

Prediction of Plasma Confinement Indices by Gaussian Process Regression

Sora YABUMOTO^{1)*}, Shinsuke SATAKE^{1,2)}, Hiroyuki YAMAGUCHI^{1,2)}

¹⁾ The Graduate University for Advanced Studies, SOKENDAI, 322-6 Oroshi-cho, Toki, Gifu 509-5292, Japan

²⁾ National Institute for Fusion Science, National Institutes of Natural Sciences, 322-6 Oroshi-cho, Toki 509-5292, Japan

(Received 12 June 2024 / Accepted 18 February 2025)

To optimize the design of a helical fusion reactor by varying the shape of the magnetic coils, several requirements related to the performance of the reactor should be satisfied under various constraints. To address this multi-objective optimization problem, we utilized Gaussian process regression (GPR) for machine learning to develop a surrogate model capable of predicting the dependence of the objective functions on the parameters representing the coil shape. This study demonstrates that the dependence of objective functions, such as plasma volume and the Mercier criterion, on the shape of helical coil windings can be predicted by GPR.

© 2025 The Japan Society of Plasma Science and Nuclear Fusion Research

Keywords: fusion reactor design, machine learning, Gaussian process, Mercier criterion

DOI: 10.1585/pfr.20.1403026

1. Introduction

To realize a commercial helical fusion reactor, it is necessary to choose a device that simultaneously satisfies various requirements [1–3] for a wide variety of magnetic field configurations. These include the conditions for the confinement performance of the reactor and engineering constraints. For example, the former includes MHD stability, suppression of neoclassical and turbulent transport, and high fusion energy gain. The latter includes the distance between the coils and plasma to ensure blanket space, distance between helical coils, and maximum curvature and torsion of the superconducting coils [4]. In addition, the construction and operation of a reactor must satisfy various constraints such as seismic resistance and radiation protection and must be economically feasible. Thus, the fusion reactor design is an example of a multi-objective optimization problem. In such cases, there are many trade-off relationships between requirements, and it is generally difficult to understand the relationships between a large number of design parameters and objective functions. In fusion reactor design, the parameter dependence of such complex and nonlinear objective functions must be explored, and thus, a reasonable design can be achieved. The ultimate goal of our study is to solve a multi-objective optimization problem using machine learning.

In previous studies on magnetic field configuration optimization, it has been a general method that the shape of the last closed-flux surface (LCFS) was varied and the MHD stability and the reduction of plasma transport were optimized. The shape of the coils that produced the optimized magnetic

configuration was then considered as another optimization problem [5, 6]. Most of these studies aimed to achieve a certain symmetry in the magnetic field configuration because it is theoretically known that configurations with good symmetry in Boozer coordinates have low neoclassical transport levels (*e.g.*, quasi-axisymmetric and quasi-helically symmetric configurations) [7].

This study proposes a different approach that uses the winding law of helical coils as a parameter to generate various magnetic-field configurations and optimize the objective functions. We consider a variety of helical coil shapes different from simple ones, such as LHD, and attempt to predict magnetic field configurations that satisfy the aforementioned requirements using Gaussian process regression (GPR). The advantage of this method is that, by imposing engineering constraints on the coils in advance, it is possible to search for optimized configurations from a large number of technically feasible ones. A previous study shows that various magnetic field configurations, including quasisymmetric configurations, can be created using helical coils with more degrees of freedom compared to conventional devices [8]. In addition, the performance of helical fusion reactors was evaluated using a modified winding law for LHD-like helical coils [1]. Therefore, it is of practical significance to consider a continuous winding coil method to optimize fusion reactors. Generally, the machine learning methods used for optimization include the gradient method, genetic algorithms (GA), neural networks (NN), and Gaussian process regression (GPR). In research on optimized reactor designs, gradient-based methods have been adopted in STELLOPT and FOCUS codes [9, 10], for example. The optimization of fusion reactors using a GA was proposed in a recent study [11]. The goal of our research

*Corresponding author's e-mail: yabumoto.sora@nifs.ac.jp

is to perform a multi-objective optimization of magnetic field configurations using GPR as a surrogate model to predict the confinement performance. As a preparatory step, we implemented GPR for certain objective functions to be used in the optimization study and evaluated the prediction performance of GPR. The objective functions selected were plasma volume, Mercier criterion, and Shafranov shift, which are indicators of MHD stability. Moreover, the feasibility of predicting the dependence of these objective functions on coil-shape parameters using GPR was investigated.

The remainder of this paper is structured as follows: Sec. 2 describes the method, including the calculation of the magnetic configurations and objective functions, and GPR. Section 3 presents the results of calculation and GPR. Finally, the results are summarized in Sec. 4.

2. Method

2.1 Objective of optimization

We considered the design of a fusion reactor with a pair of helical coils based on LHD [12]. The characteristics of the plasma confinement of LHD are that the magnetic configuration in which the magnetic axis shifts outward has good MHD stability, while the inward-shifted one has good particle confinement and a low neoclassical transport level [13]. One of our objectives was to determine a possible configuration in which both good MHD stability and low neoclassical transport could be achieved by changing the helical coil winding. A previous study showed that a continuously winding coil can create a quasi-isodynamic configuration [8]. Our research goal is to create a surrogate model for optimized configuration research by adopting GPR to learn the dependency of objective functions, such as MHD stability and neoclassical transport, on the coil shape parameters. This study aimed to verify whether such learning is possible.

