CONFERENCE PRE-PRINT

EXTRAPOLATIVE PREDICTABILITY OF PLASMA TURBULENT TRANSPORT VIA A MULTI-FIDELITY DATA FUSION APPROACH

S. Maeyama
National Institute for Fusion Science
Toki, Japan
Email: maeyama.shinya@nifs.ac.jp

M. Honda, E. Narita
Graduate School of Engineering, Kyoto University
Kyoto, Japan

F. Miwakeichi, K. Yano, A. Okuno
The Institute of Statistical Mathematics,
Tokyo, Japan

Abstract

Predicting turbulent transport in fusion plasmas is essential for forecasting the performance of next-generation fusion reactors. This study presents a multi-fidelity modeling framework based on nonlinear autoregressive Gaussian process (NARGP) regression, combining high-fidelity experimental data with low-fidelity simulation results. By integrating a Neural Kernel Network (NKN) into the NARGP framework, our model captures complex nonlinear correlations across fidelity levels and achieves significantly improved extrapolative performance. The approach is validated on both synthetic benchmark problems and a real-world dataset from the Joint European Torus (JET), demonstrating its robustness and practical applicability in plasma transport modeling.

1. INTRODUCTION

Understanding and predicting turbulent transport remains a central challenge in magnetic confinement fusion research. In magnetically confined plasmas, microscopic turbulence dominates transport processes and directly impacts the plasma confinement performance. Therefore, the ability to quantitatively predict turbulent transport is essential for optimizing the design and operational scenarios of next-generation fusion devices. First-principles simulations based on gyrokinetic theory have been widely used as high-fidelity tools for capturing the physics of plasma turbulence [1]. While these simulations offer high accuracy, they require substantial computational resources, making it difficult to explore wide plasma parameter spaces comprehensively. In addition, although several validation studies based on first-principles simulations well explain the experimental results, there are sometimes quantitative discrepancies between simulation predictions and experimental observations. As alternatives, theory-based quasilinear models or machine learning surrogate models based on experimental data have also been actively investigated. Models grounded in experimental measurements are valuable because they reflect the actual behavior of operational devices. However, the availability of the training dataset is, of course, limited to the plasma parameter range within existing devices.

In practice, transport predictions are most urgently needed for unexplored regions, such as those corresponding to future devices or high-performance regimes that have not yet been achieved experimentally. In these settings, the model's extrapolative performance becomes critically important. However, existing models face significant challenges: the experiment-based surrogate model tends to exhibit poor generalization outside the training data. In contrast, the simulation-based model can be quantitatively insufficient, despite having the inherent advantage of theory-based extrapolation.

To address these challenges, this study introduces a multi-fidelity data fusion framework for turbulent transport modeling. In this approach, low-fidelity data are abundant but less accurate, while high-fidelity data are accurate but sparse or only applicable to a limited parameter range. By jointly leveraging the complementary strengths of both and learning the correlations across fidelity levels, the proposed method aims to improve predictive performance in extrapolative regimes significantly. This work builds on previous demonstrations of multi-fidelity modeling for turbulent transport [1] and extends the methodology to target improved extrapolative capability.

2. METHODS

2.1. Gaussian process regression

Gaussian process (GP) regression [2] is a nonparametric Bayesian inference method that estimates the relationship between input variables $X = \{\boldsymbol{x}_0, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_{n-1}\}$ and output variables $Y = \{y_0, y_1, \ldots, y_{n-1}\}$. In a GP regression framework, the output function $y = f(\boldsymbol{x})$ is typically assumed to follow a zero-mean GP prior $f \sim \mathcal{GP}(f \mid 0, k(\boldsymbol{x}, \boldsymbol{x}'))$ with a kernel (covariance) function $k(\boldsymbol{x}, \boldsymbol{x}')$. Under a given dataset X and Y, the posterior distribution of the response $f_* = f(\boldsymbol{x}_*)$ at a new input point $\boldsymbol{x}_*$ is given by a normal distribution. Their posterior mean μ_* and posterior variance σ_*^2 are given by

$$\mu_*(\boldsymbol{x}_*) = \boldsymbol{k}_*^T \cdot \widehat{K}^{-1} \cdot \boldsymbol{\eta}, \tag{1}$$

$$\sigma_*^2(\boldsymbol{x}_*) = k(\boldsymbol{x}_*, \boldsymbol{x}_*) - \boldsymbol{k}_*^T \cdot \widehat{K}^{-1} \cdot \boldsymbol{k}_*, \tag{2}$$

where $k_{*i} = k(\boldsymbol{x}_*, \boldsymbol{x}_i)$, $K_{ij} = k(\boldsymbol{x}_i, \boldsymbol{x}_j) + \sigma^2\delta_{ij}$ and $\eta_i = y_i$. In this way, GP regression models the correlation between inputs through the kernel function and simultaneously provides uncertainty estimates of the prediction within a Bayesian framework.

