Application of multiple regression for sensitivity analysis of helium line emissions to the electron density and temperature in Magnum-PSI

Shin Kajita,[1, a)] Daisuke Nishijima,[2] Keisuke Fujii,[3] Gijs Akkermans,[4] and Hennie van der Meiden[4]

[1)] *Institute of Materials and Systems for Sustainability, Nagoya University, Nagoya 464-8603, Japan*

[2)] *Center for Energy Research, University of California, San Diego, 9500 Gilman Dr., La Jolla, CA 92093-0417, USA*

[3)] *Graduate School of Engineering, Kyoto University, Kyoto 615-8540, Japan*

[4)] *NWO Institute DIFFER, Dutch Institute for Fundamental Energy Research, De Zaale 20, 5612 AJ Eindhoven, The Netherlands*

(Dated: 7 April 2021)

Helium line intensities have been utilized to measure the electron density, n_e, and temperature, T_e, by comparing measured line intensities to a collisional-radiative model (CRM). In this study, we use multiple regression analysis to train a model of the helium line intensities and n_e/T_e obtained from a Thomson scattering system in the linear plasma device Magnum-PSI; based on the trained model, we predict n_e and T_e from line intensities. We show that this method can also obtain radial profiles of n_e and T_e. We discuss appropriate selections of line pairs for the prediction based on the multiple regression analysis. A big advantage of this method against the standard technique using CRM is that modeling of atomic population distributions is not required, which sometimes needs to take into account various effects such as radiation trapping, transport of helium atoms in metastable states, etc.

I. INTRODUCTION

Line emissions from helium (He) atoms have been utilized to measure the electron density, n_e, and temperature, T_e, in various fusion devices[1–4]. The basic principle of the measurement is to fit a relative population distribution obtained from a collisional radiative model (CRM) to a measured one[5]. This method has also been used in various linear devices, and comparisons have been made to other diagnostics such as an electrostatic probe and laser Thomson scattering (TS)[6–11].

For the optimization process of CRM calculations, in addition to the dependence of the relative population distribution on n_e and T_e, it is sometimes important to take into account several other effects including high energy electrons[6], radiation trapping[12], plasma fluctuations[13,14], and transport of He atoms in metastable states[14]. However, it is not straightforward to model these effects inclusively. For instance, concerning the effect of radiation trapping, various investigations have been conducted in terms of neutral He density and temperature[10,11], the radius and radial profile of the optical escape factor[9,15]. In particular, it is not easy to assess the influence far from the plasma column center in a linear plasma device, because the emissions from the central region significantly disturb the population distribution at the edge.

Recently, Nishijima and his colleagues have applied a machine learning method to the relations between He line intensities and n_e/T_e, and successfully reproduced radial profiles of n_e and T_e from optical emission spectroscopy (OES) data[16]. This method requires another

[a)] Electronic mail: kajita.shin@nagoya-u.jp

reliable diagnostic tool, but does not require any sophisticated modeling of population distribution. Since n_e was limited up to $\sim 4 \times 10^{18}$ m^{-3} in Ref. 16, it is of interest to check the validity of the method in a higher n_e range, covering the divertor strike point region in fusion devices. Although a machine learning technique without a physics backbone cannot compete with a modeling that is able to treat all the relevant physics correctly, it can be a useful tool when the physics has yet to be fully understood. Also, the difference between machine learning and physics-based methods can give us clues to understand the physics further.

In this study, we analyze OES data collected at a higher n_e range of $10^{19} - 10^{21}$ m^{-3} in the Magnum-PSI linear device using multiple regression analysis to predict n_e and T_e. Previously, in Magnum-PSI, it was found from the comparison between the OES data and CRM that n_e and T_e deduced from the OES were sometimes not consistent with those from TS[11]. In this study, in addition to a line of sight (LoS) observing the center of the plasma column, we will try to deal with radial profile data and discuss the robustness and limitation of the OES data. Moreover, based on the multiple regression analysis, selection of lines appropriate for the prediction of n_e and T_e is discussed.