2.2 Calculation of various magnetic configurations

This section explains the method for producing the training data. The winding law of a helical coil in the (R, Z, ϕ) -coordinates is given as follows:

$$\left\{ \begin{array}{l} R(\phi) = R_{ax}(\phi) \\ \quad + r_{00}\epsilon_r \cos(\lambda(\phi) + \alpha_r \sin(\lambda(\phi) + \phi_{0r})) \\ \quad [1 + \epsilon_{2r} \cos(\lambda(\phi) + \phi_{2r})], \\ Z(\phi) = Z_{ax}(\phi) \\ \quad + r_{00}\epsilon_z \sin(\lambda(\phi) + \alpha_z \sin(\lambda(\phi) + \phi_{0z})) \\ \quad [1 + \epsilon_{2z} \cos(\lambda(\phi) + \phi_{2z})], \\ R_{ax}(\phi) = R_{00}(1 + \epsilon_{aR} \cos(\lambda(\phi) + \phi_{aR})), \\ Z_{ax}(\phi) = R_{00}\epsilon_{aZ} \sin(\lambda(\phi) + \phi_{aZ}), \\ \lambda(\phi) = N\phi + \phi_{00}. \end{array} \right. \quad (1)$$

The parameters mean as follows: ϵ_r and ϵ_z are magnifications of the minor radius of the helical coil, ϵ_{2r} and ϵ_{2z} are the scalar values to provide the bumpiness of helical windings, α_r and α_z are the helical pitch modulation parameters,

and ϕ_{0r} , ϕ_{0z} , ϕ_{2r} , and ϕ_{2z} are the relative initial phases, respectively. In addition, $\lambda(\phi)$ represents the phase angle of the helical coils. Another coil is given by a symmetric law $(R(\phi), -Z(\phi), -\phi)$, which makes the produced magnetic field stellarator symmetry. There are two pairs of circular vertical field coils (OV and IV) of which positions are fixed at $(R, Z) = (5.75, \pm 1.60)$ and $(1.70, \pm 1.00)$, respectively. The LHD is equipped with three pairs of vertical field coils; however, one of the pairs was omitted for simplicity. FFHR-c1, which is a design of LHD-like reactor, also omits one of these pairs [1].

The range of parameters in the law is listed in Table 1, where OV and IV are the magnification factors for the coil currents of the OV and IV coils, respectively. The base constant parameters are: $R_{00} = 3.90$ m, $r_{00} = 0.975$ m, $\phi_0 = 90^\circ$. The coil current of the helical coils is 5.636 MA, and the base values for the OV and IV coil current are -3.615 MA and 3.013 MA, respectively, where the positive sign indicates the counter-clockwise direction when the device is viewed from the top. In this study, the helical coil current is kept constant. The “base case,” defined as a configuration that closely resembles the shape of the LHD, has the following parameters: $\alpha_r = \alpha_z = 0.1$, $\epsilon_r = \epsilon_z = 1.0$, $\epsilon_{2r} = \epsilon_{2z} = 0.0$, and $\phi_{0r} = \phi_{0z} = 0.0$.

Engineering constraints are imposed on the distance between the helical coils, between the helical coil and the vertical field coil, and between the helical coil and the LCFS of the plasma. We also imposed an upper limit on the curvature and torsion of the helical coils and a lower limit on the plasma volume. The constraints are as follows: the maximum curvature $\kappa_{\text{lim}} = 3.0 \text{ m}^{-1}$; the maximum torsion $\tau_{\text{lim}} = 20 \text{ m}^{-1}$, the maximum length of helical coils is 100 m, the minimum distance between helical coils is 0.4 m, and the maximum distance between them on a $\phi = \text{const.}$ plane is 4.0 m, and the minimum distance between the helical coils and LCFS is 0.15 m. The constraints κ_{lim} and τ_{lim} are not strictly implemented, as the maximum values of these parameters for the LHD helical coils are roughly estimated to be 0.7 and 1, respectively, assuming an HTS tape width $h = 5$ mm. $\tau_{\text{lim}} = 20$ is estimated from the evaluation of the constraints for the edge strain of HTS, $\epsilon \simeq 0.5(h\tau)^2 < 0.005$ [4, 14]. The minimum

Table 1. Variable parameters in the coil shape calculation.

Parameter	Range
OV	0.70, 0.80, 0.90, 1.00, 1.10, 1.20
IV	0.80, 0.90, 1.00
α_r	-0.20, -0.10, 0.00, +0.10, +0.20
α_z	-0.20, -0.10, 0.00, +0.10, +0.20
$\epsilon_r = \epsilon_z$	1.00, 1.10, 1.20
$\epsilon_{2r} = \epsilon_{2z}$	0.00, 0.10, 0.20, 0.30
ϕ_{0r}	-10.0, 0.0, +10.0
ϕ_{0z}	-10.0, 0.0, +10.0
$\phi_{2r} = \phi_{2z}$	0.0, 15.0, 30.0, 45.0, 60.0, 75.0, 90.0, 105.0, 120.0, 135.0, 150.0, 165.0, 180.0

plasma volume is set at 8.0 m³. As a reference, the LHD volume is approximately 25 m³. However, in the parameter scan demonstrated in Sec. 3, the range of coil parameters was restricted in reality only by the plasma volume because we adopted loose limits on the distances, coil curvature, and torsion. In the present test, ϵ_{aR} and ϵ_{aZ} were fixed to zero to reduce the dimension, and we limited parameters to $\epsilon_r = \epsilon_z = \epsilon$, $\epsilon_{2r} = \epsilon_{2z} = \epsilon_2$, and $\phi_{2r} = \phi_{2z} = \phi_2$. To study the beta dependence, the central beta values in MHD equilibrium calculations are varied, $\beta = 0.5\%$, 1.5% , 2.5% . We did not try higher- β case because the calculation at the large beta value is not reliable since we used the fixed boundary calculation of VMEC [15]. In the parameter scan, the calculation of degeneracy cases was omitted. The degeneracy is, for example, in the case with $\epsilon_{2r} = 0$, the value of ϕ_{2r} can be ignored. In such a case, only one representative case ($\phi_{2r} = 0$) is calculated and the same result is referenced to for $\phi_{2r} \neq 0$ cases to learn the dependence of objective functions by GPR.

