TGLF-SINN: Deep Learning Surrogate Model for Accelerating Turbulent Transport Modeling in Fusion

Yadi Cao[*]
University of California, San Diego
yac066@ucsd.edu

Futian Zhang[*]
University of California, San Diego
f6zhang@ucsd.edu

Wesley Liu[*]
University of California, San Diego
wyl002@ucsd.edu

Tom Neiser
General Atomics
neisert@fusion.gat.com

Orso Meneghini
General Atomics
meneghini@fusion.gat.com

Lawson Fuller
University of California, San Diego
lawson@lawsonfuller.com

Sterling Smith
General Atomics
smithsp@fusion.gat.com

Raffi Nazikian
General Atomics
nazikian@fusion.gat.com

Brian Sammuli
General Atomics
sammuli@fusion.gat.com

Rose Yu
University of California, San Diego
roseyu@ucsd.edu

Abstract

Fusion simulations require accurate modeling of core turbulent transport in tokamaks, but state-of-the-art numerical simulators like TGLF (Trapped Gyro-Landau Fluid) are too slow for interactive device simulation due to coupling requirements with other components. Neural network surrogates offer speedup but require massive datasets to cover diverse transport flux variations, limiting adaptation to full-scale gyrokinetic simulations. We propose **TGLF-SINN (Spectra-Informed Neural Network for TGLF)** with three innovations: (1) principled Feature tuning to compress wide-ranging outputs, (2) physics-informed regularization using turbulent energy spectrum, and (3) Bayesian Active Learning (BAL) for smart training data selection. Our approach achieves 12.4% better accuracy than SOTA NN surrogates in the offline setting, while maintaining competitive performance with only 25% of training data via BAL. In flux matching with neoclassical simulations, we demonstrate $45\times$ speedup compared to optimized numerical methods while preserving accuracy, enabling interactive device simulation, and accelerating fusion energy research.

---

[*] Equal contribution.

39th Conference on Neural Information Processing Systems (NeurIPS 2025) Workshop: .

1 Introduction

Fusion energy centers around the simulation of turbulent transport behavior in the core plasma region of magnetic confinement devices—tokamaks, as turbulent transport (i.e., how energy and particles move through the plasma) directly impacts reactor performance [1, 2, 3].

The full-scale approach for simulating turbulent transport is gyrokinetic simulation, which solves complex physics equations but requires hours of supercomputer time per evaluation [4, 5, 6]. To address this bottleneck, physicists developed reduced-order models like TGLF (Trapped Gyro-Landau Fluid) [7, 8, 9] that approximate gyrokinetic results in seconds. However, TGLF still becomes expensive in whole-device simulations that require hundreds of transport queries per time step.

Neural networks as surrogate models to TGLF then offer a promising solution by cutting the single inference time to microseconds [10, 11]. However, existing approaches still rely on significantly large (2.5 million evaluations), clean datasets; this is because the wide variation in transport fluxes (values spanning many orders of magnitude) across different plasma conditions is hard to train and capture, and training tends to be unstable. These strong requirements thus make training for the full-scale method (gyrokinetic) impossible due to the prohibitive cumulative cost of simulation.

Tackling these challenges, we propose **TGLF-SINN (Spectra-Informed Neural Network)** to address these data efficiency and stability challenges through three key innovations:

- **Feature tuning**: Systematic preprocessing and loss design to handle the extreme variation in transport fluxes, making neural network training more stable.
- **Spectral Regularization**: Leveraging physics knowledge by using intermediate turbulent transport spectrum information as additional training regularization.
- **Bayesian Active Learning (BAL)**: Smart sample selection to minimize training data requirements while maximizing information gain and maintaining accuracy.

Our experiments demonstrate that TGLF-SINN achieves 12.4% better accuracy than the state-of-the-art baseline in the offline setting. The model also shows improved robustness against data noise and scarcity. Through Bayesian Active Learning, we achieve competitive performance using only 25% of the training data. In downstream fusion simulation workflows, flux-matching with neoclassical simulations, our surrogate provides a 45× speedup while maintaining accuracy comparable to the original numerical simulator.

2 Related Work

Physics-Based Transport Models. TGLF [7, 8] computes transport by simulating a linearized version of gyrokinetics, evaluated at different wavelengths, and applying empirical "saturation rules" to estimate nonlinear effects. It represents a careful balance between physics fidelity and computational efficiency, validated extensively on experimental data [10, 11]. TGLF serves as our numerical solver baseline in the flux-matching experiments.

Neural Surrogates for Physics Simulations. Machine learning has shown promise for accelerating physics simulations across domains [12, 13, 14]. In fusion, two main approaches exist: (1) learning complete plasma evolution using CNNs and LSTMs [15, 16], which struggles with high dimensionality due to multiphysics coupling, and (2) surrogating individual components like transport models [17, 10, 11, 18], which offers more flexibility and easier data collection.

The current state-of-the-art, TGLF-NN [10, 11], follows the latter approach using an encoder-ResNet-decoder architecture but requires ensemble averaging to handle output variation, leading to slow and unstable training. Alternative approaches like QuaLiKiz-NN [18, 19] have limitations in model assumptions (e.g., do not allow negative growth rates). TGLF-NN serves as our neural surrogate baseline in experiments.

Active Learning for Expensive Simulations. Bayesian Active Learning strategically selects training samples to maximize information gain, particularly valuable when data generation is expensive [20, 21]. This is highly relevant for fusion applications where generating training data requires substantial computational resources. We use BAL to study the data efficiency of our approach.

3 Methodology

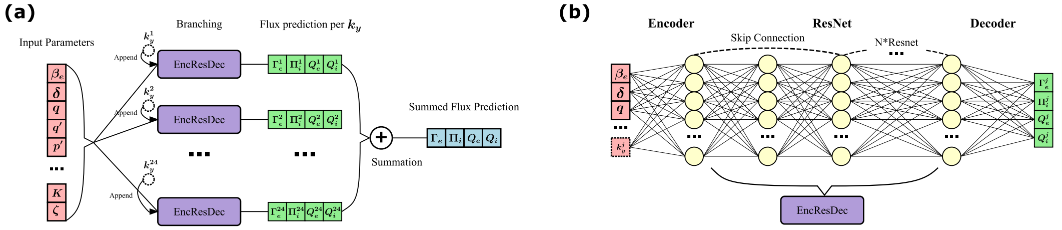


Figure 1: **Overview of TGLF-SINN architecture.** (a) TGLF-SINN takes 31 input features and splits into 24 branches. Each branch produces 4 transport predictions for a specific wavelength. These are summed for final predictions. (b) The shared neural network backbone.

Figure 1 shows our approach: TGLF-SINN takes 31-dimensional physics parameters consistent with TGLF-NN and the numerical solver, but we split a single flux prediction into 24 parallel branches, each predicting the contributions from one specific wavelength. These contributions are summed to produce the final flux predictions.

3.1 Dataset and Preprocessing

We generated datasets by systematically sampling the 31-dimensional physics parameter space and querying the TGLF numerical solver. Two dataset variations were created: a smaller 330,000-sample set and a comprehensive 3-million-sample set. Since TGLF simulations can fail and produce erroneous numbers, we apply iterative Median Absolute Deviation (MAD) filtering (see details in appendix C.1) to remove outlier problematic samples, reducing dataset size by 10-15%.

3.2 Feature tuning for Extreme Value Ranges

Transport predictions span many orders of magnitude, creating instability. Unlike previous ad-hoc approaches, we adopt principled Feature tuning to handle the large variation in flux magnitudes. Let $\boldsymbol{y} = \{\Gamma_e, \Pi_i, Q_e, Q_i\}$ denote the concatenated fluxes and $\boldsymbol{y}^j = \{{\Gamma_e}^j(k_y^j), {\Pi_i}^j(k_y^j), Q_e^j(k_y^j), Q_i^j(k_y^j)\}$ denote the contributions at the j-th wavenumber for $j = 1, \ldots, N_y$. We apply inverse hyperbolic sine ($\sinh^{-1}$) transformations followed by standardization to both the final flux predictions $\boldsymbol{y}' = (\mathcal{S} \circ \sinh^{-1})(\boldsymbol{y})$ and the per-wavelength contributions $\boldsymbol{y}'^j = (\mathcal{S} \circ \sinh^{-1})(\boldsymbol{y}^j)$, where $\mathcal{S}$ represents standardization across the training dataset. The choice of $\sinh^{-1}$ over $\log$ is motivated by the need to accommodate possible negative predictions in TGLF.