II. PREPARATION

A. Data set

In this study, the data set in Ref. 11 that had 24 discharges at different discharge currents and gas pressures is used. The magnetic field strength was 1.2 T. Details of the experimental device and experimental setup can be found in Refs. 11 and 17; here, a short explanation of the setup is provided. Pure He plasmas were produced in the linear plasma device Magnum-PSI. Figure 1 shows a schematic representing the field of views of the OES and TS seen from the target to the source. The second harmonic of Nd:YAG laser pulses (532 nm) pass through the plasma from the bottom to the top of the device[18]. The laser TS signal is collected from a side field of view and is detected by a high etendue transmission grating spectrometer that equips with an intensified charge coupled device. The Rayleigh peak, which is much narrower than TS, can be separated from the TS signals. The signal intensity is calibrated by Rayleigh scattering, enabling to measure n_e, while T_e is evaluated from the spectrum broadening. The minimum measurable n_e and T_e are 1×10^{17} m^{-3} and 0.07 eV, respectively. The radial profiles of n_e and T_e can be measured along the laser path. In this study, we assume the Thomson scattering system gives unbiased n_e and T_e data.

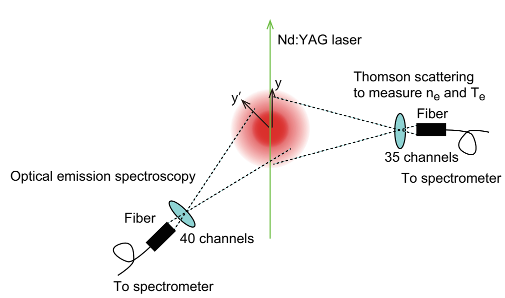

FIG. 1. A schematic representing the field of views of the OES and TS seen from the target to the source. Although the coordinates are different between TS (y) and OES (y'), we will treat the data under the assumption that the parameter variation in the azimuthal direction was not significant for simplicity.

Figure 2 shows a typical observed emission spectrum. The wavelength coverage of the spectrometer used in this study is $\sim$165 nm in a single acquisition. At each plasma condition, two spectra were taken at two wavelength ranges: $\sim$365-530 nm and $\sim$660-825 nm. Note

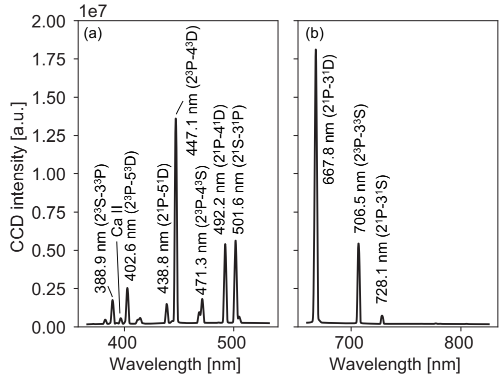

FIG. 2. A typical emission spectrum from the Magnum-PSI device used in this study.

TABLE I. The nine line intensities used in this study. The wavelengths and upper/lower states for the transitions are shown.

Index	Wavelength [nm]	Upper state	Lower state
1	728.1 nm	3^1S	2^1P
2	706.5 nm	3^3S	2^3P
3	501.6 nm	3^1P	2^1S
4	388.9 nm	3^3P	2^3S
5	667.8 nm	3^1D	2^1P
6	492.2 nm	4^1D	2^1P
7	447.1 nm	4^3D	2^3P
8	438.8 nm	5^1D	2^1P
9	402.6 nm	5^3D	2^3P

that the strongest visible line at 587.6 nm was not measured in this study, as it often saturates when it is measured with other lines. And the line emission at 471.3 nm was not used, because there was an overlap with another line probably from some impurity. Other emission lines observed with a decent intensity are included in the analysis. Table I describes the wavelengths and transitions of the observed nine He atomic lines in the wavelength range of 388-728 nm. Those lines were observed by a Czerny-Turner type spectrometer. The intensities have been calibrated using a standard lamp; the calibration is, in principle, not necessary for the method used in this study. The axial (in the direction parallel to the magnetic field) position of the field of view for the OES was the same as the TS measurement, but it was rotated 135 degree in the clockwise direction from the LoS of the TS system, as shown in Fig. 1. The number of fiber channels was 40 for OES and 35 for TS. We will treat the data in the coordinate system shown in Fig. 1 with y for TS and y' for OES under the assumption that the parameter variation in the azimuthal direction was not significant for simplicity.

