

CONFERENCE PRE-PRINT

NEURAL OPERATOR SURROGATE MODELS OF PLASMA EDGE SIMULATIONS: FEASIBILITY AND DATA EFFICIENCY

N. Carey
UKAEA
Abingdon, UK
Email: Naomi.Carey@UKAEA.uk

L. Zanisi
UKAEA
Abingdon, UK

S. Pamela
UKAEA
Abingdon, UK

V. Gopakumar
UKAEA
Abingdon, UK

J. Omotani
UKAEA
Abingdon, UK

J. Buchanan
UKAEA
Abingdon, UK

J. Brandstetter
Johannes Kepler University
Linz, Austria

Fabian Paischer
Johannes Kepler University
Linz, Austria

Gianluca Galletti
Johannes Kepler University
Linz, Austria

Paul Setinek
Johannes Kepler University
Linz, Austria

Abstract

Simulation-based plasma scenario development plays a crucial role in designing next-generation tokamaks and fusion power plants. However, the inclusion of high-fidelity simulations of transient MHD events such as Edge Localized Modes (ELMs) in highly iterative applications remains computationally prohibitive, limiting their use in design and control workflows. Understanding these phenomena is vital, as they govern heat flux on plasma-facing components, influencing reactor performance and material lifetime. The paper explored Fourier Neural Operators (FNOs) as surrogate models to accelerate plasma simulations from the JOREK MHD code.

FNOs were trained on single-step rollouts and evaluated in terms of long-term predictive accuracy in an auto-regressive manner. To mitigate the computational burden of dataset generation, a transfer learning strategy was explored, leveraging low-fidelity simulations to improve performance on high-fidelity datasets. The paper’s results showed that FNOs effectively captured initial plasma evolution, including blob movement for JOREK. However, long rollouts accumulated errors and exhibited sensitivity to certain physical phenomena, leading to non-monotonic error spikes. Transfer learning significantly reduced errors for small dataset sizes and short rollouts, achieving an order-of-magnitude reduction when transferring from low- to high-fidelity datasets. However, its effectiveness diminished with longer rollouts and larger dataset sizes. Attempts to transfer models to previously unseen variables in simulations were unsuccessful, underscoring the limitations of transfer learning in this context.

The paper’s findings demonstrate the promise of neural operators for accelerating fusion-relevant PDE simulations. However, they also highlight key challenges: improving long-term accuracy to mitigate error accumulation, capturing critical physical behaviors, and developing robust surrogates that effectively leverage multi-fidelity, multi-physics datasets.

1. INTRODUCTION

Modeling the plasma behaviour is fundamental to ensuring reliable divertor and core performance in present experiments and in next-generation devices such as ITER and STEP [1, 2]. Edge dynamics, such as transient magnetohydrodynamic (MHD) events like edge localized modes (ELMs), play a central role in determining heat and particle fluxes to plasma-facing components. However, resolving these dynamics requires solving strongly coupled partial differential equations (PDEs), which is computationally intensive even for reduced-order models (e.g. DIV1D [3]). High-fidelity solvers such as JOREK for MHD instabilities remain prohibitively expensive for highly iterative tasks, such as scenario optimization or real-time control. Additionally, numerical solvers often have to be specifically tailored to the equations and spatial geometry they address [4], are challenging to deploy quickly due to their complex software infrastructure and library dependencies, and these solvers may fail to converge when integrated into larger simulation frameworks, often due to model mismatches or the need for expert manual intervention.

Neural surrogate models offer a promising route to accelerate these workflows by providing rapid approximations to PDE solutions. Convolutional neural network (CNN) surrogates have demonstrated orders-of-magnitude speedups for specific codes [5, 6, 7], but they lack discretization invariance, restricting their generality. Neural Operators (NOs) [8], and in particular the Fourier Neural Operator (FNO) [9], overcome this limitation by learning mappings between function spaces rather than between fixed discretizations. FNOs have shown strong performance on a variety of PDE problems [9, 10], including some simplified fusion-relevant cases [11, 12, 13, 14, 15, 16, 7], suggesting their potential for more complex settings. Nevertheless, their application to fusion plasma simulations remains limited, and important open questions remain regarding long-term predictive stability. Additionally, creating simulation data to train surrogate models is often prohibitively expensive, this work explores the potential of transfer learning across codes and fidelity levels to help mitigate the dataset requirements to train.