3.3 Physics-Informed Spectral Regularization

TGLF computes the total turbulent fluxes of particles Γ, momentum Π, and heat Q as sums over wavelength-specific contributions using quasi-linear theory and empirical saturation rules [8]:

$$\Gamma = \sum_{j=1}^{N_y} \Gamma^j(k_y^j), \quad \Pi = \sum_{j=1}^{N_y} \Pi^j(k_y^j), \quad Q = \sum_{j=1}^{N_y} Q^j(k_y^j), \tag{1}$$

where the superscript ^{j} denotes the contribution from the j-th wavelength k_y^j, and $N_y = 24$ is the total number of wavelengths considered.

We leverage this physics structure by appending the wavenumbers $\{k_y^1, \ldots, k_y^{N_y}\}$ to the 31-dimensional input parameters $\boldsymbol{x}$ to enable per-wavenumber predictions. The supervised loss on the turbulent transport flux is defined as:

$$\mathcal{L}_f = \|\boldsymbol{y}' - \sum_{j=1}^{N_y} f_\theta\left(\boldsymbol{x}, k_y^j\right)\|^2, \tag{2}$$

where $\boldsymbol{y} = (\Gamma, \Pi, Q)$ is the prediction target, $\boldsymbol{y}'$ is the value after Feature tuning in Section 3.2, and f_θ is the neural network with trainable parameters θ. In addition, we impose a regularization term $\mathcal{L}_s$

on the per-wavelength fluxes:

$$\mathcal{L}_s = \sum_{j=1}^{N_y} \| \boldsymbol{y}'^{j} - f_\theta\left(\boldsymbol{x}, k_y^j\right) \|^2. \tag{3}$$

Combined, the overall objective function reads:

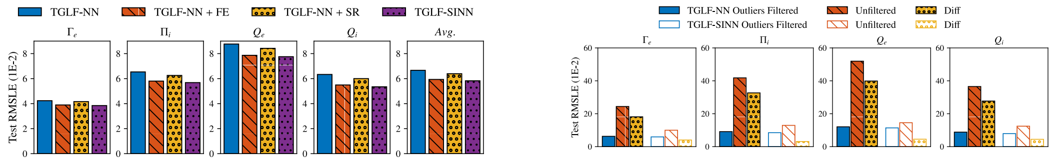
$$\mathcal{L} = \mathcal{L}_f + \mathcal{L}_s. \tag{4}$$

Importantly, $\mathcal{L}_f$ and $\mathcal{L}_s$ are not redundant: $\mathcal{L}_f$ supervises the final aggregated flux prediction (post-summation), while $\mathcal{L}_s$ supervises each individual wavelength contribution before summation. This dual supervision acts as a physics-informed regularization that constrains the model to learn physically meaningful spectral decompositions. While a perfect model would theoretically minimize both losses simultaneously, in practice this regularization provides additional constraints that improve generalization, especially with limited training data, as demonstrated in our ablation studies (Section 4.1).

3.4 Bayesian Active Learning

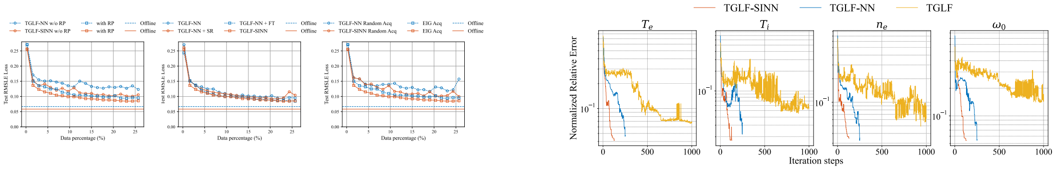
To reduce training data requirements, we employ Bayesian Active Learning with Expected Information Gain (EIG) acquisition. Starting with a small initial dataset, we iteratively: (1) use model uncertainty to identify informative candidate samples, (2) generate labels via physics simulations, and (3) retrain the model. We incorporate physics knowledge by sampling candidates according to parameter distributions at different spatial locations in the device, improving sample efficiency.

4 Experimental Results



(a) Offline model's performance gain

(b) Robustness against noisy and sparse data



(c) Performance gain by BAL

(d) Faster coupling using our method

Figure 2: **Comprehensive experimental results.** (a) TGLF-SINN achieves the lowest error across all transport channels (FT = Feature Tuning, SR = Spectral Regularization). (b) it shows robustness against datasets that are sparser and contain erroneous results. (c) Bayesian Active Learning ablation results: (left) TGLF-SINN achieves the fastest error reduction, (middle) EIG outperforms random sampling, (right) physics-informed candidate selection (RP = Radius-based Proposing) improves performance. (d) Flux-matching convergence: TGLF-SINN converges in 129 iterations and enables 45× speedup for coupling between turbulent transport and neoclassical simulations.

4.1 Offline Performance and Robustness

On the standard benchmark dataset, TGLF-SINN reduces root mean squared logarithmic error from 6.66×10^{-2} to 5.83×10^{-2}, achieving 12.4% improvement over the state-of-the-art baseline (Figure 2 (a)). Feature tuning and spectral regularization contribute 11.0% and 4.1% improvements respectively.

To test robustness, we evaluated performance using smaller datasets (major perturbations only) without outlier filtering to mimic real-world data limitations. While the baseline's error increases

363% with this data contamination, TGLF-SINN degrades only 51%, demonstrating superior robustness (Figure 2 (b)). Comprehensive ablation studies across all model combinations and dataset contaminations are provided in Appendix A.

4.2 Active Learning Results

Bayesian Active Learning with EIG acquisition significantly reduces data requirements. Using only 25% of the full dataset, TGLF-SINN achieves error only 0.0165 higher than the baseline's full-data performance—a 12.4% improvement over the baseline with the same active learning settings (Figures 2 (c)).

In the end of this section, we note that while EIG acquisition incurs additional computational cost compared to random sampling, this overhead is justified by the significant reduction in required training data (25% of full dataset vs. 100%). Furthermore, when extending this approach to higher-fidelity simulations such as CGYRO [22, 23]—where a single evaluation takes hours on supercomputers—the acquisition overhead becomes negligible compared to the cost saved by requiring fewer simulation evaluations. For TGLF, where evaluations are relatively fast (seconds), the benefit is less pronounced but still demonstrates the viability of the BAL framework for future high-fidelity applications.

4.3 Downstream Application: Flux-Matching

We integrated TGLF-SINN into fusion simulation workflows, and validated on real DIII-D tokamak experimental data. TGLF-SINN converges in 129-412 iterations compared to 251-364 for TGLF-NN, while the numerical TGLF solver fails to converge (Figure 2 (d)). The $45\times$ speedup (20 seconds vs 15 minutes) enables practical whole-device simulation while maintaining accuracy. We provide the complete results (including convergence analysis, profile analysis, stored heat energy calculation, etc.) in Appendix E.

5 Conclusion and Future Work

We presented TGLF-SINN, a neural surrogate for fusion transport modeling that addresses key challenges: handling extreme value ranges, leveraging physics-informed priors, and Bayesian active learning to reduce training data requirements while still achieving superior accuracy and robustness. In downstream fusion simulation workflows, TGLF-SINN enables a $45\times$ speedup while maintaining accuracy comparable to the original numerical simulator. However, our method still has limitations. For example, the current implementation does not explicitly enforce fundamental conservation laws (energy, momentum, particle conservation) as hard constraints. Future work could construct constrained layers to guarantee physical consistency. Additionally, systematic biases observed in flux-matching applications stem from underlying TGLF ROM limitations in certain plasma regimes (e.g., plasma edge regions with large ρ). Addressing these requires learning directly from higher-fidelity CGYRO simulations [22, 23] in regions where TGLF's quasi-linear assumptions break down. Our demonstrated data efficiency and robustness make this extension tractable, establishing a pathway toward more accurate and physically faithful surrogates for fusion simulation.

Acknowledgments

This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Fusion Energy Sciences, using the DIII-D National Fusion Facility, a DOE Office of Science user facility, under Awards DE-FC02-04ER54698 (DIII-D), DE-SC0024426 (FDP), DE-SC0018990 (MAST-U), DE-SC0022031 (ALF), DE-SC0017992 (AToM), DE-FG02-95ER54309 (GA Theory), and the DOE Office of Workforce Development for Teachers and Scientists (WDTS) under the Science Undergraduate Laboratory Internships (SULI) program. This work used computational resources of the National Energy Research Scientific Computing Center (NERSC), a Department of Energy User Facility using NERSC awards FES-ERCAP 30971 and FES-ERCAP 33303, and the computational resources of NRP Nautilus, which is supported by National Science Foundation (NSF) awards CNS-1730158, ACI-1540112, ACI-1541349, OAC-1826967, OAC-2112167, CNS-2100237, CNS-2120019.