Figure 3(a) shows the emission profile of the nine lines in a logarithmic scale, and Fig. 3(b) shows the relative emission profiles normalized to the intensities at $y' = 0$ mm in a linear scale. Figure 3(c) shows the radial profiles of n_e and T_e measured by TS. In this study, we split radial profiles to three regions: core ($r < 3$ mm), transition region ($3 < r < 7$ mm), and periphery ($r > 7$ mm), where r is the distance from the center of the plasma column. The profile at 501.6 nm has a relatively strong intensity at the periphery similar to the other devices[10]. This is because the upper state of the transition at 501.6 nm is 3^1P, which is also associated with a resonance line at 53.7 nm (1^1S–3^1P), and the photoexcitation by the radiation from the center can increase the population at the periphery[19,20].

Figure 4 shows the distributions of T_e and n_e that will be used for the analysis in this

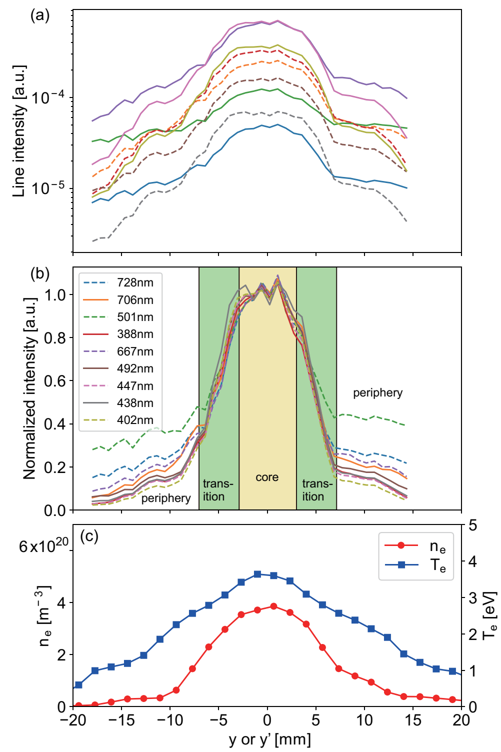

FIG. 3. (a) The emission profile of the nine lines in a logarithmic scale, (b) the relative emission profiles normalized to the intensities at $y' = 0$ mm in a linear scale, and (c) the radial profiles of n_e and T_e measured by TS.

study; different markers represent the three radial regions (core, transition, and periphery). In the core and transition regions, n_e and T_e are mainly in the ranges of 10^{20}-10^{21} m^{-3} and 0.2-3 eV, respectively. In the periphery, the density is lower than the core and transition region, and n_e and T_e are mainly in the ranges of 10^{19}-10^{20} m^{-3} and 0.2-2 eV, respectively.

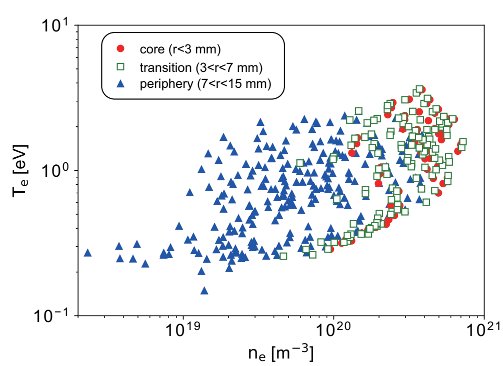

FIG. 4. The distributions of T_e and n_e that will used in this study from the three regions (core, transition, and periphery).

B. Methods

In this study, we mainly use multiple regression analysis (`LinearRegression` in Scikit-learn[21]) and partially use a non-linear model, Kernel ridge regression (`KernelRidge` in Scikit-learn). The main reason for using a linear model is that we can interprete the coefficients in an easy manner.