2.2. Nonlinear auto-regressive Gaussian process regression

Nonlinear auto-regressive Gaussian process (NARGP) regression [3] is a multi-fidelity method designed to improve predictive performance by integrating multiple datasets $(X^{(t)}, Y^{(t)})$ with hierarchically different levels of fidelity ($t = 0,1, \ldots$), although a standard GP regression treats the relationship between a single dataset of input X and output Y. The NARGP modeling procedure begins with applying a standard GP regression on the lowest-fidelity dataset, from which the posterior prediction of the lowest-fidelity function $f_0(\boldsymbol{x})$ is obtained. Then, for the higher-fidelity data, a new GP is constructed as a function of the extended input space consisting of the original input $\boldsymbol{x}$ and the posterior prediction from the one-order lower-fidelity GP. Repeating this procedure enables the model to integrate multi-fidelity information in an auto-regressive structure recursively. Specifically, for the t-th fidelity function,

$$f_t(\boldsymbol{x}) = g_t\big(\boldsymbol{x}, f_{t-1}(\boldsymbol{x})\big), \tag{3}$$

$$g_t \sim \mathcal{GP}\left(f_t \mid 0, k_t\left(\big(\boldsymbol{x}, f_{t-1}(\boldsymbol{x})\big), \big(\boldsymbol{x}', f_{t-1}(\boldsymbol{x}')\big)\right)\right), \tag{4}$$

where k_t is the kernel function at the t-th fidelity dataset, defined in the extended input space $(\boldsymbol{x}, f_{t-1})$. In this work, we employ a separated-variable form of the kernel function proposed in Ref. [3],

$$k_t = k_{t\rho}(\boldsymbol{x}, \boldsymbol{x}')k_{tf}\big(f_{t-1}(\boldsymbol{x}), f_{t-1}(\boldsymbol{x}')\big) + k_{tw}(\boldsymbol{x}, \boldsymbol{x}'), \tag{5}$$

where $k_{t\rho}$, k_{tf}, and k_{tw} are kernel functions having different hyperparameters.

2.3. Kernel functions

It is well known that the performance of GP regression is highly sensitive to the choice of kernel functions. The design of the kernel function essentially defines the prior assumptions about the types of functions the model considers plausible, directly specifying properties such as smoothness, linearity, and periodicity. Therefore, selecting an appropriate kernel structure is fundamentally important for enhancing both the interpretability and predictability of GP models. Since the central importance of the present study is an extrapolative prediction, it is desirable to use a kernel capable of simultaneously representing both short-term variations and long-term trends [2]. Accordingly, automatic learning or flexible construction of kernel structures has been explored [4,5].

In this study, we consider and compare the following types of kernel functions:

(a) Radial basis function (RBF) kernel

The RBF kernel is one of the most widely used kernels,

$$k(\boldsymbol{x}, \boldsymbol{x}') = \sigma_r^2 \exp\left[-\frac{1}{2}\sum_{i=0}^{d-1} w_i(x_i - x_i')^2\right], \tag{6}$$

where σ_r^2 represents the variance parameter, and w_i corresponds to the relevance (inverse length scale) in each direction of the d-dimension input vector $\boldsymbol{x}$.

(b) Spectral mixture (SM) kernel

The SM kernel is constructed based on a Gaussian mixture model in the frequency domain and is theoretically capable of approximating any stationary kernel [6,7].

$$k(\boldsymbol{x}, \boldsymbol{x}') = \sum_{q=0}^{Q-1} w_q \cos\left(2\pi\boldsymbol{\mu}_q \cdot (\boldsymbol{x} - \boldsymbol{x}')\right) \exp\left(-2\pi^2 \sum_{i=0}^{d-1} v_{q,i}(x_i - x_i')^2\right), \tag{7}$$

where w_q, $\boldsymbol{\mu}_q$, and $\boldsymbol{v}_q$ denote the weight, frequency, and mean relevance of each spectral component. Increasing the number of mixture components Q allows for approximation of more complex structures. We set $Q = 20$ in the following analysis.