References

[1] P. Bonoli, L. C. McInnes, C. Sovinec, D. Brennan, T. Rognlien, P. Snyder, J. Candy, C. Kessel, J. Hittinger, L. Chacon, and . others. Report of the workshop on integrated simulations for magnetic fusion energy sciences. Office of Fusion Energy Sci. and the Office of Adv. Sci. Comput. Res., Tech. Rep, 2015.

[2] E. Suchyta, S. Klasky, N. Podhorszki, M. Wolf, A. Adesoji, C. Chang, J. Choi, P. E. Davis, J. Dominski, S. Ethier, and . others. The exascale framework for high fidelity coupled simulations (effis): Enabling whole device modeling in fusion science. The International Journal of High Performance Computing Applications, 36(1):106–128, 2022.

[3] O. Meneghini, T. Slendebroek, B.C. Lyons, K. McLaughlin, J. McClenaghan, L. Stagner, J. Harvey, T.F. Neiser, A. Ghiozzi, G. Dose, J. Guterl, A. Zalzali, T. Cote, N. Shi, D. Weisberg, S.P. Smith, B.A. Grierson, and J. Candy. FUSE (Fusion Synthesis Engine): A Next Generation Framework for Integrated Design of Fusion Pilot Plants. arXiv, 2024.

[4] A. Dimits, B. Cohen, N. Mattor, W. Nevins, D. Shumaker, S. Parker, and C. Kim. Simulation of ion temperature gradient turbulence in tokamaks. Nuclear fusion, 40(3Y):661, 2000.

[5] J. Candy and R. Waltz. An eulerian gyrokinetic-maxwell solver. Journal of Computational Physics, 186(2):545–581, 2003.

[6] N. Howard, C. Holland, A. White, M. Greenwald, and J. Candy. Multi-scale gyrokinetic simulation of tokamak plasmas: enhanced heat loss due to cross-scale coupling of plasma turbulence. Nuclear Fusion, 56(1):014004, 2015.

[7] G. Staebler, J. Kinsey, and R. Waltz. Gyro-landau fluid equations for trapped and passing particles. Physics of Plasmas, 12(10), 2005.

[8] G. Staebler, J. Kinsey, and R. Waltz. A theory-based transport model with comprehensive physics. Physics of Plasmas, 14(5), 2007.

[9] J. Kinsey, G. Staebler, and R. Waltz. The first transport code simulations using the trapped gyro-landau-fluid model. Physics of Plasmas, 15(5), 2008.

[10] O. Meneghini, S. P. Smith, P. B. Snyder, G. M. Staebler, J. Candy, E. Belli, L. Lao, M. Kostuk, T. Luce, T. Luda, and . others. Self-consistent core-pedestal transport simulations with neural network accelerated models. Nuclear Fusion, 57(8):086034, 2017.

[11] T. Neiser, O. Meneghini, S. Smith, J. McClenaghan, D. Orozco, J. Hall, G. Staebler, E. Belli, and J. Candy. Database generation for validation of TGLF and retraining of neural network accelerated TGLF-NN. In Bulletin of the American Physical Society, 2022.

[12] T. Pfaff, M. Fortunato, A. Sanchez-Gonzalez, and P. Battaglia. Learning mesh-based simulation with graph networks. In International conference on learning representations, 2020.

[13] D. Wu, L. Gao, X. Xiong, M. Chinazzi, A. Vespignani, Y. Ma, and R. Yu. Deepgleam: a hybrid mechanistic and deep learning model for covid-19 forecasting. arXiv preprint arXiv:2102.06684, 2021.

[14] Y. Cao, M. Chai, M. Li, and C. Jiang. Efficient learning of mesh-based physical simulation with bi-stride multi-scale graph neural network. In International conference on machine learning, pages 3541–3558. PMLR, 2023.

[15] J. Abbate, R. Conlin, and E. Kolemen. Data-driven profile prediction for diii-d. Nuclear Fusion, 61(4):046027, 2021.

[16] M. E. Fenstermacher, . DIII-D Team:, J. Abbate, S. Abe, T. Abrams, M. Adams, B. Adamson, N. Aiba, T. Akiyama, P. Aleynikov, E. Allen, S. Allen, H. Anand, J. Anderson, Y. Andrew, T. Andrews, D. Appelt, R. Arbon, N. Ashikawa, A. Ashourvan, M. Aslin, Y. Asnis, M. Austin, D. Ayala, J. Bak, I. Bandyopadhyay, S. Banerjee, K. Barada, L. Bardoczi, J. Barr, E. Bass, D. Battaglia, A. Battey, W. Baumgartner, L. Baylor, J. Beckers, M. Beidler, E. Belli, J. Berkery, T. Bernard, N. Bertelli, M. Beurskens, R. Bielajew, S. Bilgili, B. Biswas, S. Blondel, J. Boedo, I. Bogatu, R. Boivin, T. Bolzonella, M. Bongard, X. Bonnin, P. Bonoli, M. Bonotto, A. Bortolon, S. Bose, N. Bosviel, S. Bouwmans, M. Boyer, W. Boyes, L. Bradley, R. Brambila, D. Brennan, S. Bringuier, L. Brodsky, M. Brookman, J. Brooks, D. Brower, G. Brown, W. Brown, M. Burke, K. Burrell, K. Butler, R. Buttery, I. Bykov, P. Byrne, A. Cacheris, K. Callahan, J. Callen, G. Campbell, J. Candy, J. Canik, P. Cano-Megias, N. Cao, L. Carayannopoulos, T. Carlstrom,