This work explores FNO-based surrogate modeling for a widely used plasma simulation code: JOREK [17], which captures large-scale MHD instabilities. The study focuses on two core objectives: (1) assessing long-term predictive accuracy of FNOs in autoregressive rollouts, and (2) investigating transfer learning as a strategy to improve data efficiency when high-fidelity simulation data are scarce. Specifically, we examine transfer learning across physics fidelity levels within JOREK, and across variables when the target dataset includes outputs absent from the source dataset.

2. METHODS

2.1. Dataset

We use two training datasets from JOREK [17], previously adopted in several studies [18, 19, 20]. The JOREK simulation code is used to simulate large-scale MHD instabilities in the core and edge plasma. All simulations were run in simplified 2D slab geometry; extension to 3D toroidal cases is left for future work. Dataset details are summarized in Table 1.

TABLE 1. DATASET SIZES, VARIABLES, AND SPATIAL DIMENSIONS

Dataset	Size	Trajectory length	Variables	Dimensions
Electrostatic JOREK	2000 traj.	200 timesteps	2	100x100
Reduced-MHD JOREK	11391 slices + 20 traj.	10 per slice / 200 per traj	4	100x100

JOREK datasets simulate filamentary blobs in the tokamak edge. Electrostatic JOREK evolves density ρ , electric potential Φ , temperature T , and an auxiliary toroidal vorticity ω . Reduced-MHD JOREK adds poloidal magnetic flux Ψ and toroidal current zj for higher-fidelity physics. Trajectories are 200 timesteps (0.15 μ s each) on a 200x200 Bezier grid.

Datasets were down-sampled spatially and/or temporally due to overresolution during generation for numerical stability and then split between subsets for training/testing. Min-max normalization scaled each variable to $[-1,1]$ to handle large differences in magnitude between variables.

Detailed explanation for the dataset generation and pre-processing can be found in [21]. An example of the simulations can be seen in Fig. 1.

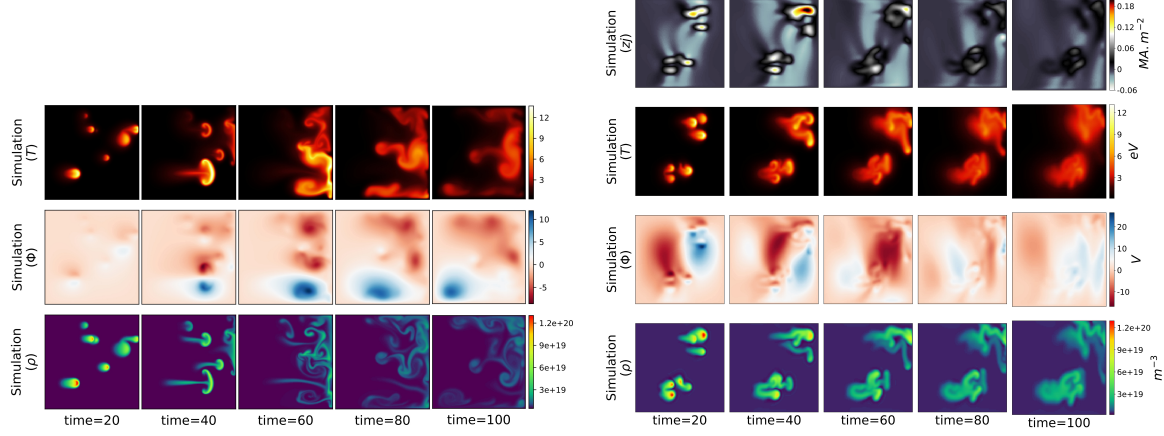


FIG. 1. Example rollout for both datasets: (a) electrostatic JOREK dataset, and (b) reduced-MHD JOREK dataset.

2.2. Model and Training

We employed the PDEArena platform [22], which provides a standardized codebase for implementing neural PDE surrogates and allows seamless adaptation of existing models or integration of new architectures. The data loader was extended to incorporate the simulation datasets used in this study.

Neural operators (NOs) [8, 23] map between continuous function spaces, making them suitable for approximating PDE solutions. In this framework, a neural operator G_θ learns the mapping from input function $a \in A$ to output $u \in U$, $G_\theta(a) = u$. A neural operator is structured as $u = G_\theta(a) = L_{lift} \circ K_1 \circ \dots \circ K_T \circ L_{proj}(a)$, where L_{lift} and L_{proj} are local linear layers projecting to and from a latent space, and K_t are layers combining local linear operations with non-local integral kernel operators. In the Fourier Neural Operator (FNO) [9], the kernel is computed in Fourier space as $(\kappa(a; \phi)v_t)(x) = \mathcal{F}^{-1}(\mathcal{F}(\kappa_\phi) \cdot \mathcal{F}(v_t))(x)$, and the Fourier layer applies learnable weights R in Fourier space and W in Euclidean space, with bias b and nonlinearity σ : $y = \sigma(\mathcal{F}^{-1}(R\mathcal{F}(x)) + Wx + b)$. A modified FNO configuration (‘FNO-128-32m’) was used, with three Fourier blocks (each comprising two layers) and grid discretization concatenated with field data. To enforce non-negativity for variables such as density and temperature, a modified ReLU activation was applied at the output. Only FNO results are presented in detail in this manuscript, but U-net baseline can be found in [21] in the appendix.