(c) Neural kernel network (NKN)

NKN is a framework that composes kernel functions using a neural network architecture [8]. In NKN, the first layer defines a set of basic kernels, such as RBF, linear, or periodic kernels. These kernels are then composed using a combination of additive and multiplicative operations, respectively, in linear and product layers. Each layer's output remains a valid kernel, ensuring the overall model retains the properties required for kernel-based learning. This architecture enables the automatic learning of rich kernel structures directly from data. In the following analysis, we employ eight primitive kernels (two RBF, two linear, two periodic, and two rational quadratic functions, which are instantiated with different initial hyperparameters) as the first layer, and construct the subsequent architectures by stacking three linear layers and two product layers alternately.

3. RESULTS

3.1. Importance of kernel design in single-fidelity Gaussian process regression

We first applied GP regression to a single-fidelity dataset to examine how kernel design affects extrapolation performance. As a benchmark dataset, we used the time series of atmospheric CO_2 concentrations measured at Mauna Loa, which is publicly available in the OpenML repository [9]. This dataset is widely used as a canonical example because it exhibits multiple temporal scales of short-term seasonal fluctuations and a long-term trend.

Figure 1(a) shows the regression results using an RBF kernel with a long correlation length. While the long-term increasing trend is well captured, the model fails to reproduce the short-term seasonal oscillations. In this case, the GP model absorbs the short-term variations into the noise term and thus does not effectively utilize the available information for prediction. Conversely, when using an RBF kernel with an artificially short correlation length, as shown in Fig. 1(b), the model can capture the short-term oscillations but fails to represent the long-term trend in the extrapolated region. These results highlight a fundamental limitation of single-scale kernels: they are unable to simultaneously model both short- and long-term structures, which poses a critical challenge for extrapolation tasks.

To overcome this limitation, we tested general-purpose kernels such as the SM kernel and NKN, both of which can represent multiple correlation scales. These kernels successfully captured both the short-term oscillations and the long-term trend, thereby significantly improving extrapolation performance. In our numerical experiments, the SM kernel demonstrated sufficient expressive power; however, it was also sensitive to initialization and prone to convergence issues, particularly in capturing the long-term trend [Fig. 1(c)]. In contrast, the NKN model demonstrated more robust training behavior and consistently reproduced both short- and long-term features across extrapolative test data [Fig. 1(d)]. Based on these findings, we adopt the NKN as the primary general-purpose kernel in the following sections and evaluate its effectiveness through comparisons with the baseline RBF kernel.

These results clarify the importance of kernel design in extrapolative regression tasks. Complex and flexible kernels that can express multiple correlation scales are fundamentally important for extrapolation prediction in GP regressions.

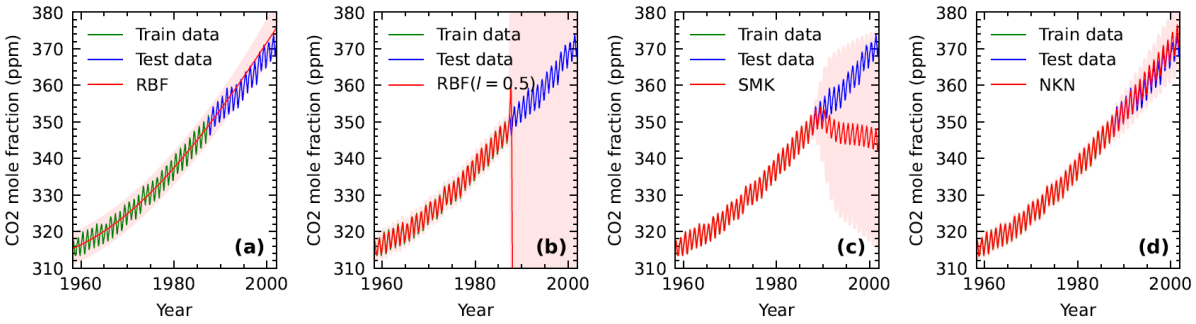

FIG. 1. Extrapolative prediction of the Mauna Loa CO2 dataset by using single-fidelity GP regression with different kernels: (a) RBF kernel with the length scale $l = w^{-1} = 107.2$, (b) RBF kernel with the length scale $l = 0.5$, (c) SM kernel, and (d) NKN kernel.