In the regression analysis, the relations between n_e or T_e from TS and line intensities are trained. The TS data was interpolated to the corresponding OES radial position. Because only the relative population distribution is important, the intensities are normalized to the sum of the used ones as

$$I_i = \frac{\iota_i}{\sum_j \iota_j}, \tag{1}$$

where ι_i is the line intensity with the index number of i, which is shown in table I. Then, we take the logarithm for all the above values, and they are used for the input values of the model. In other words, in the multiple regression analysis, we assume that n_e and T_e can be expressed in the following form:

$$n_e \text{ or } T_e = C_0 I_1^{C_1} I_2^{C_2} \cdots I_n^{C_n}, \tag{2}$$

where C_0, C_1, C_2, $\cdots$, and C_N are coefficients. When taking logarithm on both sides, Eq. (2) becomes

$$\log n_e \text{ or } \log T_e = \log C_0 + C_1 \log I_1 + C_2 \log I_2 + \cdots + C_n \log I_n. \tag{3}$$

Although there is no theoretical support that ne and Te can be expressed using Eq. (2), Nishijima *et al.*[16] have recently shown that the power function model using the relation in Eq. (2) can be well used for the prediction of n_e and T_e in the ionizing plasmas in the PISCES-A linear device. Thus, in this study, we mainly use the linear model, and the deviation from Eq. (2) is discussed using a non-linear model. First, in Sec. III, among the available 24 discharge data, 2/3 (16 discharges) of them are selected randomly and used for training, and remained 1/3 (8 discharges) is used for testing.

III. MULTIPLE REGRESSION ANALYSIS

First, we use all the radial profile data at once for training and testing. Figure 5(a,b) shows predicted n_e and T_e, respectively, as a function of the measured TS values. As mentioned in Sec. II B, randomly selected two-thirds of available discharges were used for training and remaining discharges were used for testing. We repeated the random selection three times and plotted in Fig. 5(a,b) with different markers. The scattering of predicted n_e is not small and has a range from 10^{19}-10^{20} m^{-3} at TS-measured $n_e \approx 2\times10^{19}$ m^{-3}. The scattering of the predicted T_e also has a wide range, and the maximum scattering is roughly a factor of five.

To specify the origin of the scattering, the data set was separated to the three regions (core, transition, and periphery), and training and testing were performed in the three regions separately. Figure 6(a,b) plots the predicted n_e and T_e, respectively, as a function of measured TS values for the three regions (core: circles, transition: triangles, and periphery: crosses). The blue and red colors represent the different sets of data used for training and testing. It is seen that the residual errors are relatively small in the core region, while they gradually increase with increasing radius.

In Table II, the multiple correlation coefficient, R, and the residual errors, e, are presented. In this study, e is defined as

$$e = \frac{1}{n} \sum_j \frac{|X_j - x_j|}{X_j}, \tag{4}$$

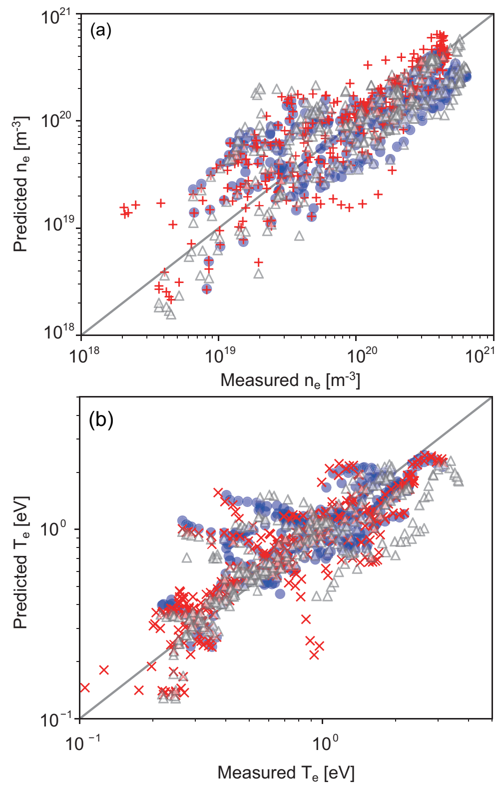

FIG. 5. Predicted (a) n_e and (b) T_e of test data as a function of TS-measured n_e and T_e, respectively. All the radial profile data were analyzed together. Different markers represent the results using randomly chosen three different set of data for training/testing.

TABLE II. The multiple correlation coefficient, R, and the residual errors assessed from the multiple regression analysis in Figs. 5 and 6. Here, *All* in the region of interest means that all the three regions were used for the analysis together, and *Core, Transition,* and *Periphery* mean the three regions were used for the analysis separately.