FNO hyperparameters largely followed [9] and PDEArena defaults, with minor adjustments to learning rate and gradient clipping. Models were trained using input sequences of T_{in} time steps to predict T_{out} steps simultaneously; typically $T_{in} = 20$, $T_{out} = 5$, except for reduced-MHD JOREK ($T_{in} = 5$). Larger T_{out} was tested but did not improve performance and consistently lost finer spatial details, consistent with prior observations [24]. Models were trained using mean squared error (MSE) averaged over space, time, and fields. To ensure comprehensive coverage of the simulations dynamics through the entire trajectory, the data the model was trained on was created by sampling timeslices starting at random points along the trajectory. Early stopping was applied based on validation performance.

For longer predictions, an autoregressive rollout was used with the model’s output is recursively fed as input for subsequent predictions (Fig. 2). Rollouts began at $t = 20$ to accommodate input sequence length. Errors were computed on min-max normalized fields to prevent numerical overflow for accumulated error. Teacher-forcing tests, in which only true PDE states are used as inputs, were performed as a diagnostic to isolate model approximation error from accumulation effects.

Transfer learning employed a fine-tuning framework. Pre-training was performed on a source dataset (D_S) with abundant samples, followed by fine-tuning on a smaller target dataset (D_T), with learning rate reduction and gradient clipping adjusted for stability. For target datasets with additional variables, a two-stage approach was used: Stage 1 fine-tuned common variables to adapt to the target domain, then Stage 2 trained new variables, allowing gradual adaptation and avoiding architectural modifications.

Additional details for model implementation and training can be found in [21].

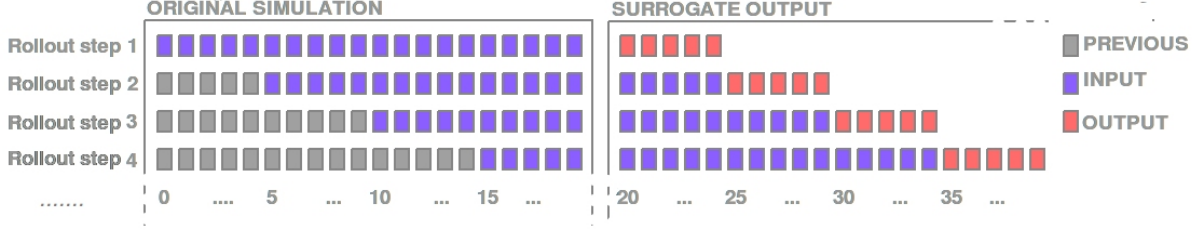


FIG. 2. Example rollout using surrogate with $T_{in} = 20$ and $T_{out} = 5$. For the initial rollout step, the surrogate uses only timesteps from the original simulation, but rollout steps beyond that start using an increasing portion of generated rollout steps.

3. RESULTS

3.1. Baseline model performance

FNOs were trained on two full simulation datasets (Electrostatic JOREK and Reduced-MHD JOREK), converging to the test single-step MSE shown in Table 2. Electrostatic JOREK models excluded auxiliary variables, while Reduced-MHD JOREK included all fields. Reduced-MHD JOREK exhibited the largest overall error, with significant variability across variables.

TABLE 2. MSE ON MINIMUM OUTPUT LENGTH (MEANING 5 TIMESTEPS) FOR EACH RESCALED DATASET VARIABLE AVERAGED ACROSS DIFFERENT STARTING POINTS FOR ALL DATASET TRAJECTORIES

Variable	Electrostatic JOREK (MSE $\pm$ STD)	Reduced-MHD JOREK (MSE $\pm$ STD)
Temperature (T)	$4.34 \pm 11.1 \times 10^{-8}$	$3.21 \pm 13.14 \times 10^{-4}$
Electric potential (Φ)	$2.95 \pm 13.0 \times 10^{-6}$	$1.48 \pm 6.39 \times 10^{-4}$
Density (ρ)	$5.80 \pm 22.7 \times 10^{-6}$	$1.36 \pm 4.52 \times 10^{-4}$
Vorticity (ω)		$5.00 \pm 20.0 \times 10^{-6}$
Magnetic flux (Ψ)		$3.90 \pm 13.5 \times 10^{-5}$
Current (zj)		$1.41 \pm 5.62 \times 10^{-4}$

Trajectory rollouts (density is shown as an example for each model in Fig. 3) show that surrogates capture global features (e.g., density blobs) but finer details diverge over longer rollouts. Both JOREK models exhibit error peaks.