3.2. Multi-fidelity extrapolative prediction of a one-dimensional test function

To evaluate the effectiveness of multi-fidelity regression, we compare the conventional single-fidelity GP regression and multi-fidelity NARGP regressions in a one-dimensional test function problem. Here, we define the low-fidelity function $f_l(x)$ by the Airy function and the high-fidelity function $f_h(x)$ as a nonlinear transform of it:

$$f_l(x) = 2\mathrm{Ai}(-2\pi x), \tag{8}$$

$$f_h(x) = \left(x - \sqrt{2}\right) f_l^2(x). \tag{9}$$

For the training dataset, we employed 100 equidistant sample points of $f_l(x)$ over the entire input domain $0 \leq x < 2$, providing a sufficient representation of the low-fidelity function profile. In contrast, the high-fidelity data consists of only 15 randomly sampled points of $f_h(x)$, and no data is available in the region $x > 1.25$, which therefore constitutes an extrapolation problem for the high-fidelity function.

We first performed the conventional single-fidelity GP regression using only the high-fidelity data with an RBF kernel, as shown in Fig. 2(a). While the model provides a good fit within the training domain, its predictive accuracy in the extrapolated region $x > 1.25$ significantly degrades, making reliable estimation difficult. Next, we applied an NKN as the kernel of the single-fidelity GP regression, as shown in Fig. 2(b). NKN constructs a composite kernel from a set of base kernels, thereby offering a certain degree of expressiveness. However, the high-fidelity function $f_h(x)$ exhibits non-stationary oscillations with gradually varying frequency, which makes it difficult for the NKN to represent the function structure effectively. While the model still provides a reasonable fit within the training range, it does not improve extrapolation performance outside of this range. This result highlights that increasing kernel expressiveness alone is insufficient to guarantee extrapolation accuracy.

In contrast, the application of NARGP leads to significantly improved predictions even in the extrapolated region, as shown in Fig. 2(c). This is achieved by modeling the high-fidelity function in the extended input space, $f_h(x) = g_h(x, f_l(x))$, effectively leveraging the correlation structure between fidelity levels. As shown in the figure, the high-fidelity function in terms of the input parameter x exhibits the complex oscillation $f_h(x)$. In the NARGP expression, such an oscillatory behavior is primarily captured through the oscillatory structure of the low-fidelity function $f_l(x)$, whereas the high-fidelity function in the extended input space, $g_h(x, f_l)$, is relatively smooth. Thanks to the smoothness of the function g_h with respect to x, NARGP makes the extrapolation plausible.

Through this transformation, NARGP simplifies the underlying functional dependence of the target quantity. We here refer to this effect as "dependence simplification." This one-dimensional example demonstrates that NARGP enables high-accuracy predictions in extrapolated regions by exploiting the correlation between low-fidelity and high-fidelity datasets, and hopefully reduces the complexity of the regression problem through an appropriate embedding in the extended input space.

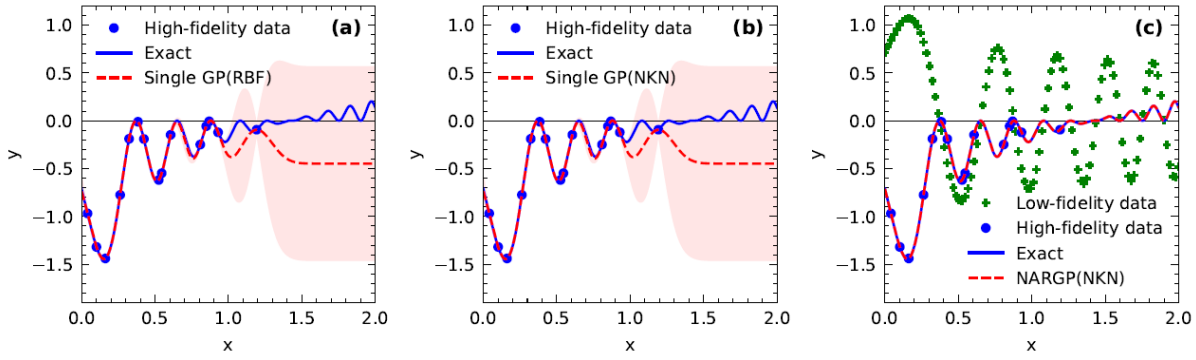

FIG. 2. Extrapolative prediction of a one-dimensional problem. (a) Single-fidelity GP regression with an RBF kernel, (b) Single-fidelity GP regression with an NKN kernel, and (c) NARGP with an NKN kernel.

