Transport model surrogates via Gaussian process regression with neural network kernels

M. Honda[1], E. Narita[1,2], and S. Maeyama[2]

[1]*Kyoto University, Kyoto, Kyoto 615-8530, Japan*

[2]*National Institute for Fusion Science, Toki, Gifu 509-5292, Japan*

Introduction

Neural network (NN) surrogate models are widely used for simulating turbulent transport models like TGLF [1, 2], but they typically require extensive datasets, leading to larger model sizes, longer training times, and a lack of uncertainty quantification. To address these limitations, we explore Gaussian Process Regression (GPR), a nonparametric Bayesian method that provides uncertainty estimates and performs well in small-data regimes. The predictive performance and uncertainty estimates of GPR heavily depend on the kernel choice, which defines input space similarity and acts as a prior for the model function. While data characteristics often guide kernel selection, TGLF surrogate construction involves high-dimensional, complex input-output relationships, making a priori kernel determination challenging.

Expressivity of deep kernels

The connection between deep NNs and GPs, established via the central limit theorem, has enabled closed-form expressions for kernels corresponding to error-function (ERF) and ReLU activations. These form the basis of Neural-Net-induced Gaussian Processes (NNGPs) [3]. NNGPs allow for varying weight variances across input neuron connections and separate tuning of weight/bias variances per layer, parameterizing layer-specific scale factors. The high expressivity of NNGPs, often realized through deep ReLU/ERF kernels, arises from a hierarchical, layer-by-layer iterative construction. Adopting these kernels allows GPR to approximate deep learning behavior while retaining probabilistic interpretability. Here, we refer to kernels corresponding to infinitely wide deep neural networks as "deep kernels".

Since no existing GPR library supported deep kernel implementation to our knowledge, we created "dgpr", a custom GPR library in Python using JAX for automatic differentiation and GPU acceleration. It supports full GP and SVGP with various kernels, including deep kernels. It also supports the Intrinsic Coregionalization Model for multi-output regression. It shows good agreement in benchmarks against GPy and GPflow.

To evaluate the performance of deep kernels, we conducted a reproducibility test using the following true function shown in Fig. 1(a): $f(x,y) = \sin(x) + \cos(y^2),\ x,y \in [-3,3]$. This func-