W. Carrig, T. Carter, W. Cary, L. Casali, M. Cengher, G. Cespedes Paz, R. Chaban, V. Chan, B. Chapman, I. Char, A. Chattopadhyay, R. Chen, J. Chen, X. Chen, X. Chen, J. Chen, M. Chen, J. Chen, Z. Chen, M. Choi, W. Choi, G. Choi, L. Chousal, C. Chrobak, C. Chrystal, Y. Chung, R. Churchill, M. Cianciosa, J. Clark, M. Clement, S. Coda, A. Cole, C. Collins, W. Conlin, A. Cooper, J. Cordell, B. Coriton, T. Cote, J. Cothran, A. Creely, N. Crocker, C. Crowe, B. Crowley, T. Crowley, D. Cruz-Zabala, D. Cummings, M. Curie, D. Curreli, A. Dal Molin, B. Dannels, A. Dautt-Silva, K. Davda, G. De Tommasi, P. De Vries, G. Degrandchamp, J. Degrassie, D. Demers, S. Denk, S. Depasquale, E. Deshazer, A. Diallo, S. Diem, A. Dimits, R. Ding, S. Ding, W. Ding, T. Do, J. Doane, G. Dong, D. Donovan, J. Drake, W. Drews, J. Drobny, X. Du, H. Du, V. Duarte, D. Dudt, C. Dunn, J. Duran, A. Dvorak, F. Effenberg, N. Eidietis, D. Elder, D. Eldon, R. Ellis, W. Elwasif, D. Ennis, K. Erickson, D. Ernst, M. Fasciana, D. Fedorov, E. Feibush, N. Ferraro, J. Ferreira, J. Ferron, P. Fimognari, D. Finkenthal, R. Fitzpatrick, P. Fox, W. Fox, L. Frassinetti, H. Frerichs, H. Frye, Y. Fu, K. Gage, J. Galdon Quiroga, A. Gallo, Q. Gao, A. Garcia, M. Garcia Munoz, D. Garnier, A. Garofalo, A. Gattuso, D. Geng, K. Gentle, D. Ghosh, L. Giacomelli, S. Gibson, E. Gilson, C. Giroud, F. Glass, A. Glasser, D. Glibert, P. Gohil, R. Gomez, S. Gomez, X. Gong, E. Gonzales, A. Goodman, Y. Gorelov, V. Graber, R. Granetz, T. Gray, D. Green, C. Greenfield, M. Greenwald, B. Grierson, R. Groebner, W. Grosnickle, M. Groth, H. Grunloh, S. Gu, W. Guo, H. Guo, P. Gupta, J. Guterl, W. Guttenfelder, T. Guzman, S. Haar, R. Hager, S. Hahn, M. Halfmoon, T. Hall, K. Hallatschek, F. Halpern, G. Hammett, H. Han, E. Hansen, C. Hansen, M. Hansink, J. Hanson, M. Hanson, G. Hao, A. Harris, R. Harvey, S. Haskey, E. Hassan, A. Hassanein, D. Hatch, R. Hawryluk, W. Hayashi, W. Heidbrink, J. Herfindal, J. Hicok, D. Hill, E. Hinson, C. Holcomb, L. Holland, C. Holland, E. Hollmann, J. Hollocombe, A. Holm, I. Holmes, K. Holtrop, M. Honda, R. Hong, R. Hood, A. Horton, L. Horvath, M. Hosokawa, S. Houshmandyar, N. Howard, E. Howell, D. Hoyt, W. Hu, Y. Hu, Q. Hu, J. Huang, Y. Huang, J. Hughes, T. Human, D. Humphreys, P. Huynh, A. Hyatt, C. Ibanez, L. Ibarra, R. Icasas, K. Ida, V. Igochine, Y. In, S. Inoue, A. Isayama, O. Izacard, V. Izzo, A. Jackson, G. Jacobsen, A. Jaervinen, A. Jalalvand, J. Janhunen, S. Jardin, H. Jarleblad, Y. Jeon, H. Ji, X. Jian, E. Joffrin, A. Johansen, C. Johnson, T. Johnson, C. Jones, I. Joseph, D. Jubas, B. Junge, W. Kalb, R. Kalling, C. Kamath, J. Kang, D. Kaplan, A. Kaptanoglu, S. Kasdorf, J. Kates-Harbeck, P. Kazantzidis, A. Kellman, D. Kellman, C. Kessel, K. Khumthong, E. Kim, H. Kim, J. Kim, S. Kim, J. Kim, H. Kim, K. Kim, C. Kim, W. Kimura, M. King, J. King, J. Kinsey, A. Kirk, B. Kiyan, A. Kleiner, V. Klevarova, R. Knapp, M. Knolker, W. Ko, T. Kobayashi, E. Koch, M. Kochan, B. Koel, M. Koepke, A. Kohn, R. Kolasinski, E. Kolemen, E. Kostadinova, M. Kostuk, G. Kramer, D. Kriete, L. Kripner, S. Kubota, J. Kulchar, K. Kwon, R. La Haye, F. Laggner, H. Lan, R. Lantsov, L. Lao, A. Lasa Esquisabel, C. Lasnier, C. Lau, B. Leard, J. Lee, R. Lee, M. Lee, M. Lee, Y. Lee, C. Lee, J. Lee, S. Lee, M. Lehnen, A. Leonard, E. Leppink, M. Lesher, J. Lestz, J. Leuer, N. Leuthold, X. Li, K. Li, E. Li, G. Li, L. Li, Z. Li, J. Li, Y. Li, Z. Lin, D. Lin, X. Liu, J. Liu, Y. Liu, T. Liu, Y. Liu, C. Liu, Z. Liu, C. Liu, D. Liu, A. Liu, D. Liu, A. Loarte-Prieto, L. Lodestro, N. Logan, J. Lohr, B. Lombardo, J. Lore, Q. Luan, T. Luce, T. Luda Di Cortemiglia, N. Luhmann, R. Lunsford, Z. Luo, A. Lvovskiy, B. Lyons, X. Ma, M. Madruga, B. Madsen, C. Maggi, K. Maheshwari, A. Mail, J. Mailloux, R. Maingi, M. Major, M. Makowski, R. Manchanda, C. Marini, A. Marinoni, A. Maris, T. Markovic, L. Marrelli, E. Martin, J. Mateja, G. Matsunaga, R. Maurizio, P. Mauzey, D. Mauzey, G. Mcardle, J. Mcclenaghan, K. Mccollam, C. Mcdevitt, K. Mckay, G. Mckee, A. Mclean, V. Mehta, E. Meier, J. Menard, O. Meneghini, G. Merlo, S. Messer, W. Meyer, C. Michael, C. Michoski, P. Milne, G. Minet, A. Misleh, Y. Mitrishkin, C. Moeller, K. Montes, M. Morales, S. Mordijck, D. Moreau, S. Morosohk, P. Morris, L. Morton, A. Moser, R. Moyer, C. Moynihan, T. Mrazkova, D. Mueller, S. Munaretto, J. Munoz Burgos, C. Murphy, K. Murphy, C. Muscatello, C. Myers, A. Nagy, G. Nandipati, M. Navarro, F. Nave, G. Navratil, R. Nazikian, A. Neff, G. Neilson, T. Neiser, W. Neiswanger, D. Nelson, A. Nelson, F. Nespoli, R. Nguyen, L. Nguyen, X. Nguyen, J. Nichols, M. Nocente, S. Nogami, S. Noraky, N. Norausky, M. Nornberg, R. Nygren, T. Odstrcil, D. Ogas, T. Ogorman, S. Ohdachi, Y. Ohtani, M. Okabayashi, M. Okamoto, L. Olavson, E. Olofsson, M. Omullane, R. Oneill, D. Orlov, W. Orvis, T. Osborne, D. Pace, G. Paganini Canal, A. Pajares Martinez, L. Palacios, C. Pan, Q. Pan, R. Pandit, M. Pandya, A. Pankin, Y. Park, J. Park, J. Park, S. Parker, P. Parks, M. Parsons, B. Patel, C. Pawley, C. Paz-Soldan, W. Peebles, S. Pelton, R. Perillo, C. Petty, Y. Peysson, D. Pierce, A. Pigarov, L. Pigatto, D. Piglowski, S. Pinches, R. Pinsker, P. Piovesan, N. Piper, A. Pironti, R. Pitts, J. Pizzo, U. Plank, M. Podesta, E. Poli, F. Poli, D. Ponce, Z. Popovic, M. Porkolab, G. Porter, C. Powers, S. Powers, R. Prater, Q. Pratt, I. Pusztai, J. Qian, X. Qin, O. Ra, T. Rafiq, T. Raines, R. Raman, J. Rauch, A. Raymond, C. Rea,

M. Reich, A. Reiman, S. Reinhold, M. Reinke, R. Reksoatmodjo, Q. Ren, Y. Ren, J. Ren, M. Rensink, J. Renteria, T. Rhodes, J. Rice, R. Roberts, J. Robinson, P. Rodriguez Fernandez, T. Rognlien, A. Rosenthal, S. Rosiello, J. Rost, J. Roveto, W. Rowan, R. Rozenblat, J. Ruane, D. Rudakov, J. Ruiz Ruiz, R. Rupani, S. Saarelma, S. Sabbagh, J. Sachdev, J. Saenz, S. Saib, M. Salewski, A. Salmi, B. Sammuli, C. Samuell, A. Sandorfi, C. Sang, J. Sarff, O. Sauter, K. Schaubel, L. Schmitz, O. Schmitz, J. Schneider, P. Schroeder, K. Schultz, E. Schuster, J. Schwartz, F. Sciortino, F. Scotti, J. Scoville, A. Seltzman, S. Seol, I. Sfiligoi, M. Shafer, S. Sharapov, H. Shen, Z. Sheng, T. Shepard, S. Shi, Y. Shibata, G. Shin, D. Shiraki, R. Shousha, H. Si, P. Simmerling, G. Sinclair, J. Sinha, P. Sinha, G. Sips, T. Sizyuk, C. Skinner, A. Sladkomedova, T. Slendebroek, J. Slief, R. Smirnov, J. Smith, S. Smith, D. Smith, J. Snipes, G. Snoep, A. Snyder, P. Snyder, E. Solano, W. Solomon, J. Song, A. Sontag, V. Soukhanovskii, J. Spendlove, D. Spong, J. Squire, C. Srinivasan, W. Stacey, G. Staebler, L. Stagner, T. Stange, P. Stangeby, R. Stefan, R. Stemprok, D. Stephan, J. Stillerman, T. Stoltzfus-Dueck, W. Stonecipher, S. Storment, E. Strait, D. Su, L. Sugiyama, Y. Sun, P. Sun, Z. Sun, A. Sun, D. Sundstrom, C. Sung, J. Sungcoco, W. Suttrop, Y. Suzuki, T. Suzuki, A. Svyatkovskiy, C. Swee, R. Sweeney, C. Sweetnam, G. Szepesi, M. Takechi, T. Tala, K. Tanaka, X. Tang, S. Tang, Y. Tao, R. Tao, D. Taussig, T. Taylor, K. Teixeira, K. Teo, A. Theodorsen, D. Thomas, K. Thome, A. Thorman, A. Thornton, A. Ti, M. Tillack, N. Timchenko, R. Tinguely, R. Tompkins, J. Tooker, A. Torrezan De Sousa, G. Trevisan, S. Tripathi, A. Trujillo Ochoa, D. Truong, C. Tsui, F. Turco, A. Turnbull, M. Umansky, E. Unterberg, P. Vaezi, P. Vail, J. Valdez, W. Valkis, B. Van Compernolle, J. Van Galen, R. Van Kampen, M. Van Zeeland, G. Verdoolaege, N. Vianello, B. Victor, E. Viezzer, S. Vincena, M. Wade, F. Waelbroeck, J. Wai, T. Wakatsuki, M. Walker, G. Wallace, R. Waltz, W. Wampler, L. Wang, H. Wang, Y. Wang, H. Wang, Z. Wang, H. Wang, Z. Wang, Y. Wang, G. Wang, S. Ward, M. Watkins, J. Watkins, W. Wehner, Y. Wei, M. Weiland, D. Weisberg, A. Welander, A. White, R. White, S. Wiesen, R. Wilcox, T. Wilks, M. Willensdorfer, H. Wilson, A. Wingen, M. Wolde, M. Wolff, K. Woller, A. Wolz, H. Wong, S. Woodruff, M. Wu, Y. Wu, S. Wukitch, G. Wurden, W. Xiao, R. Xie, Z. Xing, X. Xu, C. Xu, G. Xu, Z. Yan, X. Yang, S. Yang, T. Yokoyama, R. Yoneda, M. Yoshida, K. You, T. Younkin, J. Yu, M. Yu, G. Yu, Q. Yuan, L. Zaidenberg, L. Zakharov, A. Zamengo, S. Zamperini, M. Zarnstorff, E. Zeger, K. Zeller, L. Zeng, M. Zerbini, L. Zhang, X. Zhang, R. Zhang, B. Zhang, J. Zhang, J. Zhang, L. Zhao, B. Zhao, Y. Zheng, L. Zheng, B. Zhu, J. Zhu, Y. Zhu, Y. Zhu, M. Zsutty, and M. Zuin. Diii-d research advancing the physics basis for optimizing the tokamak approach to fusion energy. Nuclear Fusion, 62(4):042024, apr 2022.