3.3. Multi-fidelity extrapolative prediction of a two-dimensional test function

We next examine an extrapolation scenario using a two-dimensional test function. In this case, the high-fidelity function $f_h(x_0, x_1)$ is defined as an axisymmetric Bessel function, while the low-fidelity function $f_l(x_0, x_1)$ is its envelope approximation,

$$f_l(x_0, x_1) = \frac{1 - r^2/4}{1 + r^4} + \sqrt{\frac{2}{\pi |r|}} \frac{r^4}{1 + r^4}, \tag{10}$$

$$f_h(x_0, x_1) = J_0(r), \tag{11}$$

where J_0 is the zeroth-order Bessel function and $r = \sqrt{x_0^2 + x_1^2}$. Sampling points for training data is shown in Fig. 3 (a). The high-fidelity data is restricted to the region $x_0 < 0$, and there are no data points available for $x_0 > 0$. Therefore, any prediction in the region $x_0 > 0$ corresponds to an extrapolation problem with regard to the high-fidelity function. However, in the multi-fidelity framework, we assume that the low-fidelity function $f_l(x_0, x_1)$ is available over a whole range of $x_0 < 0$ and $x_0 > 0$.

First, we performed conventional single-fidelity GP regression with an RBF kernel using only the high-fidelity data in Fig. 3(b). As anticipated, the prediction error in the extrapolation region $x_0 > 0$ is significantly large, and the model fails to provide accurate results. Furthermore, as shown in Fig. 3(e), even when using an NKN kernel, the single-fidelity GP prediction in the extrapolated region remains poor. These results again confirm the fundamental limitations of extrapolation using a single GP model, regardless of kernel expressiveness.

In contrast, we applied NARGP by utilizing not only high-fidelity but also low-fidelity data. Figure 3(c) shows that the result using NARGP with an RBF kernel does not improve the extrapolation performance. An analysis of the optimized hyperparameters of the RBF kernel revealed that the model had not sufficiently captured the dependence on the low-fidelity data f_l, and the prediction effectively reproduced a single-fidelity GP regression based solely on x_0 and x_1. On the other hand, Fig. 3(f) shows the result of NARGP using an NKN kernel, where the prediction accuracy in the extrapolated region was significantly improved. Taken together with the results of the previous sections, we conclude that combining a multi-fidelity data integration algorithm with complex and flexible kernels that handle multiple length scales is indispensable for improving extrapolative predictions.

This improvement in extrapolation performance by NARGP is due to the replacement of the extrapolative dependence of the high-fidelity function by a simplified expression. NARGP expresses the high-fidelity function $f_h(x_0, x_1)$ in the extended parameter spaces $g_h(x_0, x_1, f_l(x_0, x_1))$. In the current problem setup, this function can be expressed simply as a function of the low-fidelity function f_l alone, that is, $f_h(x_0, x_1) = g_h(x_0, x_1, f_l(x_0, x_1)) = g_h(f_l)$. By modeling the high-fidelity function in this form, the model can exclude the direct dependence on the extrapolating parameter x_0, and therefore, convert the original extrapolation problem into an interpolation problem, specifically, learning the relationship between f_l and f_h. In this study, we refer to this effect as "Interpolation Recasting." That is, NARGP achieves high predictive performance in extrapolated regions by transforming a difficult extrapolation problem (with respect to x_0) into a tractable interpolation problem based on the correlation between low- and high-fidelity functions.