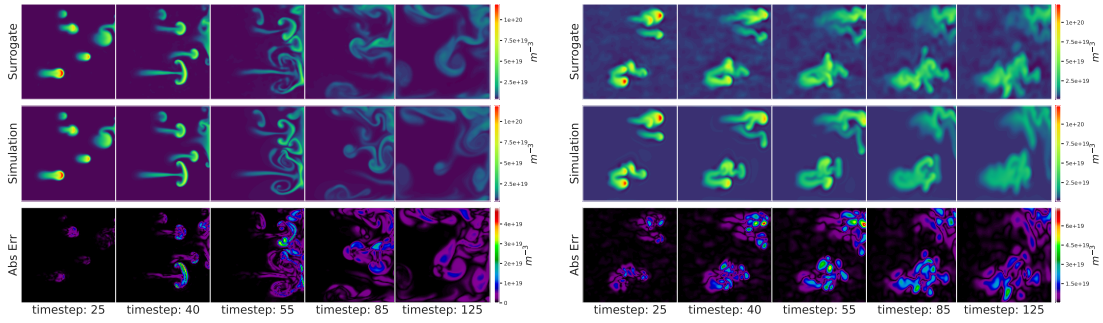


FIG. 3. An example run for the variable particle density ρ (m^{-3}) plotted at specific rollout timesteps for datasets (LEFT) electrostatic JOREK, and (RIGHT) reduced-MHD JOREK. The corresponding examples for rest of the variables can be found in [21]

3.2. Impact of trajectory location with JOREK datasets

Pronounced error spikes occur in both Electrostatic and reduced-MHD JOREK datasets at specific simulation points, independent of cumulative input error: around $t = 50$ for electrostatic JOREK and $t = 30$ for reduced-MHD JOREK. This error spike is shown in Fig. 4 for electrostatic JOREK dataset, and the corresponding figure for reduced-MHD JOREK can be found in [21]. Trajectories in both datasets show consistent physical behavior: blobs initialized at random positions move toward the boundary, interact with it, and dissipate. This indicates that specific physical events, rather than gradual error accumulation, dominate the observed spikes.

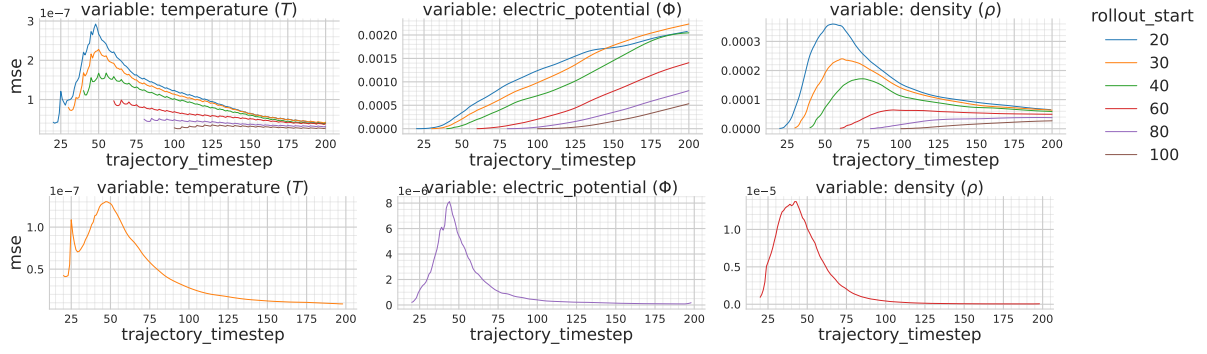


FIG. 4. Long-rollout FNO errors for electrostatic JOREK fields: (TOP ROW) autoregressive rollouts from different start points, (BOTTOM ROW) teacher-forcing testing.