In the present method, to vary the configuration, some of the generated magnetic configurations had large magnetic islands. We developed a method to detect and remove them by image recognition of the Poincaré plots of magnetic field lines [16]; however, we did not use it in the present study for simplicity. Thus, the training data contained relatively large magnetic islands. However, in the MHD equilibrium calculation using the ideal MHD code VMEC, nested flux surfaces without magnetic islands is obtained even if there are magnetic islands in the vacuum magnetic field.

2.3 Objective functions

The objective functions chosen here to demonstrate the prediction capability of the GPR are the plasma volume (V), minimum value of the Mercier criterion in the minor radius direction ($D_{\text{Merc}}(\min)$), and Shafranov shift (S). The input parameters for the GPR specify the helical coil shape, as shown in Table 1. Because the Mercier criterion,

$$D_{\text{Merc}} > 0, \quad (2)$$

is a condition for ideal MHD stability [17], we treat its minimum value in the minor radius direction $D_{\text{Merc}}(\min)$ as an objective function such that the entire plasma volume becomes MHD stable. Here, we observe $D_{\text{Merc}}(\min)$ in the range $0.15 < (r/a)^2 < 0.95$. This is due to the unreliability of the calculation of D_{Merc} in the VMEC near the magnetic axis and LCFS, as highlighted in [18]. The Shafranov shift is defined as

$$S \equiv \frac{R_{\text{ax, avg}}(\beta) - R_{\text{ax, avg}}(0.0)}{a}, \quad (3)$$

where $R_{\text{ax, avg}}(\beta)$ is the averaged position of magnetic axis in the major radius direction at each β value, $R_{\text{ax, avg}}(0.0)$ is one at $\beta = 0.0$, and a is the average of the minor radius. The Shafranov shift is not a direct index of plasma confinement, but the variation in the magnetic configuration is smaller because the shift is smaller in the high-beta case. Therefore, we chose the value of S as a candidate for the optimization

index. In this study, we did not use any indices for neoclassical and turbulent transport because the cost of evaluating their objective functions is relatively high.

2.4 Gaussian process

In this section, we briefly describe the Gaussian process. For N predictions $\mathbf{y} = (y_1, \dots, y_N)$, the linear regression model $\mathbf{y} = \Phi(\mathbf{x}) \cdot \mathbf{w}$ is used. Here, \mathbf{x}_i , $i = 1, 2, \dots, N$ are the input parameters, and $\mathbf{w} = (w_1, w_2, \dots, w_L)$ is the weight vector, and

$$\Phi(\mathbf{x}) = \begin{pmatrix} \phi_1(x_1) & \cdots & \phi_L(x_1) \\ \vdots & \ddots & \vdots \\ \phi_1(x_N) & \cdots & \phi_L(x_N) \end{pmatrix} \quad (4)$$

is a design matrix, and $\phi_i(x) = \exp(x^2/\sigma_i^2)$ is a feature vector. Assuming that \mathbf{w} follows Gaussian distribution, $\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \lambda^2 \mathbf{I})$, the expectation value of \mathbf{y} follows a multivariate normal distribution,

$$\mathbf{y} \sim \mathcal{N}(\mathbf{0}, \lambda^2 \Phi \Phi^T). \quad (5)$$

Instead of evaluating a covariance matrix $\lambda^2 \Phi \Phi^T = \mathbf{K}$, a kernel function $K_{n, n'} = \phi(\mathbf{x}_n)^T \phi(\mathbf{x}_{n'}) = k(\mathbf{x}_n, \mathbf{x}_{n'})$ such as radial basis function (RBF) kernel $k(\mathbf{x}_n, \mathbf{x}_{n'}) = \theta_1 \exp(-|\mathbf{x}_n - \mathbf{x}_{n'}|^2/\theta_2)$ can be used. Then in GPR, the goal is to optimize two hyperparameters θ_1 and θ_2 , and we do not need to compute the high-dimensional vector \mathbf{w} . In this study, the objective functions were the simulation results and did not contain any observation noise. In fact, they generally contain noise. Therefore, assuming that an objective function has the error ϵ , which follows Gaussian distribution, $\mathcal{N}(0, \sigma^2)$, the prediction follows a multivariate normal distribution,

$$\mathbf{y} \sim \mathcal{N}(\mu, \mathbf{K} + \sigma^2 \mathbf{I}). \quad (6)$$

In this case, a kernel function is represented by

$$k'(\mathbf{x}_n, \mathbf{x}_{n'}) = k(\mathbf{x}_n, \mathbf{x}_{n'}) + \sigma^2 \delta(n, n').$$

Thus, a hyperparameter $\theta_3 = \sigma^2$ is added. Including the noise term in the kernel function contributes to enhanced numerical stability of the GPR and prevents overfitting of the hyperparameters.