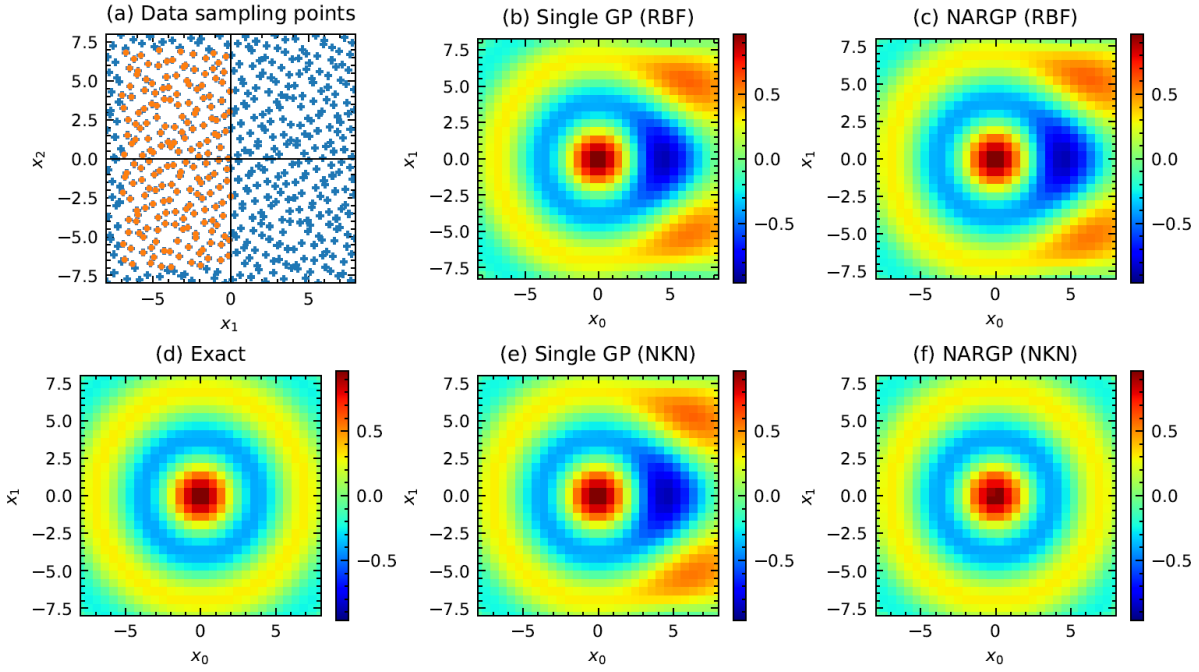

FIG. 3. Extrapolative prediction for a 2D test function. (a) Sampling points of training data: blue crosses show 512 sampling points for low-fidelity functions, while orange dots show 196 sampling points for high-fidelity functions. (b) Single GP prediction with an RBF kernel using only the high-fidelity data. (c) NARGP prediction with an RBF kernel using the low- and

high-fidelity data. (d) Exact profile of the high-fidelity function. (e) Single GP prediction with NKN kernel. (f) NARGP prediction with NKN kernel.

3.4. Multi-fidelity extrapolative prediction of plasma turbulence transport dataset

3.4.1. Problem setup

Finally, we apply the proposed multi-fidelity regression approach to a real experimental dataset on plasma turbulent transport in magnetic confinement fusion devices, JET. In this dataset, 12 local plasma parameters are used as input features, $\boldsymbol{x} = (R/L_n, R/L_{Te}, R/L_{Ti}, n_i/n_e, T_e/T_i, \beta, \nu_{ee}, q, \hat{s}, \varepsilon, \kappa, \delta)$, where the normalized inverse density and electron/ion temperature gradient scale lengths, R/L_n, R/L_{Te}, R/L_{Ti}, the ion-to-electron density ratio n_i/n_e, the electron-to-ion temperature ratio T_e/T_i, the plasma beta value β, the electron-electron collision frequency ν_{ee}, the magnetic geometrical factors such as the safety factor q, the magnetic shear $\hat{s}$, the inverse aspect ratio ε, the elongation κ, and the triangularity δ. The experimentally measured turbulent diffusion coefficient is used as the high-fidelity data, denoted as $y = D_{exp}$, which we aim to predict. To formulate the multi-fidelity framework, we use the linearly most unstable wavenumber k and its growth rate γ in local gyrokinetic microinstability analysis as low-fidelity data. For detailed information on the dataset, refer to our previous publications [2, 11].

In this study, we intentionally restrict the training high-fidelity dataset to regions with relatively high diffusion coefficients (67 points having the D_{exp} value larger than its median of the total 135 data points), and use the model to extrapolate into unseen regions with lower diffusion levels. This setup mimics a realistic and practical scenario in fusion reactor design. The experimental data are always limited to the parameter range of existing devices. In contrast, it is necessary to estimate turbulent transport levels in untested regimes to predict the future performance of magnetic confinement fusion plasma.

3.4.2. Results

The regression results of the conventional GP regression with the RBF kernel are shown in Fig. 4(a). While the model achieves good agreement with the experimental data in the training region, the prediction accuracy drops significantly on the test data. Since the training data only has large values of D_{exp}, the GP model tends to overestimate the transport level on the test data. Next, we replaced the kernel with NKN and retrained the single-fidelity GP model. As shown in Fig. 4(b), the improvement in extrapolation performance is insufficient. This result confirms that the conventional single-fidelity GP regression has general difficulty with extrapolation, even when using a more expressive kernel function.