Field-dependent error patterns emerge in Electrostatic JOREK: electric potential shows steady error growth over time, temperature peaks around $t = 50$, and density exhibits a mix of both behaviors. These patterns persist under teacher forcing (Fig. 4), indicating systemic model limitations beyond input-error accumulation. A similar pattern of behaviour is observed for the different variables for the reduced-MHD JOREK model. Periodic error spikes at five-step intervals in Electrostatic JOREK temperature suggest minor discontinuities due to the model’s chunked prediction strategy.

Using the model trained on the electrostatic JOREK dataset, analysis of the pointwise error (Fig. 5) revealed distinct spatial patterns. Pointwise error involves computing the error at each spatial point for a given time step, then averaging this error across the validation dataset to assess the model’s accuracy across space and time. Errors were predominantly localized near the boundaries, particularly the right-hand side wall during early timesteps. The error peaks at $t = 50$, coinciding with the point at which blobs collided with the wall (see Fig. 6). Over time, these localized errors expanded to encompass other boundaries. This spatial distribution suggests that boundary interactions present a challenge for the neural operator. It is important to note that within the neural operator, there is no explicit consideration of boundary conditions, other than the observed behavior at the edges of the spatial dimensions for the simulation in the dataset.

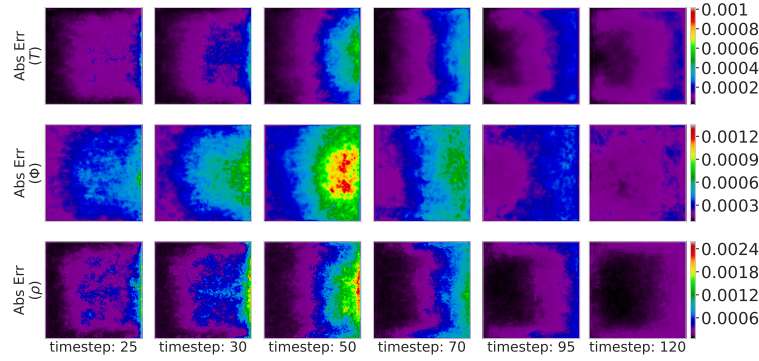


FIG. 5. This figure illustrates the pointwise model error for the electrostatic JOREK dataset at specific time steps, averaged across the validation dataset. All errors are calculated from the rescaled fields.

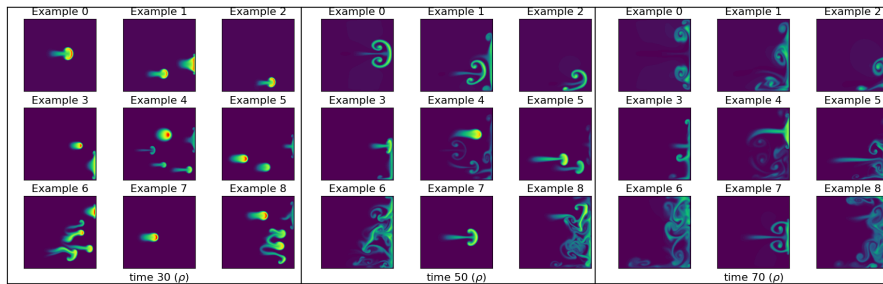


FIG. 6. Electrostatic JOREK FNO rollout examples for density; blobs typically impact the wall around $t = 50$.

However, this pattern was not observed in the reduced-MHD JOREK dataset (Fig. 7), where blobs took a much longer time to travel to the right boundary and diffused or dissipated before hitting it (for an example), and the peak error at $t = 30$ occurred earlier in the trajectory well before the blobs have hit the right wall. This divergence suggests that boundary effects alone cannot fully explain the observed error spikes, as other factors must be contributing to the dynamics.