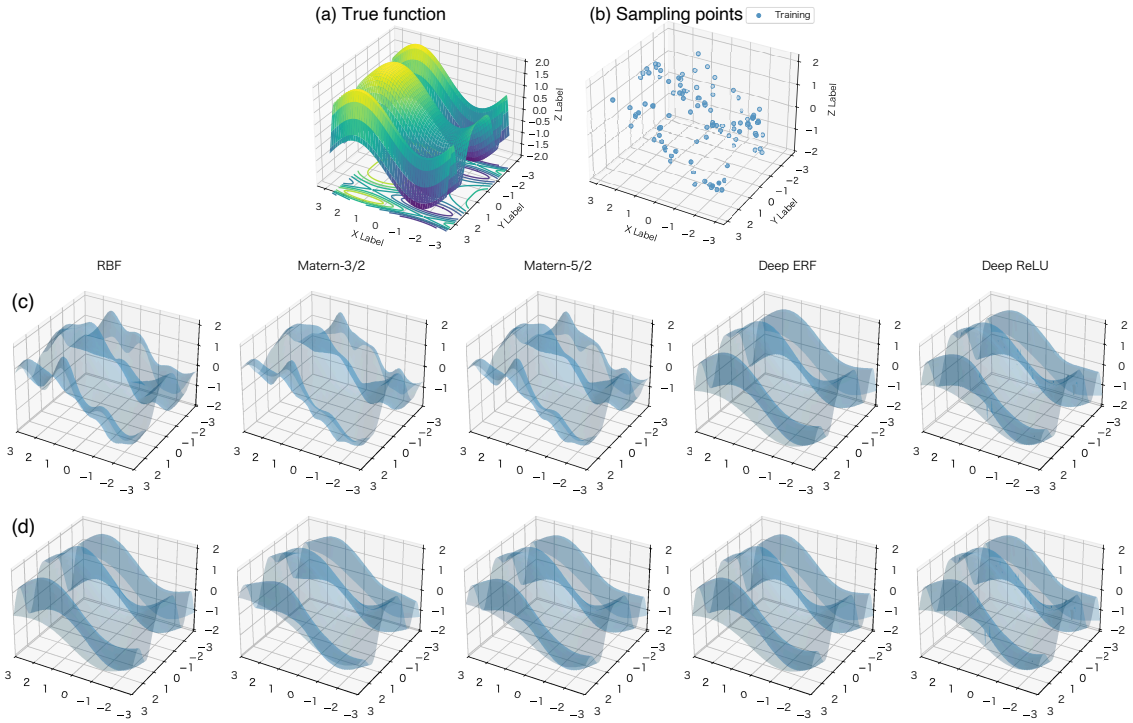

Figure 1: Performance of deep kernels. Panel (a) shows a plot of the true function $f(x,y) = \sin(x) + \cos(y^2)$, and panel (b) displays 100 randomly sampled data points from $f(x,y) + \varepsilon$ where $\varepsilon \sim \mathcal{N}(0, 0.1^2)$. Panel (c) presents the regression results using five different kernels: RBF, Matérn-3/2, Matérn-5/2, deep ERF, and deep ReLU kernels, while panel (d) shows the regression results with Automatic Relevance Determination (ARD) enabled. Since deep kernels do not support ARD, their results remain unchanged between panels (c) and (d).

tion, $f(x,y)$, was sampled with additive Gaussian noise of standard deviation 0.1, and 100 points were randomly selected from the resulting dataset (Fig. 1(b)). GPR was then performed using various kernels, and the quality of the reconstruction was assessed using the coefficient of determination, R^2.

As shown in Fig. 1(a), the true function exhibits different periodicities in the x- and y-directions. As illustrated in Fig. 1(c), standard kernels such as the Radial Basis Function (RBF) and Matérn kernels, which assume a single length scale, fail to capture the anisotropic features of the function. Consequently, their regression performance is limited, with R^2 scores of 0.856, 0.831, and 0.842, respectively. In contrast, the deep ERF kernel and deep ReLU kernel with three hidden layers effectively capture the differing characteristic scales along each input dimension, achieving much higher R^2 scores of 0.974 and 0.963, respectively. When Automatic Relevance Determination (ARD) is enabled, which allows each input dimension to have its own length scale, the performance of the RBF and Matérn kernels improves significantly, as shown

in Fig. 1(d). The resulting R^2 values are 0.977, 0.937, and 0.958, making their performance comparable to that of the deep kernels.

These results demonstrate that deep kernels are capable of accurately capturing anisotropic structure without prior consideration tailored to the input data, even when the underlying function exhibits distinct behaviors across input dimensions. This property makes them particularly well-suited for surrogate modeling tasks where the characteristics of the input parameters vary significantly across dimensions.

Stochastic Variational Gaussian Process

Scalability is another challenge for GP-based surrogates. Applying GPR to large datasets is computationally prohibitive due to memory and computational costs. Stochastic Variational Gaussian Process (SVGP) [4, 5] offers a scalable alternative by introducing a small number of inducing points through which all predictions are made, significantly reducing these costs.

SVGP-based transport model surrogates

We construct an SVGP-based surrogate model in this section using $M = 20$, where M is the number of inducing points. From 147,529 data points generated by GOTRESS [2] with TGLF simulations, 3,000 points were randomly sampled and then split into training (2,700 points) and test (300 points) sets. Figure 2 shows the temperature profile prediction results of GOTRESS simulations using SVGP surrogates (trained on 2,700 points with deep ReLU, Matérn-3/2, and Matérn-5/2 kernels) compared to an NN surrogate and the original TGLF model. All surrogate models successfully reproduced TGLF's temperature profiles. In all cases, the simulations using surrogate models converged within 8 minutes, demonstrating a significant acceleration compared to GOTRESS with TGLF, which required 5 hours using 2,222 cores. A simulation was also conducted using a reduced training dataset of 40 points. With only 40 data points, prediction becomes feasible even with a full GP, without the need for SVGP. Remarkably, the full GP and SVGP surrogate models with deep ReLU kernel still accurately reproduced temperature profiles in GOTRESS simulations, whereas the NN surrogate failed due to the shortage of the data points, as shown in Fig. 3. This demonstrates GPR's effectiveness, particularly with deep kernels, for developing robust surrogates in low-data regimes for complex physical simulations.