[17] J. Citrin, S. Breton, F. Felici, F. Imbeaux, T. Aniel, J. Artaud, B. Baiocchi, C. Bourdelle, Y. Camenen, and J. Garcia. Real-time capable first principle based modelling of tokamak turbulent transport. Nuclear Fusion, 55(9):092001, 2015.

[18] K. L. Plassche, J. Citrin, C. Bourdelle, Y. Camenen, F. J. Casson, V. I. Dagnelie, F. Felici, A. Ho, S. Van Mulders, and J. Contributors. Fast modeling of turbulent transport in fusion plasmas using neural networks. Physics of Plasmas, 27(2), 2020.

[19] E. Fransson, A. Gillgren, A. Ho, J. Borsander, O. Lindberg, W. Rieck, M. Åqvist, and P. Strand. A fast neural network surrogate model for the eigenvalues of qualikiz. Physics of Plasmas, 30(12), 2023.

[20] B. Settles. Active Learning, volume 6 of Synthesis Lectures on Artificial Intelligence and Machine Learning. Morgan & Claypool Publishers, 2012.

[21] A. Pavone, A. Merlo, S. Kwak, and J. Svensson. Machine learning and bayesian inference in nuclear fusion research: an overview. Plasma Physics and Controlled Fusion, 65(5):053001, 2023.

[22] J. Candy, E. A. Belli, and R. Bravenec. A high-accuracy eulerian gyrokinetic solver for collisional plasmas. Journal of Computational Physics, 324:73–93, 2016.

[23] T. Neiser, O. Meneghini, S. Smith, J. McClenaghan, T. Slendebroek, D. Orozco, B. Sammuli, G. Staebler, J. Hall, E. Belli, and J. Candy. Multi-fidelity neural network representation of gyrokinetic turbulence. In APS Division of Plasma Physics Meeting Abstracts, volume 2023 of APS Meeting Abstracts, page PP11.039, January 2023.

[24] T. F. Neiser, D. Sun, B. Agnew, T. Slendebroek, O. Meneghini, B. C. Lyons, A. G. Ghiozzi, J. T. McClenaghan, G. M. Staebler, and J. Candy. TJLF: The quasi-linear model of gyrokinetic transport TGLF translated to Julia. In Bulletin of the American Physical Society, 2024.

Disclaimer

A Complete Ablation Results

This section presents comprehensive offline training results for all combinations of contributions (TGLF-NN, TGLF-NN + FT, TGLF-NN + SR, TGLF-SINN) across both datasets (Major + Minor perturbation, Major perturbation), with and without filtering. We evaluated the performance of the model using the root mean squared logarithmic error (RMSLE, Equation 5) for each flux channel and the average R^2 in all channels.

The RMSLE is defined as:

$$\text{RMSLE}(\boldsymbol{y}', \boldsymbol{y}) = \sqrt{\frac{1}{N}\sum_{i=1}^{N} \|\ln(1 + |\boldsymbol{y}'_i|) - \ln(1 + |\boldsymbol{y}_i|)\|^2}, \tag{5}$$

where N is the number of test samples, $\boldsymbol{y}'_i$ represents the predicted fluxes, and $\boldsymbol{y}_i$ represents the ground truth fluxes.

The R^2 metric is defined in Equation (6), where N is the number of test samples, $\boldsymbol{y}'_i$ represents the predicted fluxes, $\boldsymbol{y}_i$ represents the ground truth fluxes and $\bar{\boldsymbol{y}}$ represents the mean of the ground truth fluxes test.

$$R^2(\boldsymbol{y}', \boldsymbol{y}) = 1 - \frac{\sum_{i=1}^{N} (\boldsymbol{y}_i - \boldsymbol{y}'_i)^2}{\sum_{i=1}^{N} (\boldsymbol{y}_i - \bar{\boldsymbol{y}})^2}, \tag{6}$$

Table 1: **Comprehensive ablation results across all experimental configurations.** Performance metrics are reported as Mean Squared Logarithmic Error (RMSLE) ($\times 10^{-2}$) on test sets, except for the final column which shows R^2 values. Best results are highlighted in **bold**. Abbreviations: FT = Feature Tuning, SR = Spectral Regularization, Ma. = Major perturbation, Mi. = Minor Perturbation, F = Unfiltered, T = Filtered.

Model	**Dataset**	**Filtered**	Γ_e	Π_i	Q_e	Q_i	**Avg.**	R^2
TGLF-NN	Ma.	F	24.41 ± 5.67	41.8 ± 4.18	51.96 ± 5.69	36.55 ± 4.91	39.93 ± 4.72	-8.80E+11
TGLF-NN	Ma.	T	6.28 ± 1.75	9.12 ± 1.46	12.06 ± 2.35	8.82 ± 2.06	9.3 ± 1.45	0.9221
TGLF-NN	Ma. + Mi.	F	32.2 ± 9.62	47.74 ± 10.66	57.03 ± 9.62	44.69 ± 10.62	46.28 ± 10.54	-1.84E+19
TGLF-NN	Ma. + Mi.	T	4.23 ± 1.08	6.54 ± 0.94	8.76 ± 1.6	6.33 ± 1.41	6.66 ± 0.9	**0.9685**
TGLF-NN + FT	Ma.	F	10.75 ± 1.99	13.36 ± 2.43	15.46 ± 2.8	13.24 ± 2.08	13.31 ± 2	0.1022
TGLF-NN + FT	Ma.	T	6.02 ± 1.66	8.54 ± 1.5	11.38 ± 2.31	8.11 ± 1.95	8.73 ± 1.42	0.9178
TGLF-NN + FT	Ma. + Mi.	F	7.71 ± 1.3	9.64 ± 1.62	11.49 ± 2.18	9.59 ± 1.46	9.7 ± 1.46	0.4343
TGLF-NN + FT	Ma. + Mi.	T	3.9 ± 0.98	5.8 ± 0.97	7.86 ± 1.61	5.49 ± 1.22	5.93 ± 0.87	0.9646
TGLF-NN + SR	Ma.	F	21.51 ± 3.64	28.71 ± 3.73	31.34 ± 4.01	27.5 ± 3.54	27.5 ± 3.33	-4.62E+07
TGLF-NN + SR	Ma.	T	6.22 ± 1.69	8.96 ± 1.51	11.82 ± 2.37	8.68 ± 2.03	9.14 ± 1.43	0.9147
TGLF-NN + SR	Ma. + Mi.	F	21.59 ± 3.63	28.54 ± 3.33	30.8 ± 3.48	26.73 ± 3.14	27.13 ± 3.04	-1.37E+11
TGLF-NN + SR	Ma. + Mi.	T	4.16 ± 1.08	6.25 ± 0.89	8.42 ± 1.58	6 ± 1.32	6.39 ± 0.87	0.9679
TGLF-SINN	Ma.	F	9.97 ± 2	12.92 ± 2.97	14.53 ± 2.93	12.43 ± 2.21	12.57 ± 2.18	0.3760
TGLF-SINN	Ma.	T	5.93 ± 1.71	8.47 ± 1.56	11.4 ± 2.47	7.92 ± 2.03	8.66 ± 1.5	0.9016
TGLF-SINN	Ma. + Mi.	F	6.62 ± 1.32	8.33 ± 1.68	9.98 ± 2.09	8.18 ± 1.49	8.36 ± 1.43	0.6482
TGLF-SINN	Ma. + Mi.	T	**3.85±0.99**	**5.68±0.92**	**7.75±1.55**	**5.35±1.18**	**5.83±0.84**	0.9639

Table 1 presents a complete ablation of all variation of combinations (model components, dataset size, data filtering). Performance metrics are reported as RMSLE ($\times 10^{-2}$) for each flux channel (Γ_e, Π_i, Q_e, Q_i), along with the average RMSLE across all channels. The final column shows the average value of R^2 calculated on all channels.