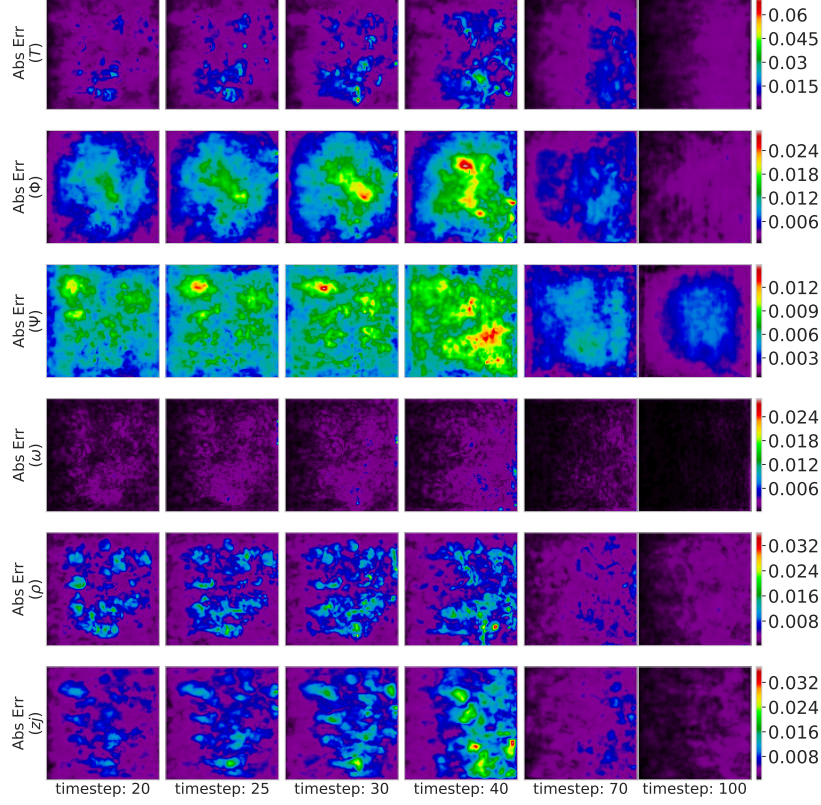


FIG. 7. This figure illustrates the pointwise model error for the reduced-MHD JOREK dataset at specific time steps, averaged across the validation dataset. The results indicate that the error is not concentrated near the boundaries unlike electrostatic JOREK. All errors are calculated from the rescaled fields.

Resampling or varying blob counts did not affect the timing of error spikes, indicating that under-representation in training data is not the cause. Normalized error metrics confirm that trends are not simply due to decreasing solution magnitudes. These additional results can be found in the appendix for [21].

The rapid acceleration and deceleration of blobs present a particular challenge for the FNO, as abrupt localized dynamics are difficult to capture. Increasing the input buffer could help but reduces efficiency, and the MSE objective may prioritize precise blob positions over more statistically relevant features like heat flux. Overall, error spikes arise from an interplay of physical dynamics and model limitations, highlighting the need to consider both in long-term surrogate predictions.

3.3. Transfer learning

This section presents the results of transfer learning experiments.

Transfer from low- to high-fidelity JOREK datasets aimed to leverage lower-fidelity data to improve performance on high-fidelity simulations. The FNO benefited substantially for short rollouts, with error reductions of up to an order of magnitude for small datasets (Fig. 8 for the results for density). The improvements decreased when the dataset size was increased and when longer rollouts were considered. This demonstrates the potential of using cheaper, low-fidelity simulations to reduce the computational burden of generating high-fidelity training data.

The case explored tested transfer learning to target datasets containing previously unseen variables, such as temperature and current. A two-step transfer approach allowed partial knowledge transfer (Fig. 9). Transfer models improved short-rollout performance but offered limited benefits for longer rollouts, and in longer rollouts negative transfer occurred. These results emphasize the challenges of applying transfer learning across datasets with non-overlapping variable sets and the need for careful adaptation to prevent performance degradation.

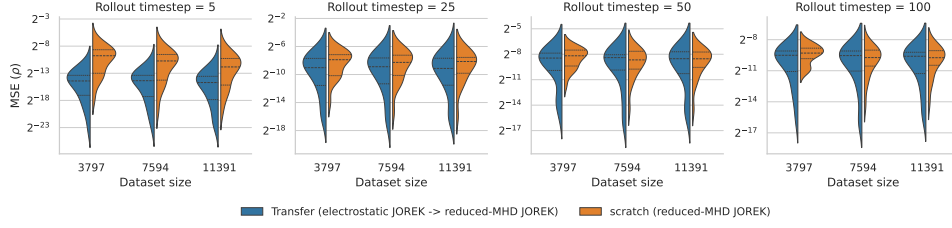


FIG. 8. Scratch and transfer model error at different timesteps and different dataset sizes on reduced-MHD JOREK for the density variable. Scratch was trained from scratch on the reduced-MHD JOREK dataset whilst transfer model was first trained on electrostatic JOREK and then finetuned on reduced-MHD JOREK. The line is the medium error and the error bars correspond to the 16th and 84th percentile.