The execution environment was an Nvidia A100 80GB GPU. The maximum number of updates in the hyperparameter optimization was set to 50. Python code was developed using the GPyTorch library. During preprocessing, the input parameter \mathbf{x} and target \mathbf{y} were normalized. The code was run on a GPU to accelerate the calculations. In this study, the kernel function was the radial basis function (RBF) kernel. Adam [19] was used to tune the hyperparameters. The accuracy of the predicted values was evaluated using the normalized root mean square error (NRMSE).

$$\text{NRMSE} = \frac{\sqrt{\frac{\sum_{i=1}^N (\hat{y}_i - y_i)^2}{N}}}{y_{\max} - y_{\min}}, \quad (7)$$

where \hat{y}_i is the predicted value, y_i is the true value, y_{\max} is the maximum value of the test data, and y_{\min} is the minimum value.

3. Result

3.1 Training data

Figure 1 shows histograms of the plasma volume and Mercier criteria. The number of computed data points was 60,933, of which approximately 51,000 had an LCFS. Histograms of the plasma volume and $D_{\text{Merc}}(\text{min})$ are shown in Fig. 1. In the dataset, 25,000 points satisfied the requirement for the plasma volume ($V > 8 \text{ m}^3$) and were able to calculate the MHD equilibrium on the VMEC. Moreover, there were 5,500 cases which had a volume exceeding that of LHD. The number of degeneracy cases corresponding to the computed data was 171,730. However, not all degeneracy cases can be used owing to a GPU memory shortage. Approximately 70,000 data points were used for learning. In this parameter range, there was a large fraction of data with $D_{\text{Merc}}(\text{min}) < 0$. This is because we lacked prior knowledge of the parameters for which $D_{\text{Merc}}(\text{min}) < 0$, and thus, determined the parameters using random numbers. The ratio of the training data to the test data was 8:2.

It was observed that the coil parameters ϕ_2 , ϵ , and ϵ_2

have relatively larger effects on V and D_{Merc} than the others. As the examples, the histograms of plasma volume and Mercier criterion at $\beta = 2.5\%$ with respect to ϵ_2 are shown in Fig. 2. In particular, the data are not uniformly distributed with respect to ϵ_2 , because the parameters are randomly chosen. With respect to the plasma volume, it tends to be larger when $\epsilon_2 = 0.0$ than when $\epsilon_2 > 0.0$. In contrast, there are more cases with $D_{\text{Merc}}(\text{min}) > 0$ at $\epsilon_2 = 0.0$ compared to those at other ϵ_2 . Therefore, concerning ϵ_2 , $\epsilon_2 = 0.0$ is optimal to have both large volume and good MHD stability.

A scatter plot of $D_{\text{Merc}}(\text{min})$ versus V is shown in Fig. 3. The large-volume cases are scattered close to $D_{\text{Merc}}(\text{min}) = 0$, whereas the small-volume cases include those with large $D_{\text{Merc}}(\text{min})$ in the negative direction. This tendency indicates that better MHD stability will be achieved in cases $\epsilon_2 > 0.0$, at the expense of volume. If Pareto-optimal solutions are sought for these objective functions, the relationship between V and $D_{\text{Merc}}(\text{min})$ is an example of a multi-objective optimization problem with conflicting relations.

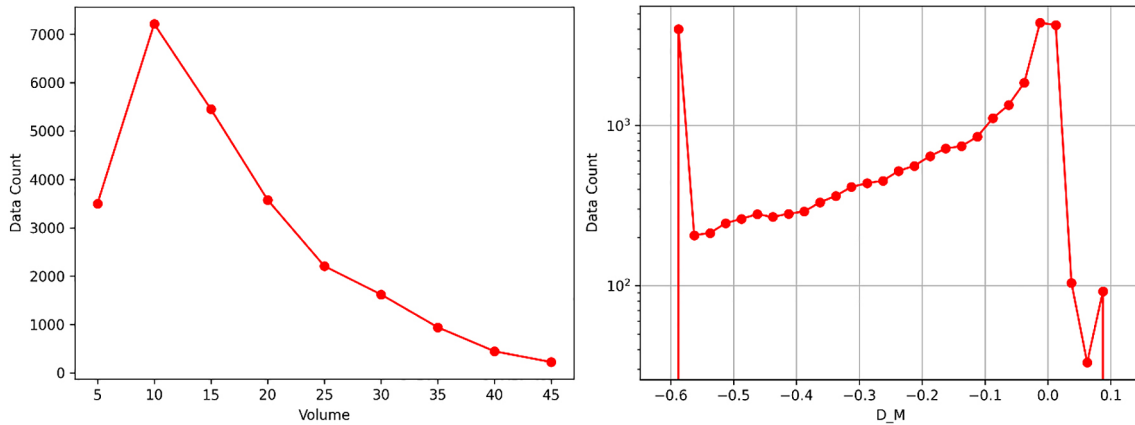


Fig. 1. The histograms of plasma volume (left image) and Mercier criterion at $\beta = 2.5\%$ (right image). The data without degeneracy cases are plotted. The vertical axis of graph of $D_{\text{Merc}}(\text{min})$ is in logarithmic scale. Only the training data which satisfy $V > 8 \text{ m}^3$ are plotted in the figures. In the right figure, the data count for $D_{\text{Merc}}(\text{min}) \leq -0.6$ are accumulated to the $D_{\text{Merc}}(\text{min}) = -0.6$ case.

