across varying input discretisation, capturing the continuous functions that govern the data. This capacity allows neural operators to handle a broad range of inputs and boundary conditions, making them ideal for modelling PDE-driven phenomena. For this work, a type of neural operator called the Fourier Neural Operator (FNO) was utilized due to its demonstrated efficiency on medium-scale PDE problems [7].

Datasets: The STORM dataset focuses on modeling turbulence and transport processes in the scrape off layer, where a vertical band of density source generates fluid turbulence in the radial direction. The JOREK datasets investigate large-scale MHD instabilities from the core and edge plasma regions [8]. Two JOREK datasets of differing fidelity were utilized: reduced-MHD JOREK, as typically implemented in routine studies, and electrostatic JOREK, which sets the magnetic component in the equations to zero. Both datasets simulate filamentary blobs in a simplified 2D rectangular slab geometry with toroidal curvature, modeling their radial outward motion due to toroidal curvature and blob pressure gradients. An example trajectory for all three datasets, along with the surrogate output, is illustrated in Fig. 1.

Towards replacing simulations with NOs. The NOs successfully replicated global plasma dynamics, such as blob and density band location (see fig 1), achieving significant speedups compared to traditional solvers (a full rollout of the surrogate is approximately 2 minutes on a single core compared to running a full simulation which takes 14 hours on 48 cores for the JOREK code as an example). Whilst FNOs captured early plasma behaviors accurately, long time predictions (generated auto regressively by feeding output from prior steps as input) resulted in the finer details being lost and revealed sensitivity to not just accumulated input errors but also trajectory-specific physical phenomena. These findings highlight the need for refined neural surrogate architectures capable of addressing long-term predictive challenges and ensuring stable and accurate simulations for practical fusion applications.

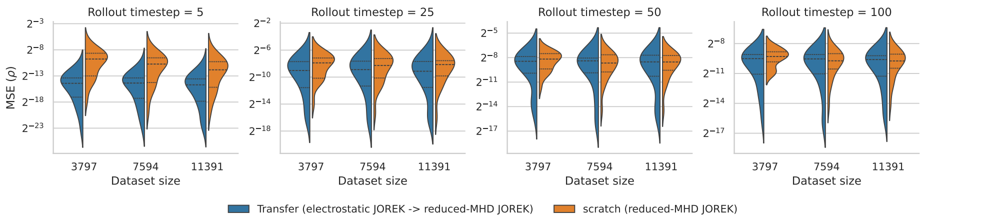

Figure 2: Comparison between model trained (orange) from scratch on the higher fidelity JOREK simulations and (blue) transferred from a lower fidelity JOREK dataset (density field) trained on different dataset sizes. Each column corresponds to a different rollout timestep, with the rollouts getting longer towards the right.

Leveraging low-fidelity simulations to inform high-fidelity surrogate models. Data from high-fidelity plasma simulations is scarce and prohibitively expensive to generate. Transfer learning allows models to leverage knowledge from a source domain to improve learning efficiency and performance in a target domain. Following the success of transfer learning in computer vision and natural language processing tasks [9, 10, 11, 12, 13], prior work has demonstrated its potential to improve data efficiency for neural operator surrogates in simulation-based tasks [14].

Transferring knowledge from low- to higher-fidelity JOREK datasets (electrostatic JOREK to reduced-MHD JOREK) achieved error reductions of up to an order of magnitude in short rollouts and by a factor of 2 in long rollouts at fixed dataset size (see fig 2). This demonstrates the potential of leveraging low-cost datasets with transfer learning to improve performance. However, this improvement diminished in long rollouts and occured much faster when the model was tranferred between datasets with differing physical characteristics (such as datasets generated from different codes). This emphasises the need for more robust transfer learning strategies to enhance long-term rollout accuracy.

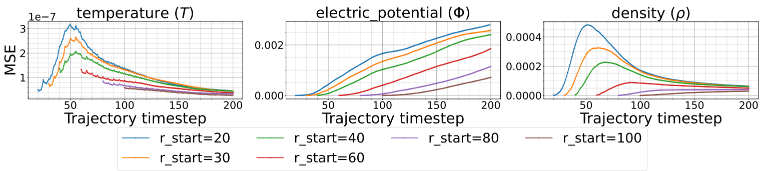

Figure 3: Long rollout error for FNO model trained on low-fidelity JOREK dataset relative to trajectory timestep. Each line is a rollout started at a different point in the trajectory (using timesteps as input prior to that point) with accumulating input error caused by the autoregressive rollout.