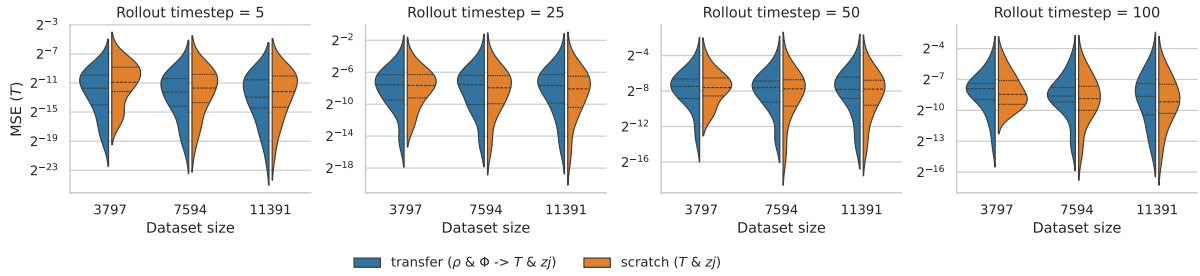


FIG. 9. Scratch and transfer model error at different timesteps and dataset sizes on reduced-MHD JOREK for the temperature variable. Scratch model was trained directly on the corresponding variables on the dataset. Transfer x2 model was first trained on electrostatic JOREK density and electric potential, finetuned on reduced-MHD JOREK density and electric potential (similar to prior section) and the transferred to reduced-MHD JOREK temperature and current. The line is the medium error and the error bars correspond to the 16th and 84th percentile.

Overall, transfer learning is most effective when source and target datasets share similarities, particularly for short-term predictions or when data is limited, while long-rollout performance remains a challenge.

4. CONCLUSION

This work demonstrates that Fourier Neural Operators (FNOs) are effective surrogate models for tokamak plasma simulations. FNOs reproduce short-term plasma dynamics from JOREK datasets with high accuracy, though errors accumulate during long rollouts, particularly near boundaries and specific trajectory events. This highlights the importance of incorporating underlying physical dynamics in surrogate evaluation.

Transfer learning was shown to improve dataset efficiency by leveraging low- to high-fidelity simulations, reducing the need for expensive high-fidelity data, particularly in short-term predictions. However, its effectiveness diminishes for long rollouts or when the source and target datasets differ in physical properties. Knowledge transfer to previously unseen variables remains challenging, sometimes leading to negative transfer effects, indicating the need for more flexible approaches to accommodate varying or non-overlapping variable sets.

Overall, FNO-based surrogates provide a promising path for accelerating plasma simulations and enabling efficient exploration of high-fidelity models.

4.1. Limitations and Future Work

Key limitations include the reliance on FNO and U-Net architectures, which are now outperformed by modern attention-based methods, and the absence of explicit enforcement of physical constraints. The MSE loss function may underrepresent fine-scale, low-amplitude features.

Future work should explore attention-based architectures, physics-informed loss functions, and advanced techniques such as LoRA, domain adaptation, and iterative refinement to improve long-term prediction and transfer learning robustness. Scaling to higher-dimensional, multiphysics simulations will require data-efficient strategies, including active learning, hybrid numerical-AI approaches, and more sophisticated transfer learning methods. Addressing these limitations is essential for developing robust, scalable, and physically consistent AI surrogates in fusion research.

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 the 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. The authors thank the reviewers for their very constructive comments that substantially helped improve the the paper.