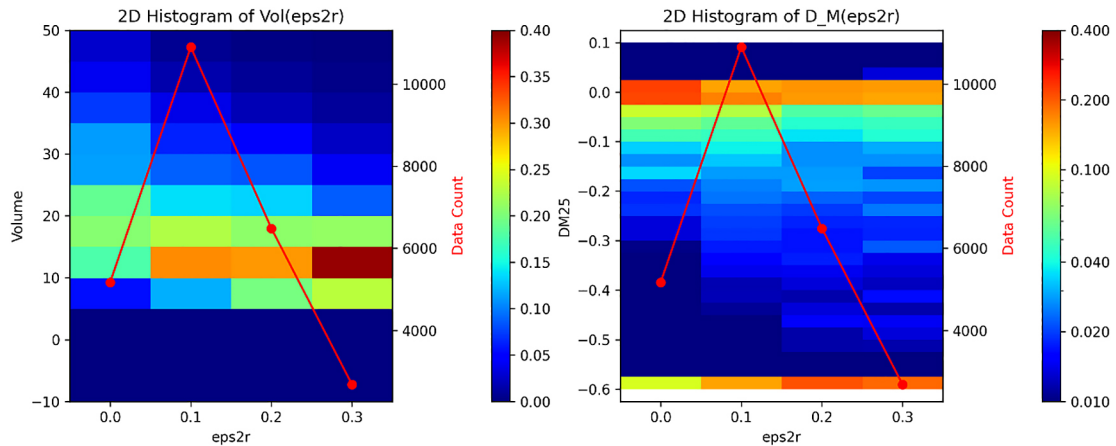


Fig. 2. The histograms of plasma volume (left image) and Mercier criterion at $\beta = 2.5\%$ (right image) with respect to $\epsilon_{2r} = \epsilon_{2z} = \epsilon_2$. The color contour represents the ratio of the data points to the total number of data points for each value of ϵ_2 , and the red line (right axis) represents the total number of data points for each value of ϵ_2 .

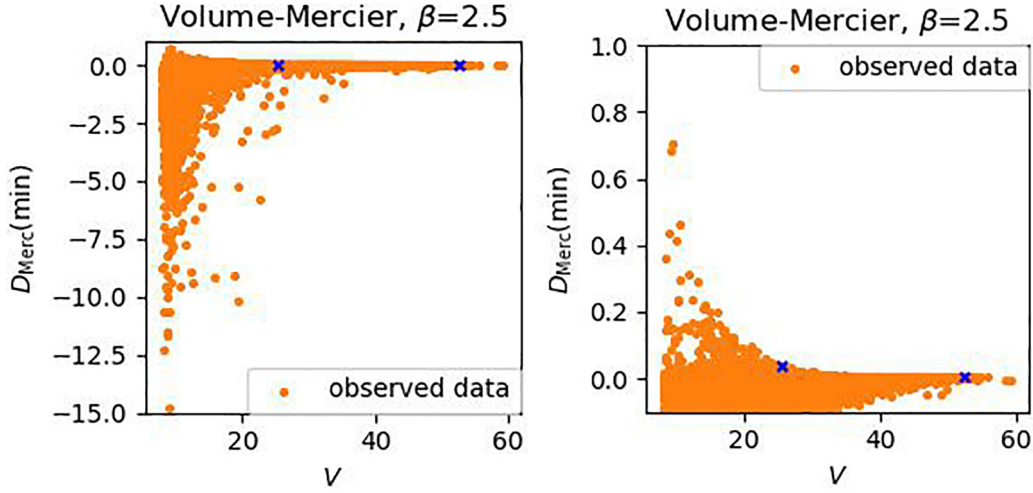


Fig. 3. The scatter plot of plasma volume and $D_{\text{Merc}}(\text{min})$. The cases with $V \leq 8$ are omitted. The enlarged image with $D_{\text{Merc}}(\text{min}) \approx 0$ is shown on the right. The blue cross marks are samples shown in Fig. 7. The right blue point has $\epsilon_2 = 0.0$, and the left one has $\epsilon_2 = 0.1$, respectively.

3.2 Prediction of GPR

As explained in Sec. 2.4, the GRP model used in this study has three hyperparameters that were optimized in 30–50 iterations. The prediction results for the plasma volume V are shown in Fig. 4. The predicted values were distributed around $V_{\text{pred}} = V_{\text{calc}}$ (green line), and the linear regression of V_{pred} using the least-squares method (orange dashed line) almost overlapped. A comparison of the standard deviation of the predicted values obtained by least-squares regression (magenta line) with the estimated prediction error for each point by GPR (blue line) indicated that the former was much smaller. The estimation of the prediction error in the GPR represents the degree of confidence in the prediction. Although the prediction error is large, the prediction points are not widely scattered from $V_{\text{pred}} = V_{\text{calc}}$ line. This indicates that GPR can effectively capture the relationship between plasma volume and coil parameters.

Next, Fig. 5 shows the prediction of Shafranov shift for three plasma β values. The predicted values exhibited a larger degree of dispersion than in the volume case. However, the green and orange dashed lines almost overlapped, and the NRMSE was 0.04. There was no significant difference in the prediction performance among these three β cases. Thus, the dependence of Shafranov shift on the coil parameters and plasma β are captured well similarly to one of the plasma volume.

Figure 6 shows the prediction of the minimum values of Mercier criterion, $D_{\text{Merc}}(\text{min})$, for three plasma β values. The predicted values for the large negative y_{calc} show large deviation from the true values around $y_{\text{calc}} = 0$ and $y_{\text{calc}} < -5$, while the overall dependence of $D_{\text{Merc}}(\text{min})$ is reproduced to some extent. For $D_{\text{Merc}}(\text{min})$, it is important that its value be positive for optimization. It is expected that the confidence level of the surrogate model for $D_{\text{Merc}}(\text{min})$ utilized in the optimization can be enhanced by adding training data for which $D_{\text{Merc}}(\text{min})$ is close to zero or positive. This was achieved by exploring the optimized configurations using the