Region of interest	R coefficient	Residual error (%)
All (n_e)	0.81±0.05	80.3
All (T_e)	0.75±0.11	37.9
Core (n_e)	0.82 ±0.11	19.8
Core (T_e)	0.89 ±0.07	17.0
Transition (n_e)	0.79 ±0.08	29.1
Transition (T_e)	0.74 ±0.15	25.1
Periphery (n_e)	0.55 ±0.13	90.7
Periphery (T_e)	0.45 ±0.19	45.5

where X_j and x_j are measured and predicted values, respectively, j is the index of the data. We repeated random selections of the test/train data 100 times, and the summation was taken for all the test data. Even with including all the radial data at once (Fig. 5), R is 0.81 for n_e and is 0.75 for T_e, indicating that the correlation is not so bad. However, e for n_e is 80%, which is not sufficiently low. When focusing on the core region, both R and e are improved. In particular, e is reduced to less than 20% for both n_e and T_e. As shifting to the outer regions, R gradually decreases and e increases. Thus, the errors are found to originate mainly from the periphery in the current data set.

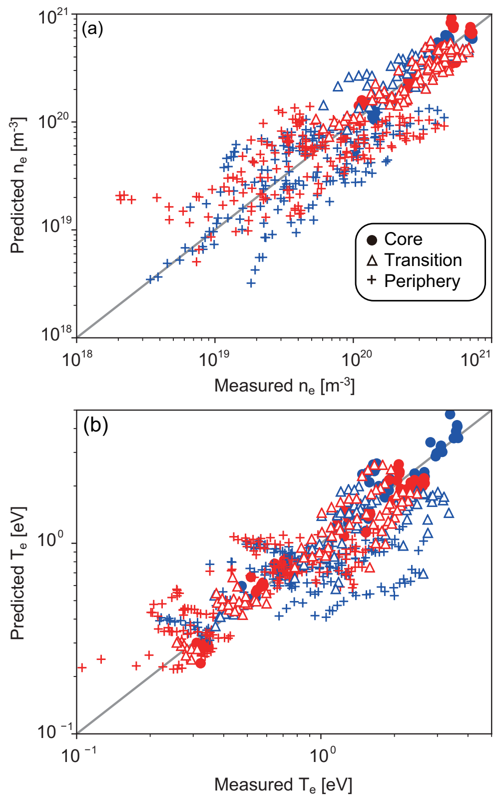

FIG. 6. Predicted (a) n_e and (b) T_e of the test data as a function of TS-measured n_e and T_e, respectively. The training and testing were performed in the three regions separately. Different markers represent the different regions (core: circles, transition: triangles, and periphery: crosses), and blue and red colors correspond to randomly chosen two different sets of data for training/testing.

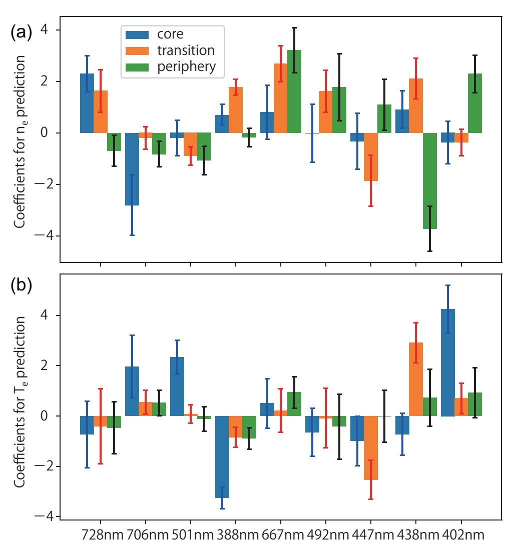

FIG. 7. Obtained coefficients of the nine lines for (a) n_e and (b) T_e predictions at the three different regions. We repeated the analysis 100 times and the average and standard deviation of the values were plotted as bars and error bars, respectively.