In contrast, Fig. 4(c) shows that a multi-fidelity regression using NARGP with an NKN kernel reduces the overestimation of transport levels in the test data. The prediction accuracy in the extrapolated region is clearly improved. Specifically, the consistency with experimental values is better in areas with small diffusion coefficients, even though such regions were not included in the training data. A key advantage of multi-fidelity modeling is that low-fidelity information (i.e., linear growth rates γ and its wavenumbers k in the present dataset) remains available even in the extrapolated regions of the high-fidelity data. By learning the nonlinear correlation between low- and high-fidelity data in the training region, the NARGP model can improve predictions in extrapolated areas by leveraging the values of the low-fidelity data and their relationships to high-fidelity outputs (e.g., small γ and large k tend to exhibit low turbulent transport D_{exp}).

3.4.3. Analysis of the outlier and the limitations of extrapolation

It should be noted that Fig. 4 (c) still contains a finite number of outliers with large prediction errors. To understand the limitations of our extrapolative prediction, we analyzed these outliers. Specifically, we evaluated how far each test point lies from the distribution of training data in the multidimensional input space using the Mahalanobis distance [12],

$$d_M(\boldsymbol{z}) = \min_{\boldsymbol{t}_i \in \Omega} \sqrt{(\boldsymbol{z} - \boldsymbol{t}_i) \cdot \Sigma^{-1} \cdot (\boldsymbol{z} - \boldsymbol{t}_i)} \tag{12}$$

where $\boldsymbol{z} = (\boldsymbol{x}, y)$ is the test data point, $\Omega = \{\boldsymbol{t}_i\}$ is the training dataset with i-th training data point $\boldsymbol{t}_i = (\boldsymbol{x}_i, y_i)$, and Σ is the covariance matrix computed using the training dataset. For each test point, we computed the Mahalanobis distance against every training point and used the smallest value as its representative distance against the training dataset. Test points with their representative distances greater than 3.0 were classified as “far” data, while the others were considered “near” data. This classification is visualized in the plots of Fig. 4 by using different symbols. The analysis reveals that all outlier points with large prediction errors tend to be far from the data, having

a large Mahalanobis distance from the dataset. In other words, outliers are points that lie outside the distribution of the training dataset.

This result indicates that while NARGP is a powerful tool to improve the extrapolative predictability, its effectiveness is fundamentally limited to regions where the correlation between low- and high-fidelity data holds. In extreme parameter regimes far from the training distribution, the predictive performance of NARGP may degrade. Therefore, for practical applications, it is essential to establish quantitative methods for identifying "extrapolatable" versus "non-predictable" regions. In this study, we propose a useful framework for such identification, utilizing the Mahalanobis distance as a quantitative measure of distance from the training distribution.

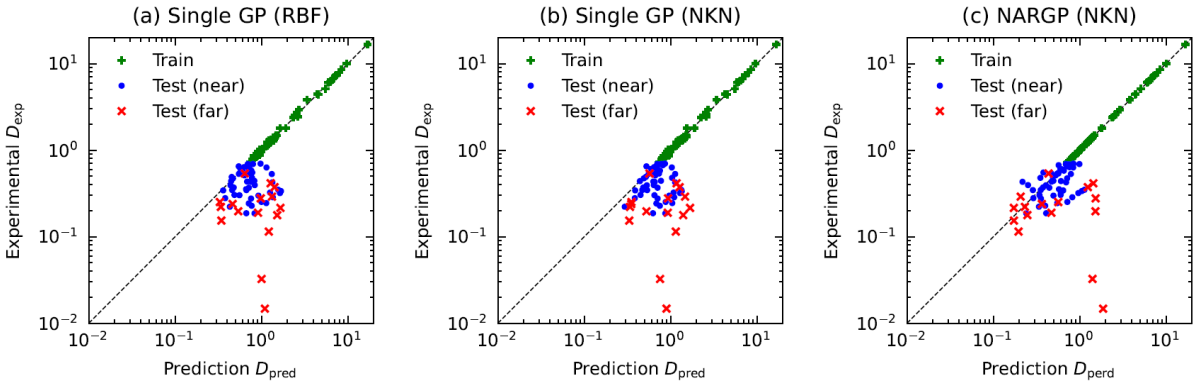

FIG. 4. Extrapolative prediction for turbulent diffusion coefficient from JET experimental dataset. Comparison of predicted and actual values for: (a) single GP prediction with an RBF kernel, (b) single GP prediction with NKN kernel, and (c) NARGP prediction with NKN kernel. Green daggers denote training data. Blue circles represent test data located near the training distribution (as measured by Mahalanobis distance), and red crosses denote test data situated far from the training data distribution.