Sensitivity to Specific Trajectory Locations: Analysis revealed that long roll-outs showed distinct error patterns depending on the dataset and field. Both models trained on JOREK datasets exhibited error spikes at specific trajectory points regardless of accumulated input error (see figure 3). As each trajectory followed consistent physical behavior (for electrostatic JOREK this being blobs being initialized at random positions, traveling to the right, interacting with the boundary, and eventually dissipating), this indicated that these were driven by localized physical events. For example, temperature and density fields experienced sharp error spikes around $t = 50$ for the electrostatic JOREK model, coinciding with the time period in which the blobs would generally impact the wall. Point-wise error analysis showed errors concentrated near boundaries, suggesting that these spikes are related to boundary effects. However, the model trained on the reduced-MHD JOREK dataset showed earlier error spikes around $t = 30$, despite the blob-wall impact occurring much later in the trajectory. This divergence suggests that boundary effects alone cannot fully explain the observed error spikes, as other factors must be contributing to the dynamics. The exact dynamics behind the observed error spikes remain unclear but these findings highlight the importance of considering both the physical dynamics of the system and the limitations of the model when analyzing long-term predictions.

ACKNOWLEDGEMENTS

This work has been part-funded by the EPSRC Energy Programme [grant number EP/W006839/1]. To obtain further information on the data and models underlying this paper please contact PublicationsManager@ukaea.uk

This work has also been part-funded by the Fusion Futures Programme. As announced by the UK Government in October 2023, Fusion Futures aims to provide holistic support for the development of the fusion sector.

REFERENCES

[1] Alberto Loarte et al. “Chapter 4: Power and particle control”. In: *Nuclear Fusion* 47 (2007), S203–S263.

[2] Thomas Eich et al. “Scaling of the tokamak near the scrape-off layer H-mode power width and implications for ITER”. In: *Nuclear Fusion* 53 (2013).

[3] M. Hoelzl et al. “The JOREK non-linear extended MHD code and applications to large-scale instabilities and their control in magnetically confined fusion plasmas”. In: *Nuclear Fusion* 61.6 (May 2021), p. 065001. ISSN: 1741-4326. DOI: 10.1088/1741-4326/abf99f. URL: http://dx.doi.org/10.1088/1741-4326/abf99f.

[4] N R Walkden et al. “Dynamics of 3D isolated thermal filaments”. In: *Plasma Phys. Control. Fusion* 58.11 (Nov. 2016), p. 115010.

[5] Nikola Kovachki et al. *Neural Operator: Learning Maps Between Function Spaces*. 2023. arXiv: 2108.08481 [cs.LG].

[6] Sifan Wang, Hanwen Wang, and Paris Perdikaris. *Learning the solution operator of parametric partial differential equations with physics-informed DeepOnets*. 2021. arXiv: 2103.10974 [cs.LG]. URL: https://arxiv.org/abs/2103.10974.

[7] Kamyar Azizzadenesheli et al. “Neural operators for accelerating scientific simulations and design”. In: *Nature Reviews Physics* 6.5 (Apr. 2024), pp. 320–328. ISSN: 2522-5820. DOI: 10.1038/s42254-024-00712-5. URL: http://dx.doi.org/10.1038/s42254-024-00712-5.

[8] S F Smith et al. “Simulations of edge localised mode instabilities in MAST-U Super-X tokamak plasmas”. In: *Nucl. Fusion* 60.6 (June 2020), p. 066021.

[9] Neil Houlsby et al. *Parameter-Efficient Transfer Learning for NLP*. 2019. arXiv: 1902.00751 [cs.LG]. URL: https://arxiv.org/abs/1902.00751.

[10] Colin Raffel et al. *Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer*. 2023. arXiv: 1910.10683 [cs.LG]. URL: https://arxiv.org/abs/1910.10683.

[11] Kuniaki Saito et al. *Maximum Classifier Discrepancy for Unsupervised Domain Adaptation*. 2018. arXiv: 1712.02560 [cs.CV]. URL: https://arxiv.org/abs/1712.02560.

[12] Karl R. Weiss, Taghi M. Khoshgoftaar, and Dingding Wang. “A survey of transfer learning”. In: *Journal of Big Data* 3 (2016). URL: https://api.semanticscholar.org/CorpusID:256407097.

[13] Jason Yosinski et al. “How transferable are features in deep neural networks?” In: *Advances in Neural Information Processing Systems*. Ed. by Z. Ghahramani et al. Vol. 27. Curran Associates, Inc., 2014. URL: https://proceedings.neurips.cc/paper_files/paper/2014/file/375c71349b295fbe2dcdca9206f20a06-Paper.pdf.

[14] Shashank Subramanian et al. *Towards Foundation Models for Scientific Machine Learning: Characterizing Scaling and Transfer Behavior*. 2023. arXiv: 2306.00258 [cs.LG]. URL: https://arxiv.org/abs/2306.00258.