Figure 7(a,b) shows the obtained coefficients of the nine lines for n_e and T_e predictions, respectively, at the three different regions. Here, to improve statistics from Fig. 6, where calculations were done only twice, we repeated the analysis with random selection of test/train data 100 times, and the average and standard deviation values are plotted as bars with error bars. It is found that the coefficient values are different in the three different regions. For example, the coefficients of 728.1 and 438.8 nm for the n_e prediction alter from positive to negative values as shifting from the core to the periphery. One of the potential causes is the effect of radiation transport. Previously, it was revealed from both experiments in the PISCES-A device and a ray tracing simulation that the intensity of singlet lines is enhanced especially at the periphery of the plasma column, because of the absorption of resonance lines emitted from the brighter plasma central region, while the enhancement of the triplet line intensity is much smaller over the whole profile[19]. For example, the population densities in 3^1P and 3^1S states increased by three orders and an order of magnitude, respectively, due to the absorption of resonance lines. This means that the absorption effect is more dominant than electron-impact excitation processes to populate the singlet states. As shown in Fig. 3(b), the 501.6 nm and 728.1 nm lines have relatively higher intensities at the periphery, because the upper states of these line are 3^1P and 3^1S, respectively. Thus, these lines are more sensitive to the absorption effect and are less sensitive to electron-impact excitation, i.e. n_e and T_e, especially at the periphery. On the other hand, lines that are less sensitive to absorption are thought to be more beneficial for the prediction of n_e and T_e at the periphery. In addition to the value itself, the relation between the value and the scattering (error bar) is also important. If the scattering is comparable or larger than the value itself, it is suggested that the value is not stable and does not have a clear sensitivity to the parameter. Concerning the coefficient of, e.g., 728.1 nm for T_e prediction, the error bars are larger than the coefficient values, suggesting that the line intensity is not sensitive to the T_e prediction. We will use Fig. 7(a,b) later for practical selection of lines.

Although the error is large at the periphery region, the fact that R is $\sim$0.5 suggests the predicted values have correlation with the measured values to some extent. Here we try to reconstruct the radial n_e and T_e profiles from the OES data. Figure 8(a,b) shows those from OES and TS. Here, one discharge was selected for test data and all the other data were used for training. For assessing the errors of n_e and T_e, we repeated the analysis (1/3 test data and 2/3 training data) used in Fig. 6(a,b) 100 times, and standard deviations of the TS n_e and T_e of the test data that have close predicted n_e and T_e ($< \pm 10\%$) values, respectively, were used. While the errors are relatively large, the predicted radial profiles of n_e and T_e from OES agree well with those of TS. It should be noted that OES coupled with CRM was not able to reproduce the radial profiles from TS[11].

IV. SELECTION OF LINES

In Sec. III, nine lines were used for the analysis as in the previous study[11], where the quality of the n_e and T_e predictions using CRM was deteriorated when reducing the number of lines. In this section, we assess the importance of each line and discuss the best selections of necessary lines.

To select lines, let us go back to Fig. 7. For the first step, we focus on the core region in this study. We can eliminate lines that have little sensitivity to both n_e and T_e. It is apparent that lines at 728.1 and 706.5 nm have sensitivity to n_e, and lines at 706.5, 501.6, 388.9, and 402.6 nm have sensitivity to T_e, while lines at 667.8, 492.2, 447.1, 438.8 nm seem not to be so sensitive to both n_e and T_e. We assess the performance for six cases (i)-(vi) with different line selections shown in Table III. Case (i) uses all the nine lines, case (ii) eliminates two lines (667.8 and 492.2 nm) from case (i), case (iii) eliminates two more lines (447.1 and 438.8 nm) from case (ii), and cases (iv)-(v) are the cases used previously[14]. The three lines used in case (vi) are the popular lines that have been used frequently[5,8–10]. In case (v), the line at 501.6 nm, which is sensitive to radiation trapping[20], is added to