B Hyperparameter Search Results

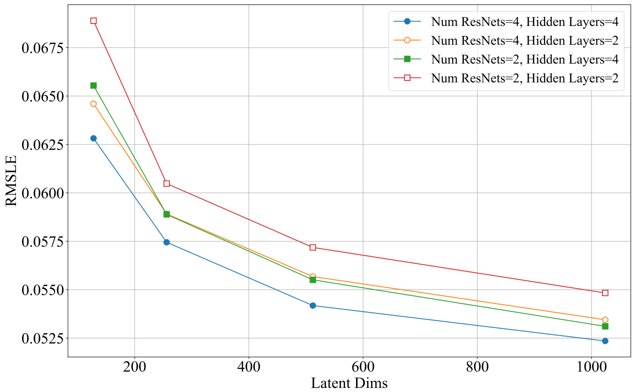


Figure 3: Impact of latent dimensions on model performance. The graph shows the test RMSLE of TGLF-SINN on the filtered Major + Minor perturbation dataset as a function of the number of latent dimensions, with all other hyperparameters held constant. Lower RMSLE values indicate superior performance.

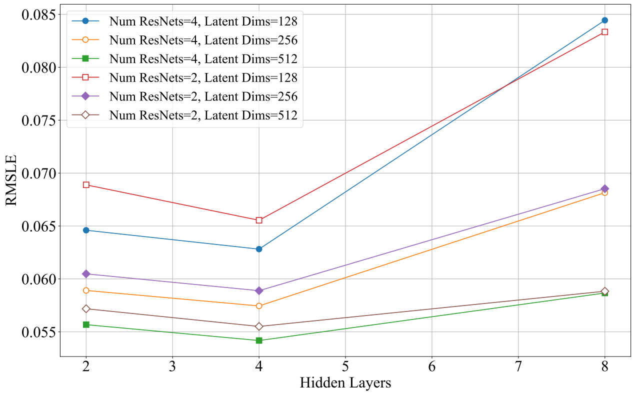


Figure 4: Effect of hidden layer depth on model performance. The graph illustrates the test RMSLE of TGLF-SINN on the filtered Major + Minor perturbation dataset as a function of the number of hidden layers, with all other hyperparameters held constant. Lower RMSLE values indicate superior performance.

We conducted extensive hyperparameter searches for TGLF-SINN using the filtered Major + Minor perturbation dataset. We noted diminishing marginal benefits when the latent dimension exceeded 256 and the number of ResNet blocks exceeded 4 (Figures 3 and 5). We also observed that the minimal loss was achieved with 4 hidden layers in the MLP (Figure 4). Collectively, these results guided our choice of hyperparameters in the implementation, balancing accuracy and computational cost. We choose 4 total ResNet blocks, 4 hidden layers per ResNet block, and 256 neurons per hidden layer. These hyperparameters remain consistent for all L-models, including TGLF-NN, as to eliminate the variable of parameter count.

C Implementation Details

C.1 Filtering Outliers

Some input combinations and output solutions are rarely encountered in real experiments (i.e., a modeling gap exists between quasilinear theory and reality). Additionally, the numerical solver, particularly the approximate eigenvalue solver using Hermite basis functions, can converge to sub-optimal solutions due to truncation or numerical errors. As a result, the generated dataset inherently

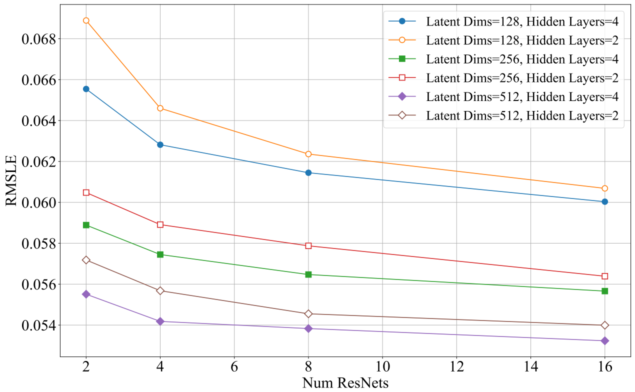


Figure 5: Influence of ResNet quantity on model performance. The graph depicts the test RMSLE of TGLF-SINN on the filtered Major + Minor perturbation dataset as a function of the number of ResNets, with all other hyperparameters held constant. Lower RMSLE values indicate superior performance.

contains noise or outliers that can destabilize training. To address this, we iteratively apply a Median Absolute Deviation (MAD) filtering approach. Specifically, we compute:

$$\boldsymbol{m} = \mathrm{med}\left(\{|\boldsymbol{x} - \tilde{\boldsymbol{x}}|, \quad \text{for } \boldsymbol{x} \in \mathbb{X}\}\right), \tag{7}$$

where $\tilde{\boldsymbol{x}} = \mathrm{med}(\mathbb{X})$ is the median of the dataset $\mathbb{X}$, and $\boldsymbol{m}$ is the median absolute deviation (MAD). We then define a distance for each data point $\boldsymbol{x}$ as:

$$\boldsymbol{d} = \kappa \frac{|\boldsymbol{x} - \tilde{\boldsymbol{x}}|}{\boldsymbol{m}}, \tag{8}$$

where $\kappa = 1.4826$ is a scaling factor that approximates the standard deviation under normality assumptions. In each iteration, any sample with $\boldsymbol{d} > \epsilon$ in any channel is removed, where $\epsilon = 200$ is a threshold hyperparameter. The iteration terminates when fewer than 1% of the data points are removed in a single iteration.

D Mean and Standard Deviation of Inputs Across Radial Locations

To obtain the statistics of input parameters for each radial location, we parsed the metadata of the shot data and grouped each shot into its corresponding radial location bracket. After parsing the entire dataset, we calculated the mean and standard deviation for each radial bracket. Figure 6 plots these statistics for each input variable against the radial profile.

E Full flux-matching experimental results

This section contains detailed flux-matching validation results on experimental DIII-D data, including both L-mode and H-mode scenarios across 4500 experimental cases. Due to complex figure formatting, the complete figures and analysis are available in the extended version of this paper.

Experimental Setup. To assess the practicality of TGLF-SINN, we integrate it into the flux matching workflow of the FUSE integrated modeling suite [3], which couples the core turbulence, the pedestal (assumed to be fixed as experimental measurements) and the neoclassical transport components. Profiles (T_e, T_i, n_e, ω_0), (electron temperature, ion temperature, electron density, and toroidal rotation frequency, respectively), are reconstructed at 16 uniformly spaced radial locations $\rho \in [0.1, 0.85]$. We compare three core transport options: 1) numerical TJLF solver [24], 2) TGLF-NN, and 3) TGLF-SINN using two representative DIII-D cases: L-mode discharge (shot #161582 at time 2300 ms) and H-mode discharge (shot #157955 at time 2700 ms).

Convergence Performance. Both NN surrogates exhibit faster and more stable convergence than the numerical solver. As shown in Figure 7, TGLF-SINN converges in 129 iterations for the L-mode and 412 iterations for H-mode, while TGLF-NN converges in 251 and 364 iterations, respectively. The numerical solver fails to converge after 1000 iterations in both scenarios.

Examining individual transport channels reveals that the numerical solver exhibits poor convergence in the ω_0 channel for the L-mode and struggles across all channels in the H-mode. We attribute the superior NN performance to its smoother function approximation compared to discrete operations in the numerical solver, resulting in reduced oscillations during gradient-based optimization.