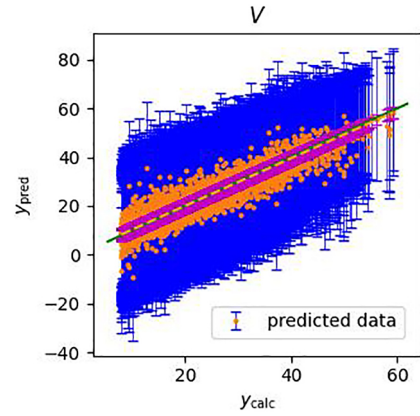


Fig. 4. Predicted and calculated values of plasma volume, V . The horizontal axis is the calculated values, and the vertical axis is the predicted values from GPR. The blue lines are estimation of expectation error $\pm\sigma$ of the GPR, and the green line shows $y_{\text{pred}} = y_{\text{calc}}$. The orange dashed line and the magenta region represent the linear least-squares regression of y_{pred} and its standard deviation $\pm\sigma$, respectively. The processing time required GPR was 5 minutes, and the NRMSE was 0.04. The variance of the predictions obtained by least-squares method was 2.25.

surrogate model and repeatedly updating the model with new training data. We plan to carry out the cycle of prediction and updating the surrogate model by Bayesian optimization (BO) [20, 21]. In BO, a GPR trained using a set of training data is used as a surrogate model to predict the coil parameters with which the objective function is expected to be maximized. The objective function ($D_{\text{Merc}}(\text{min})$ in this case) is then evaluated based on these prediction points. The results are added to the training data, and the surrogate model is updated. By repeating the cycle, the population of the training data around the maximum of the objective function increases while seeking the global maximum of the objective function, and the precision of the surrogate model around the maximum point is expected to improve.

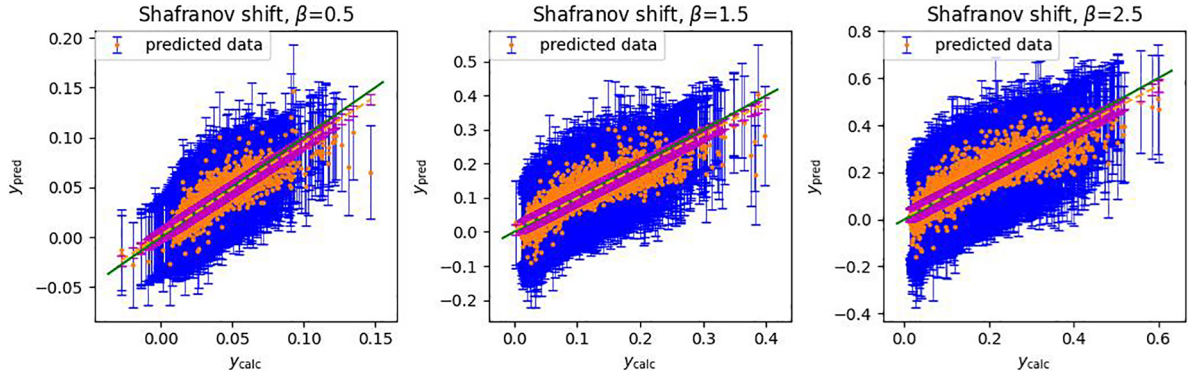


Fig. 5. Predicted and calculated values of Shafranov shift, S , for the three β -value MHD equilibria. The horizontal axis is the calculated values, and the vertical axis is the predicted values. The meaning of the lines and points are the same as those in Fig. 4. The processing time required GPR was 11 minutes, and each NRMSE was 0.04. The errors of predictions obtained by linear least-squares regression were 0.005, 0.016, and 0.026, respectively.

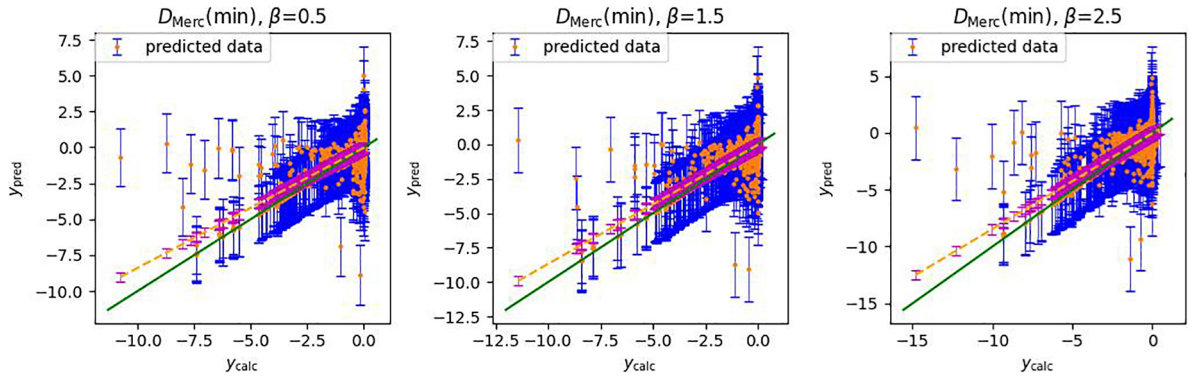


Fig. 6. Predicted and calculated values of $D_{\text{Merc}}(\text{min})$. The horizontal axis is the calculated values, and the vertical axis is the predicted values. The meaning of the lines and points are the same as in Fig. 4. The processing time required GPR was 8 minutes, and each NRMSE was 0.03. The errors of predictions obtained by linear least-squares regression were 0.32, 0.34, and 0.41, respectively.

