9^{th} Asia-Pacific Conference on Plasma Physics, 21-26 Sep, 2025 at Fukuoka

Improvement of turbulent transport model using multi-fidelity data fusion approach

S. Maeyama
National Institute for Fusion Science
e-mail (speaker): maeyama.shinya@nifs.ac.jp

Understanding and predicting turbulent transport, which critically determines plasma confinement performance in fusion devices, remains an important research challenge. Various regression models, such as those based on neural networks trained with gyrokinetic simulation databases or experimental data, have been explored in previous studies. In contrast, our approach treats multiple datasets—each differing in fidelity, size, and parameter coverage—as multi-fidelity data, and we propose applying multi-fidelity data fusion techniques to model turbulent transport. Multi-fidelity modeling is a methodology that combines a large number of low-fidelity data (which are less accurate) with a small number of high-fidelity data (which are more accurate) to improve overall prediction accuracy. We have applied nonlinear auto-regressive Gaussian process (NARGP) regression algorithm [1] to several plasma turbulence modeling problems and demonstrated that combining theory, simulation, and experimental data improves accuracy [2].

In this study, we further develop this approach to enhance prediction accuracy in extrapolative settings. In Gaussian Process (GP) regression, the choice of kernel function significantly affects the expressiveness of the model. In particular, for extrapolative tasks, it is essential to design general-purpose kernels that can simultaneously capture short-term fluctuations and long-term trends by representing multiple length scales. To that end, we introduce and compare general-purpose kernels such as spectral mixture kernels and neural kernel networks [3]. We then present one-dimensional and two-dimensional test function problems that illustrate how NARGP can improve extrapolation performance.

Finally, we apply our method to a more realistic problem: a regression task using a plasma turbulence transport dataset where local plasma parameters serve as inputs, and the experimentally observed particle diffusivity is the output. As an extrapolative scenario, we train the model using only data with large particle diffusivity and attempt to predict unseen data with small diffusivity. Figure (a) shows the result of single-fidelity GP regression using a simple radial basis kernel. While the model fits the training data well, its predictive performance for the test data is poor—it tends to predict large diffusivity values similar to the training data. In contrast, Figure (b) presents the result from multi-fidelity regression, where the use of a general-purpose kernel capable of capturing long-term trends and the incorporation of low-fidelity data (e.g., linear growth rates and wavenumbers) enable the model to predict smaller diffusivity values in the test region more accurately. We also analyze the prediction error using the Mahalanobis distance [4]. This analysis suggests that the points with large errors are located outside the distribution of the training data. These points are difficult to predict because the model has not encountered similar data during training.

In summary, our results demonstrate that multi-fidelity modeling is effective in improving the extrapolative performance of turbulent transport models. This approach provides a promising path toward unifying theoretical, numerical, and experimental insights into a single predictive framework.

Acknowledgements
This work was supported by JST, PRESTO Grant Number JPMJPR23OB, Japan.

References

[1] P. Perdikaris, et al., Proc. R. Soc. A 473, 20160751 (2017).
[2] S. Maeyama, et al., Sci. Rep. 14, 28242 (2024).
[3] S. Sun, et al., Proc. 35th Int. Conf. on Mach. Learn. 80, 4828 (2018).
[4] P. C. Mahalanobis, Proc. Nat. Inst. Sci. India 2, 49 (1936).

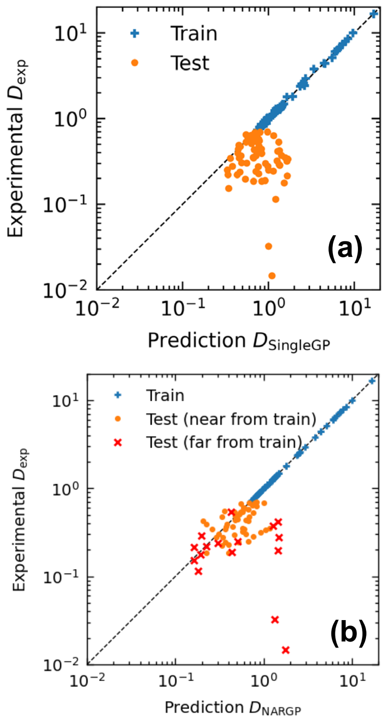

Figure. Comparison of prediction and actual values of plasma turbulent diffusion coefficients in (a) conventional GP regression only using high-fidelity data, and in (b) NARGP regression using multi-fidelity data.