Reconstruction Quality. As shown in Figure 8, TGLF-SINN and TGLF-NN consistently produce lower final mismatch values between predicted and experimental transport fluxes in both L-mode and H-mode scenarios, while TJLF exhibits significantly higher flux mismatches. The reconstructed core profiles (Figure 9) show good agreement with the experimental measurements across all channels, with deviations observed in low ρ regions (near the magnetic axis) for T_e in L-mode and both T_e and T_i in H-mode, consistent with higher experimental uncertainty near the magnetic axis.

Computational Efficiency Gain. Due to the faster inference time and fewer iterations required for convergence, the flux-matching procedure with TGLF-SINN takes 20 seconds compared to 15 minutes with the numerical solver using 128 threads, representing a 45× speed-up. TGLF-NN achieves a 600× speed-up (1.5 seconds), likely due to its native Julia implementation being more optimized for this workflow. Nevertheless, the 45× speed-up of TGLF-SINN demonstrates substantial efficiency advantages, and its translation into a Julia implementation could enable further speed-up.

Extended Dataset Validation. We extended validation to 3000 H-mode and 1500 L-mode DIII-D profiles. Figure 12a shows significant performance improvements in L-mode while H-mode performance remains comparable to TGLF-NN. Figure 12b shows that both neural surrogate models achieve 10 percentage point improvement in stored thermal energy predictions compared to the IPB98(y,2) scaling law at a typical $\rho_{\max}$ of 0.8.

The comparison of predicted and measured profiles shows similar performance between TGLF-NN and TGLF-SINN (Figure 10). Both H-modes exhibit slight underprediction of temperature and rotation in H-modes and systematic overprediction in L-mode core regions ($\rho < 0.5$). These systematic biases are observed in both surrogates, suggesting limitations in the original TGLF ROM rather than surrogate models. This may require additional CGYRO simulations to recalibrate TGLF's saturation rules, which is beyond the scope of this work.

The convergence error analysis (Figure 11) shows TGLF-NN consistently achieves smaller errors, possibly due to ensemble averaging (20 networks) versus our single network approach. Convergence errors remain stable until nonlinear increases above $\rho_{\max} = 0.85$ for H-modes and $\rho_{\max} = 0.9$ for L-modes, suggesting these as empirical upper limits for flux-matching applications.

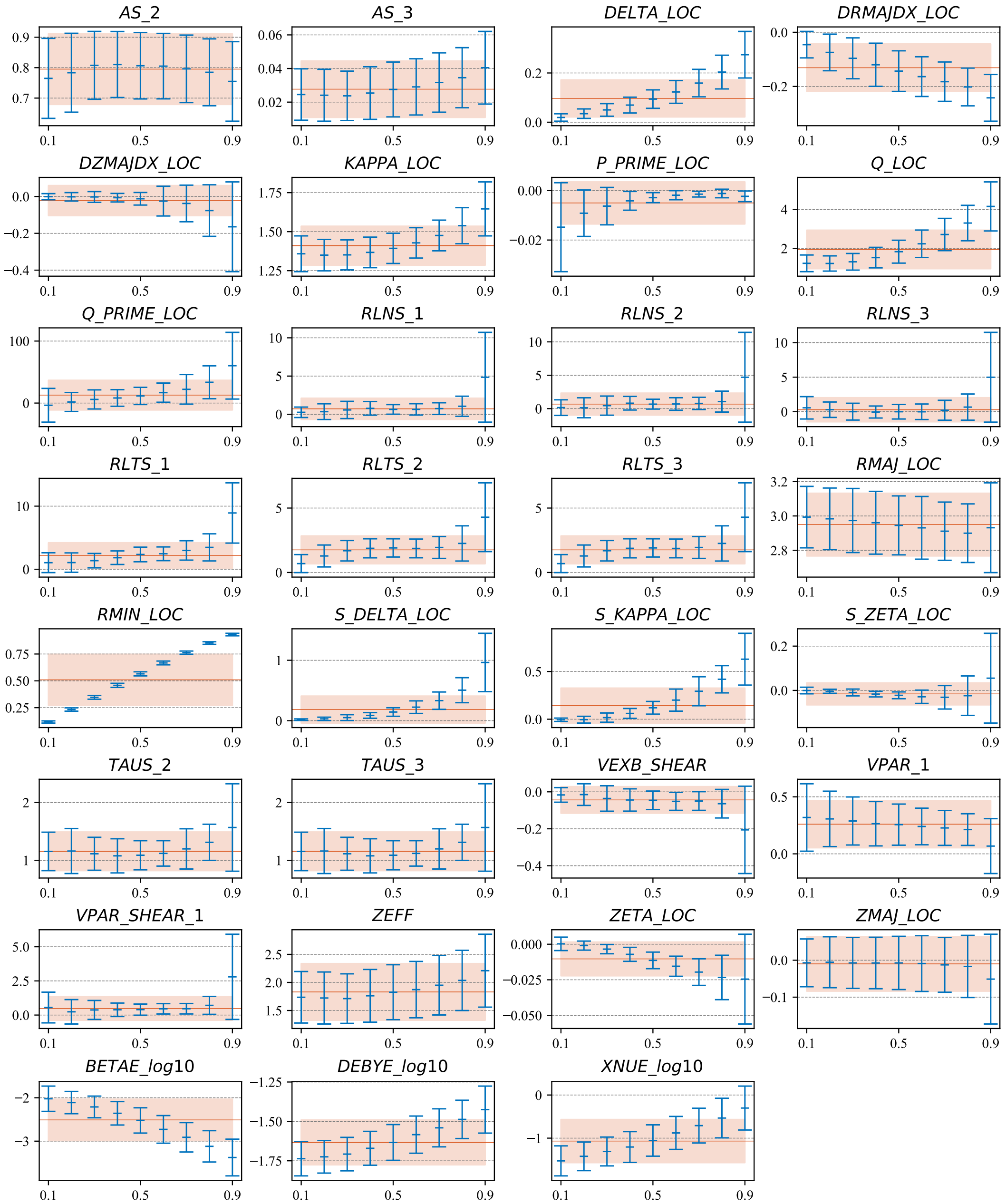
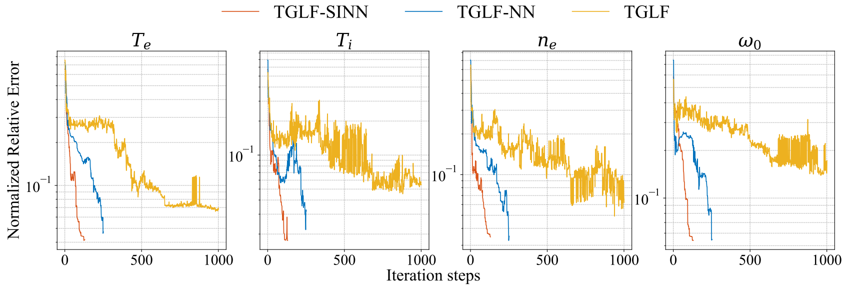
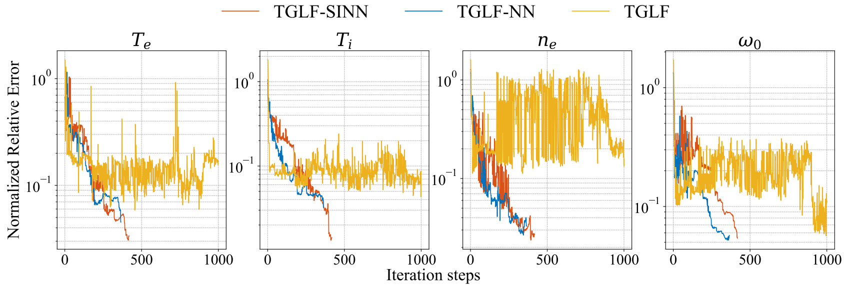


Figure 6: Radial profiles of mean values and standard deviations for input parameters. Each subplot represents a distinct input parameter, with the x-axis showing the normalized radial coordinate ρ and the y-axis displaying the parameter's value. Blue bars indicate mean values, while orange bars represent standard deviations at each radial location.