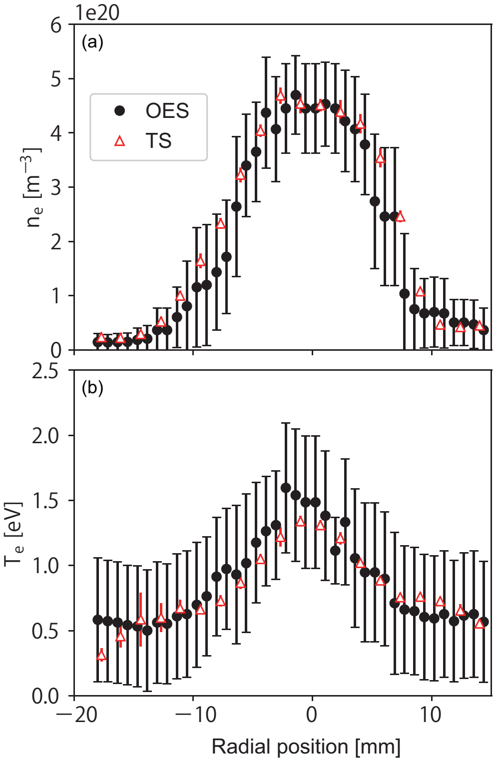

FIG. 8. Predicted radial profiles of (a) n_e and (b) T_e from OES (closed circles) compared with measured TS values (open triangles).

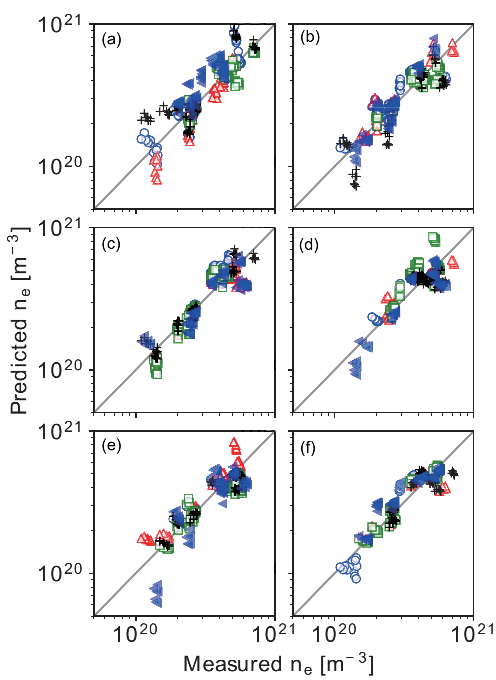

FIG. 9. (a-f) Predicted n_e in the core region as a function of the measured n_e for different line selection cases (i)-(vi), respectively. Different markers represent randomly chosen five different set of data for training/testing.

TABLE III. Six cases (i)-(vi) with different lines used for the analysis.

Case	Number of lines	Wavelengths (nm)
(i)	9	728.1, 706.5, 501.6, 388.9, 667.8, 492.2, 447.1, 438.8, 402.6
(ii)	7	728.1, 706.5, 501.6, 388.9, 447.1, 438.8, 402.6
(iii)	5	728.1, 706.5, 501.6, 388.9, 402.6
(iv)	5	728.1, 706.5, 501.6, 667.8, 447.1
(v)	4	728.1, 706.5, 501.6, 667.8
(vi)	3	728.1, 706.5, 667.8

case (vi). In case (iv), another line at 447.1 nm, which is sensitive to the recombining component[14], is added to case (v).

Figure 9(a-f) plots the predicted n_e in the core region as a function of the TS-measured n_e for cases (i)-(vi), respectively. Again, randomly chosen 2/3 (16 discharges) data from the data set was used for training and the remained 1/3 (8 discharges) data was used for testing. Different markers in Fig. 9 represent five sets of training and testing. It is seen that the predicted values are almost consistent with TS values, suggesting that the method can be used for all the cases. It is interesting to note that the quality of the prediction is good even with three lines for n_e, i.e. case (vi). Figure 10(a-f) plots the predicted T_e in the core region as a function of the measured T_e for cases (i)-(vi), respectively. While the results are almost consistent for cases (i)-(iii), the quality of the prediction is worth in cases (iv)-(vi), especially, at $T_e > 2$ eV.