References

[1] O. Meneghini, S. P. Smith, P. B. Snyder, G. M. Staebler, J. Candy, E. Belli, L. Lao, M. Kostuk, T. Luce, T. Luda, J. M. Park, and F. Poli. *Nucl. Fusion*, 57:086034, 2017.

[2] M. Honda and E. Narita. *Phys. Plasmas*, 26:102307, 2019.

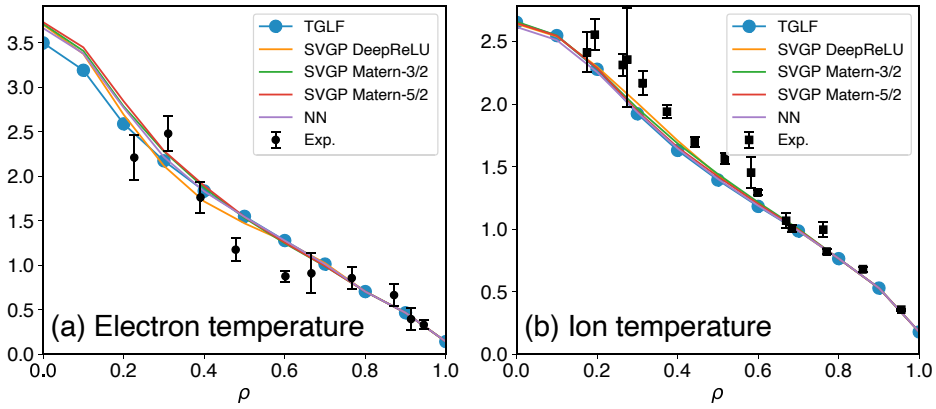

Figure 2: (a) Electron and (b) ion temperature profile predictions using GOTRESS. The results obtained using SVGP surrogate models with three different kernels and the NN surrogate model trained on 2,700 points are compared with those obtained using TGLF.

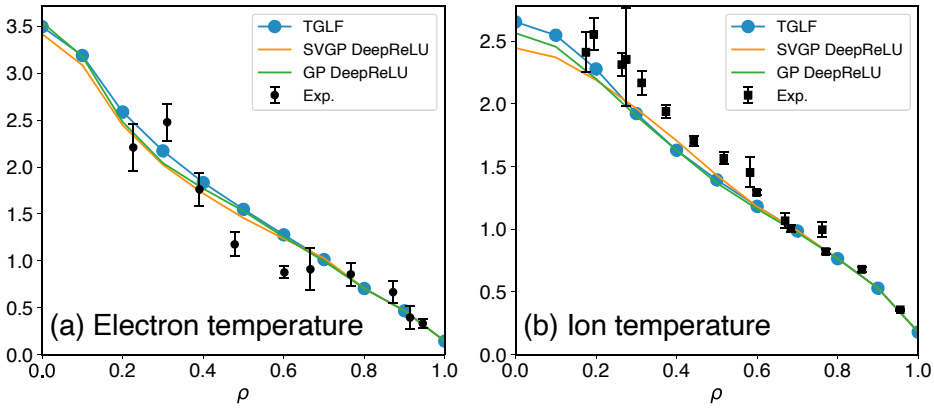

Figure 3: (a) Electron and (b) ion temperature profile predictions using GOTRESS with SVGP and GP surrogate models with the deep ReLU kernel trained on 40 points.

[3] Guofei Pang, Liu Yang, and George Em Karniadakis. *J. Comput. Phys.*, 384:270–288, 2019.

[4] James Hensman, Nicolo Fusi, and Neil D. Lawrence. Gaussian processes for big data, 2013.

[5] James Hensman, Alex Matthews, and Zoubin Ghahramani. Scalable variational gaussian process classification, 2014.