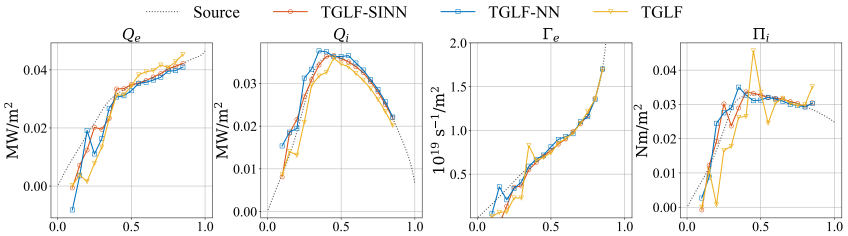


(a) L-mode convergence

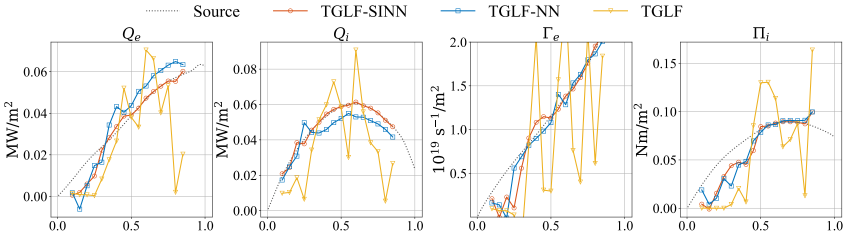


(b) H-mode convergence

Figure 7: **Convergence of flux-matching in L-mode and H-mode scenarios**. Comparison of normalized flux mismatch across iteration steps for TGLF, TGLF-NN, and TGLF-SINN. (a) L-mode: TGLF-SINN converges after 129 iterations, TGLF-NN after 251 iterations, while TGLF fails to converge after 1000 iterations. (b) H-mode: TGLF-SINN converges after 412 iterations, TGLF-NN after 364 iterations, while TGLF fails to converge after 1000 iterations.

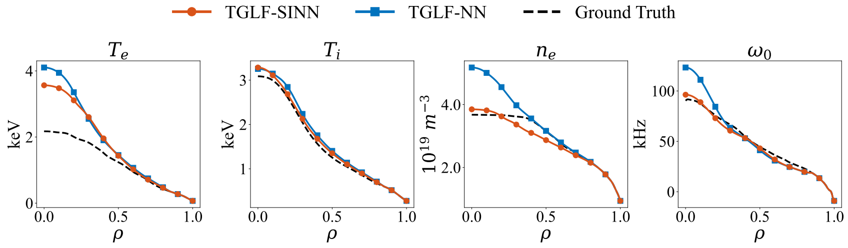


(a) L-mode flux profiles

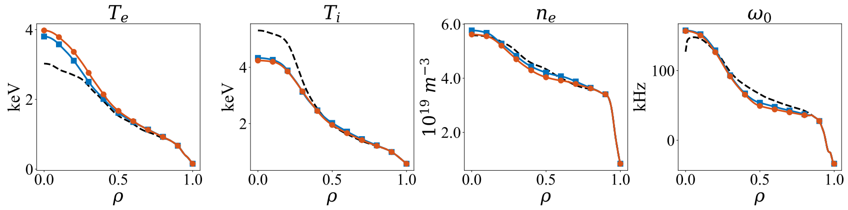


(b) H-mode flux profiles

Figure 8: **Reconstructed flux profiles in L-mode and H-mode scenarios**. Comparison of TGLF-SINN, TGLF-NN, and TGLF for predicted transport fluxes (Q_e, Q_i, Γ_e, Π) over 16 ρ locations. (a) L-mode reconstructed fluxes. (b) H-mode reconstructed fluxes. TGLF profiles are omitted due to non-convergence in both scenarios.

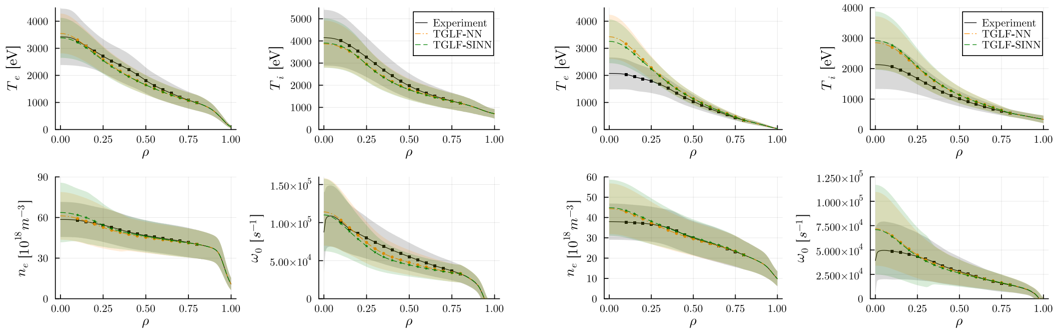


(a) L-mode core profiles



(b) H-mode core profiles

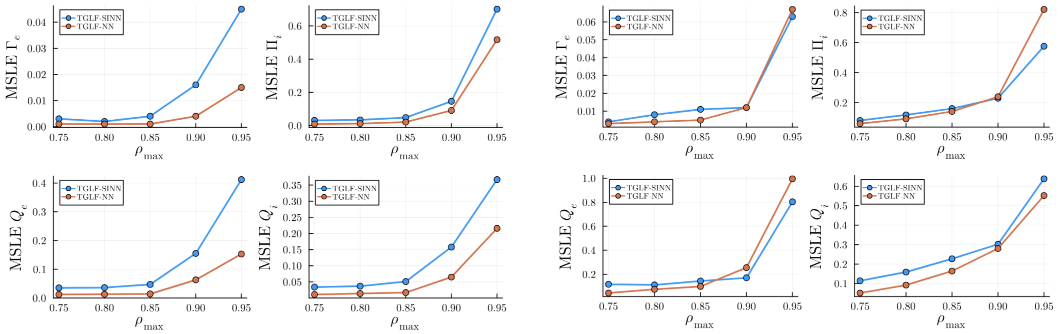
Figure 9: **Reconstructed core profiles in L-mode and H-mode scenarios**. Comparison of electron temperature, ion temperature, electron density, and toroidal rotation frequency profiles for experimental measurements, TGLF-NN, and TGLF-SINN. (a) L-mode scenario profiles. (b) H-mode scenario profiles. TGLF is omitted due to non-convergence in both scenarios.



(a) Profiles for 3000 DIII-D H-mode plasmas.

(b) Profiles for 1500 DIII-D L-mode plasmas.

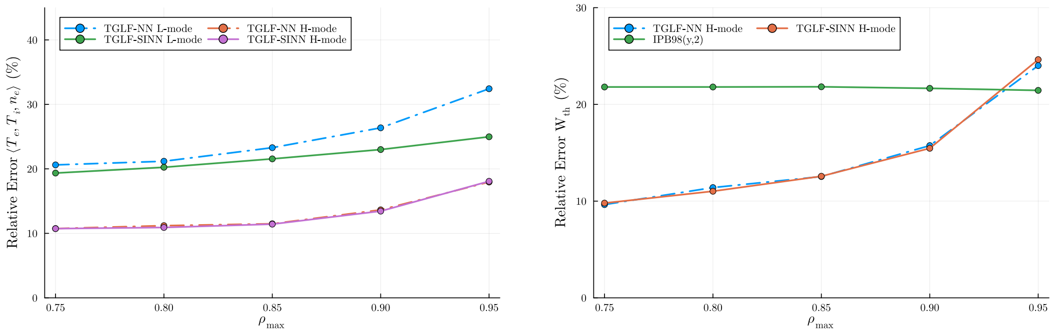
Figure 10: Comparison of measured profiles to predictions by TGLF-NN and TGLF-SINN for (a) 3000 H-modes and (b) 1500 L-modes with $\rho_{\max} = 0.80$ (solid line: mean, shaded: variance). The mean and variance of the predicted profiles are comparable, demonstrating similar performance despite the added complexity of spectra informed training.



(a) Convergence error for 3000 DIII-D H-mode plasmas.

(b) Convergence error for 1500 DIII-D L-mode plasmas.

Figure 11: Comparison of MSLE convergence errors of TGLF-NN and TGLF-SINN for (a) 3000 H-modes and (b) 1500 L-modes. The convergence errors show a nonlinear increase above $\rho_{\max} = 0.85$ for H-modes and above $\rho_{\max} = 0.9$ for L-modes, suggesting these as empirical upper limits for the ρ-range used for turbulent flux-matching.



(a) Relative error across varying radius

(b) Stored thermal energy predictions

Figure 12: **Extended dataset validation results**. (a) Relative error comparison between TGLF-SINN and TGLF-NN for reconstructed profiles in H-mode and L-mode across the DIII-D dataset. Metrics are reported as mean absolute relative error for each reconstructed profile (T_e, T_i, n_e) as a percentage. (b) Comparison of stored thermal energy predictions. TGLF-NN is more accurate by 10 percentage points compared to the IPB98(y,2) scaling law at typical $\rho_{\max}$ of 0.8, validating flux-matched predictions over empirical scaling laws.