As supporting evidence for the expectation above, the prediction of $D_{\text{Merc}}(\text{min})$ with $\beta = 0.5$ for a narrower range $D_{\text{Merc}}(\text{min}) > -0.20$ is shown in Fig. 7. The scattering of the predictions is smaller than that shown in Fig. 6. This suggests that the prediction accuracy will be improved if the relative population of the training data around $D_{\text{Merc}}(\text{min}) \sim 0$ increases, which is important for optimization. It should be noted here that in the BO, not only the information of the location of the objective function's maximum value in the current set of training data but also the distribution of the expected value and prediction error of the surrogate model across the entire parameter space is employed to predict the next parameters to be examined. Therefore, the training data of $D_{\text{Merc}}(\text{min}) < 0$ as well as those of $D_{\text{Merc}}(\text{min}) \geq 0$ are utilized for better prediction in BO with GPR.

Finally, to determine the type of magnetic configuration that tends to have a large plasma volume and good MHD stability, we arbitrarily selected two examples of magnetic configurations from the training data, as shown in Fig. 8. These cases correspond to the blue points in Fig. 3. In the case with $V = 25.4 \text{ m}^3$, Shafranov shift, $S = 0.21$, is large at finite β . The distribution of the rotational transformation, ι , changes from the distribution monotonically increasing at zero β to the one which has a local minimum around $r = 0.75a$.

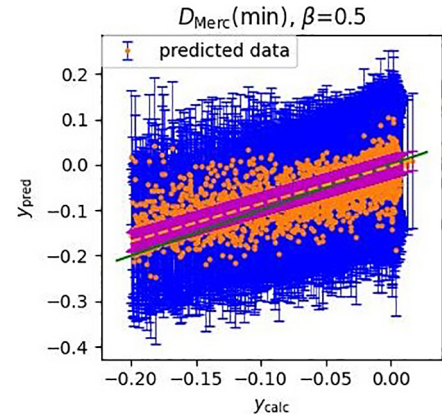


Fig. 7. Predicted and calculated values of $D_{\text{Merc}}(\text{min})$ with $\beta = 0.5$ for the data $D_{\text{Merc}}(\text{min}) > -0.20$. The meaning of the lines and points are the same as in Fig. 4. The NRMSE was 0.10, and the error of predictions obtained by linear least-squares regression was 0.02.

Therefore, the magnetic shear has a weak effect on the MHD stability for finite- β equilibrium, but the deep magnetic well in the entire plasma volume contributes to realize $D_{\text{Merc}} > 0$. On the other hand, in the case with $V = 52.5 \text{ m}^3$, the ι distribution monotonically increases similarly to one of LHD. The