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] G L Derks et al. “Benchmark of a self-consistent dynamic 1D divertor model DIV1D using the 2D SOLPS-ITER code”. In: *Plasma Physics and Controlled Fusion* 64.12 (Nov. 2022), p. 125013. ISSN: 1361-6587. DOI: 10.1088/1361-6587/ac9dbd. URL: <http://dx.doi.org/10.1088/1361-6587/ac9dbd>.
- [4] Ernst Hairer, Syvert Norsett, and Gerhard Wanner. *Solving Ordinary Differential Equations I: Nonstiff Problems*. Vol. 8. Jan. 1993. ISBN: 978-3-540-56670-0. DOI: 10.1007/978-3-540-78862-1.
- [5] S. Wiesen et al. “The new SOLPS-ITER code package”. In: *Journal of Nuclear Materials* 463 (2015). PLASMA-SURFACE INTERACTIONS 21, pp. 480–484. ISSN: 0022-3115. DOI: <https://doi.org/10.1016/j.jnucmat.2014.10.012>. URL: <https://www.sciencedirect.com/science/article/pii/S0022311514006965>.
- [6] Vignesh Gopakumar and D Samaddar. “Image mapping the temporal evolution of edge characteristics in tokamaks using neural networks”. In: *Machine Learning: Science and Technology* 1.1 (Feb. 2020), p. 015006. ISSN: 2632-2153. DOI: 10.1088/2632-2153/ab5639. URL: <http://dx.doi.org/10.1088/2632-2153/ab5639>.
- [7] Stefan Dasbach and Sven Wiesen. “Towards fast surrogate models for interpolation of tokamak edge plasmas”. In: *Nuclear Materials and Energy* 34 (2023), p. 101396. ISSN: 2352-1791. DOI: <https://doi.org/10.1016/j.nme.2023.101396>. URL: <https://www.sciencedirect.com/science/article/pii/S2352179123000352>.
- [8] Nikola Kovachki et al. *Neural Operator: Learning Maps Between Function Spaces*. 2023. arXiv: 2108.08481 [cs.LG].
- [9] Zongyi Li et al. *Fourier Neural Operator for Parametric Partial Differential Equations*. 2021. arXiv: 2010.08895 [cs.LG].
- [10] Benedikt Alkin et al. *Universal Physics Transformers: A Framework For Efficiently Scaling Neural Operators*. 2024. arXiv: 2402.12365 [cs.LG]. URL: <https://arxiv.org/abs/2402.12365>.
- [11] Vignesh Gopakumar et al. “Plasma surrogate modelling using Fourier neural operators”. In: *Nuclear Fusion* 64.5 (Apr. 2024), p. 056025. ISSN: 1741-4326. DOI: 10.1088/1741-4326/ad313a. URL: <http://dx.doi.org/10.1088/1741-4326/ad313a>.
- [12] Yoeri Poels et al. “Fast dynamic 1D simulation of divertor plasmas with neural PDE surrogates”. In: *Nuclear Fusion* 63.12 (Sept. 2023), p. 126012. ISSN: 1741-4326. DOI: 10.1088/1741-4326/acf70d. URL: <http://dx.doi.org/10.1088/1741-4326/acf70d>.
- [13] K. L. van de Plassche et al. “Fast modeling of turbulent transport in fusion plasmas using neural networks”. In: *Physics of Plasmas* 27.2 (Feb. 2020). ISSN: 1089-7674. DOI: 10.1063/1.5134126. URL: <http://dx.doi.org/10.1063/1.5134126>.
- [14] A. Ho et al. “Neural network surrogate of QuaLiKiz using JET experimental data to populate training space”. In: *Physics of Plasmas* 28.3 (Mar. 2021). ISSN: 1089-7674. DOI: 10.1063/5.0038290. URL: <http://dx.doi.org/10.1063/5.0038290>.
- [15] P Mánek et al. “Fast regression of the tritium breeding ratio in fusion reactors”. In: *Machine Learning: Science and Technology* 4.1 (Jan. 2023), p. 015008. ISSN: 2632-2153. DOI: 10.1088/2632-2153/acb2b3. URL: <http://dx.doi.org/10.1088/2632-2153/acb2b3>.
- [16] Vignesh Gopakumar et al. “Plasma surrogate modelling using Fourier neural operators”. In: *Nuclear Fusion* 64.5 (Apr. 2024), p. 056025. DOI: 10.1088/1741-4326/ad313a. URL: <https://dx.doi.org/10.1088/1741-4326/ad313a>.

- [17] 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>.
- [18] S. J. P. Pamela et al. “Neural-Parareal: Self-improving acceleration of fusion MHD simulations using time-parallelisation and neural operators”. In: *Computer Physics Communications* 307.109391 (2025). DOI: <https://doi.org/10.1016/j.cpc.2024.109391>.
- [19] Vignesh Gopakumar et al. *Fourier Neural Operator for Plasma Modelling*. 2023. arXiv: 2302.06542 [physics.plasm-ph]. URL: <https://arxiv.org/abs/2302.06542>.
- [20] N. Carey et al. *Data efficiency and long term prediction capabilities for neural operator surrogate models of core and edge plasma codes*. 2024. arXiv: 2402.08561 [physics.plasm-ph]. URL: <https://arxiv.org/abs/2402.08561>.
- [21] N. Carey et al. *Neural operator surrogate models of plasma edge simulations: feasibility and data efficiency*. 2025. arXiv: 2502.17386 [physics.plasm-ph]. URL: <https://arxiv.org/abs/2502.17386>.
- [22] Jayesh K. Gupta and Johannes Brandstetter. *Towards Multi-spatiotemporal-scale Generalized PDE Modeling*. 2022. arXiv: 2209.15616 [cs.LG].
- [23] 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>.
- [24] Phillip Lippe et al. *PDE-Refiner: Achieving Accurate Long Rollouts with Neural PDE Solvers*. 2023. arXiv: 2308.05732 [cs.LG].