4. SUMMARY

In this study, we have examined how the extrapolative prediction performance of turbulent transport models in plasma can be improved by introducing a multi-fidelity information fusion approach with a Gaussian process (GP) regression framework. First, we highlighted the critical importance of kernel design in the extrapolation problem. We demonstrated that capturing both short-term variations and long-term trends requires a flexible kernel structure that can represent multiple correlation scales. To this end, we implemented a neural kernel network (NKN) and demonstrated its effectiveness. Building upon this, we applied nonlinear auto-regressive Gaussian process (NARGP) regression as a multi-fidelity regression method, confirming that it further enhances the performance of extrapolation. We identified two underlying mechanisms that contribute to this improvement: (i) Dependence Simplification: Low-fidelity data absorb the complex parameter dependencies inherent in high-fidelity data. This enables the regression model to represent the high-fidelity target using a simpler functional form, thus facilitating extrapolation. (ii) Interpolation Recasting: By leveraging the correlation between low- and high-fidelity data, dependency on extrapolative input parameters can be effectively replaced with interpolation over the extended input space. In this way, extrapolation problems are reformulated into interpolation problems. Finally, we validated the proposed method using a real-world dataset of plasma turbulent transport. While the conventional single-fidelity GP regression suffered a severe drop in accuracy in extrapolated regions, the combination of NKN kernels and NARGP significantly improves the predictive accuracy in extrapolated regions. Furthermore, we demonstrated that the Mahalanobis distance provides a practical and quantitative means to identify feasible regions for extrapolative prediction.

In summary, this study demonstrates that the multi-fidelity information fusion approach is a powerful strategy for tackling extrapolation challenges. By integrating data of various fidelity levels, including theoretical models, simulation results, and experimental measurements, this approach provides a robust foundation for enhancing the accuracy of transport models, ultimately benefiting the development of future high-performance magnetic confinement fusion devices.

ACKNOWLEDGEMENTS

This work was partially supported by JST, PRESTO Grant Number JPMJPR23OB. This work utilized the computational resources of the plasma simulator at the National Institute for Fusion Science (NIFS) and the National Institute of Quantum Science and Technology, under the auspices of the NIFS Collaboration Research program (NIFS23KIST041).

REFERENCES

[1] GARBET, X., et al., Gyrokientic simulations of turbulent transport, Nucl. Fusion **50** (2010) 043002.

[2] MAEYAMA, S., et al., Multi-fidelity information fusion for turbulent transport modeling in magnetic fusion plasma, Sci. Rep. **14** (2024) 28242.

[3] RASMUSSEN, C. E. & WILLIAMS, C. K. I., Gaussian processes for machine learning. Adaptive computation and machine learning, MIT Press (2006).

[4] PERDIKARIS, P., et al., Nonlinear information fusion algorithms for data-efficient multi-fidelity modelling, Proc. R. Soc. A **473** (2017) 20160751.

[5] DUVENAUD, D., et al., Structure Discovery in Nonparametric Regression through Compositional Kernel Search, Proc. 30th Int. Conf. Machine Learning **28** (2013) 1166–1174.

[6] WILSON, A. G., et al., Deep Kernel Learning, Proc. 19th Int. Conf. Artificial Intelligence and Statistics **51** (2016) 370–378.

[7] WILSON, A. G., & ADAMS, R. P., Gaussian Process Kernels for Pattern Discovery and Extrapolation, Proc. 30th Int. Conf. Machine Learning **28** (2013) 1067–1075.

[8] WILSON, A. G., Correction to Spectral Mixture (SM) Kernel Derivation for Multidimensional Inputs, Carnegie Mellon University (2015). Available at: https://www.cs.cmu.edu/~andrewgw/typo.pdf

[9] SUN, S., et al., Differentiable Compositional Kernel Learning for Gaussian Processes, Proc. 35th Int. Conf. on Mach. Learn. **80** (2018) 4828–4837.

[10] OpenML, mauna-loa-atmospheric-co2, Data Set, OpenML, Data ID: 41187 (Aug. 30, 2018). Available at: https://www.openml.org/d/41187

[11] NARITA, E. et al., Modification of a machine learning-based semi-empirical turbulent transport model for its versatility, Contrib. Plasma Phys. **63** (2023) e202200152.

[12] MAHALANOBIS, P.C., On the Generalized Distance in Statistics, Proc. Nat. Inst. Sci. India **2** (1936) 49–55.