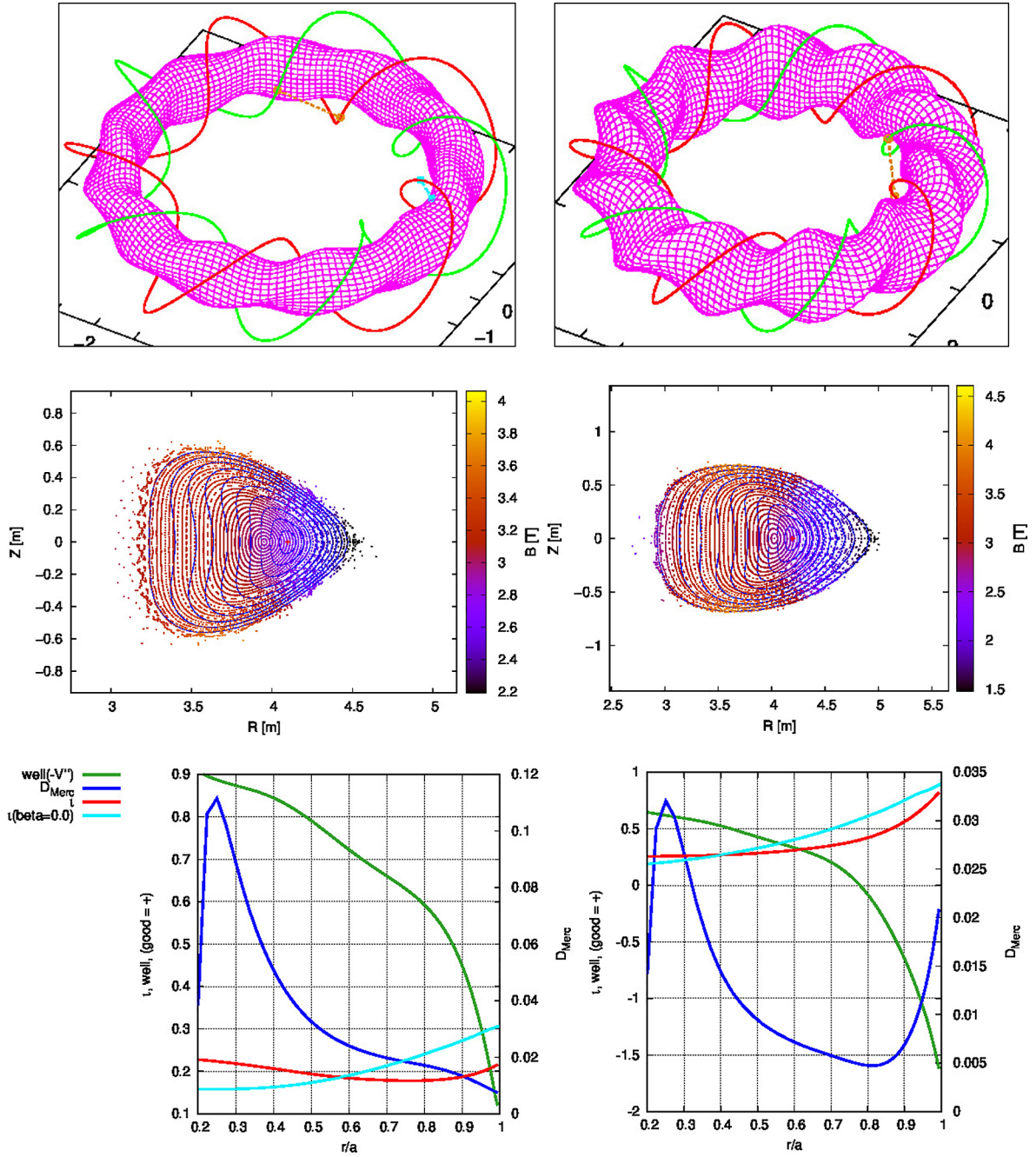


Fig. 8. Samples of helical coils and magnetic configurations, Poincaré plots of cross sections, and the corresponding D_{Merc} values. The images in the left column represent the case with $V = 25.4 \text{ m}^3$ and $\beta = 1.5\%$, whereas those in the right column correspond to the case with $V = 52.5 \text{ m}^3$ and $\beta = 1.5\%$. In figures shown at the bottom, the radial profiles of rotational transform (t) and the magnetic well ($-V''$) are also displayed.

magnetic well is negative at $r > 0.8a$, but the magnetic shear is strong in the outer region, which contributes to good MHD stability. Certain distinctive characteristics emerged among the MHD-stable configurations, although little variation in the helical coil winding was applied in the present study.

4. Summary

The predictions of V , S , and $D_{\text{Merc}}(\min)$ using GPR with machine learning captured the tendencies of the training data. Therefore, the dependence of the expected values of

each objective function on the coil-shaped parameters could be estimated using the trained GPR as a surrogate model. The prediction performance of $D_{\text{Merc}}(\min)$ was worse than those of the two objective functions. The plasma volume and Shafranov shift were scalar data per configuration; however, $D_{\text{Merc}}(\min)$ was more difficult to predict than the other two because they were the minimum values for the entire radial direction, and thus, involved a global trend in equilibrium. We consider that the prediction accuracy of $D_{\text{Merc}}(\min)$ can be improved by increasing training data with $D_{\text{Merc}}(\min) \gtrsim 0$ and re-learning the dependence on parameters while search-

ing for candidate points by a surrogate model. This study demonstrates the feasibility of estimating the objective functions of a fusion reactor design using GPR. However, the variance of each prediction obtained by GPR was still considerable, in part because of the inability to utilize a significant proportion of the data for training owing to insufficient GPU memory. To address this issue, it is essential to improve the implementation of GPR code in Python. We plan to introduce Bayesian optimization to efficiently investigate the Pareto-optimal solutions. The magnetic configurations obtained within the scanned parameter range in this study exhibited less variation than expected. Therefore, we plan to expand the range of parameters and vary the fixed parameters in the present study. To optimize plasma confinement performance and MHD stability, it is necessary to apply indices for turbulent and neoclassical transport as objective functions. Because we have established a general method for GPR here, adopting a new objective function is not difficult, given that the computational cost for the objective function is not high. By integrating these improvements, we will create surrogate models for objective functions in GPR and establish a method to solve multi-objective optimization problems.

Acknowledgments

This work was partially supported by JSPS KAKENHI (grant number: 21K03517).

- [1] T. Goto *et al.*, Nucl. Fusion **59**, 076030 (2019).
- [2] T. Goto *et al.*, Plasma Fusion Res. **16**, 1405085 (2021).
- [3] F. Warmer *et al.*, IEEE Trans. Plasma Sci. **44**, 1576 (2016).
- [4] Y. Narushima *et al.*, Plasma Fusion Res. **15**, 1405076 (2020).
- [5] D.J. Strickler *et al.*, Fusion Sci. Technol. **41**, 107 (2002).
- [6] M. Landreman, Nucl. Fusion **57**, 046003 (2017).
- [7] P. Helander, Rep. Prog. Phys. **77**, 087001 (2014).
- [8] H. Yamaguchi, Nucl. Fusion **59**, 104002 (2019).
- [9] D.A. Gates *et al.*, Nucl. Fusion **57**, 126064 (2017).
- [10] C. Zhu *et al.*, Plasma Phys. Control. Fusion **60**, 065008 (2018).
- [11] H. Yamaguchi *et al.*, Nucl. Fusion **61**, 106004 (2021).
- [12] A. Iiyoshi *et al.*, Nucl. Fusion **39**, 1245 (1999).
- [13] H. Yamada *et al.*, Plasma Phys. Control. Fusion **43**, A55 (2001).
- [14] H. Tsutsui *et al.*, IEEE Trans. Appl. Supercond. **26**, 4901704 (2016).
- [15] S.P. Hirshman *et al.*, Comput. Phys. Commun. **43**, 143 (1986).
- [16] S. Yabumoto *et al.*, Plasma Fusion Res. **18**, 1403018 (2023).
- [17] M. Landreman, J. Plasma Phys. **86**, 905860510 (2020).
- [18] D. Panici *et al.*, Plasma Phys. **89**, 955890303 (2023).
- [19] D.P. Kingma and J.L. Ba, arXiv preprint arXiv: 1412.6980 (2014).
- [20] B. Shahriari *et al.*, Proc. IEEE **104** (IEEE, 2016) 148–175.
- [21] H. Imamura and K. Matsui, *Bayesian Optimization: Fundamentals and Practice of Adaptive Experimental Design*, H. Ohtsuka, (Kindai kagaku sha Co., Ltd, Tokyo, 2023) Section 1.3. [in Japanese]