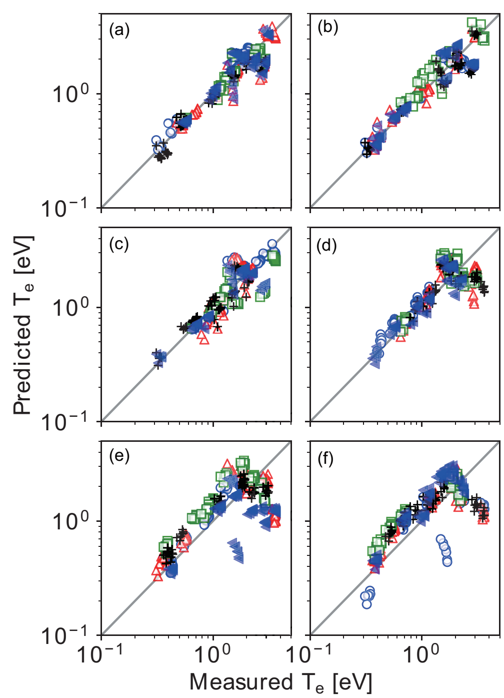

FIG. 10. (a-f) Predicted T_e in the core region as a function of the TS-measured T_e for different line selection cases (i)-(vi), respectively. Different markers represent randomly chosen five different set of data for training/testing.

Figure 11 summarizes the residual errors in n_e and T_e for the different line selection cases (i)-(iv). The error of n_e slightly decreases when the number of lines decreases from nine to three, but it does not depend strongly on the selection of the lines and is in a range of

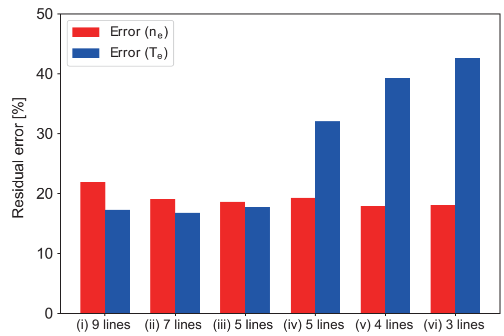

FIG. 11. The residual errors in n_e and T_e for cases (i)-(vi).

17-22%. On the other hand, the T_e error is nearly constant at ∼15% for cases (i)-(iii); it increases to ≈ 30% in case (iv) and further increased to ≈ 40% in cases (v) and (vi). This might be caused by the usage of the linear model for training the data. Thus, we examined a non-linear fit model (Kernel ridge regression) for cases (iv)-(vi). As shown in Fig. 12, the predicted T_e using the Kernel ridge regression still deviates from the TS-measured values, in particular, at $T_e > 2$ eV, as with the linear model. The residual errors of cases (iv)-(vi) are 28.1, 37.2, and 40.9%, respectively. The error was slightly improved for case (iv), but almost no improvement for cases (v) and (vi).

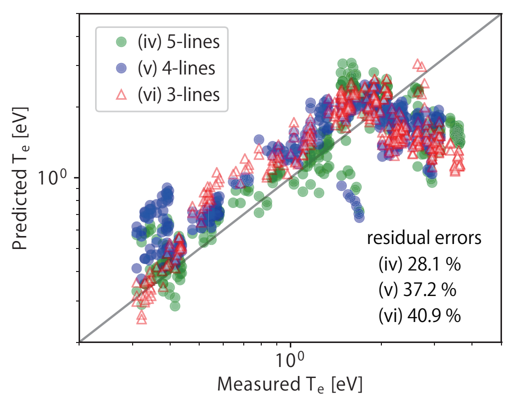

FIG. 12. Predicted T_e using the Kernel ridge regression plotted as a function of the TS-measured T_e for cases (iv)-(vi).

Fig. 13(a) plots the line intensity ratios of 728.1/706.5 nm vs. 667.8/728.1 nm with T_e as the color of markers. Here, only the core region data are used. The line intensity ratios 728.1/706.5 nm and 667.8/728.1 nm are sensitive mainly to T_e and n_e, respectively. However, high and low T_e data points overlap at 728.1/706.5 nm ∼0.06 and 667.8/728.1 nm ∼24, indicating that T_e is a multivalued function of this line intensity ratio pair. This may be explained by the mixture of ionizing and recombining components[9], since there is a minimum of the 728.1/706.5 nm ratio around the boundary between ionizing (higher T_e) and recombining (lower T_e) plasmas. While the addition of 501.6 nm to the three lines (728.1, 706.5, and 667.8 nm) does not reduce the deviation at $T_e > 2$ eV [see Fig. 10 (e) and (f)], the fit is improved by the further addition of 447.1 nm to the above four lines [see Fig. 10 (d)]. In Fig. 13(b), a ratio 402.6/501.6 nm is plotted against 501.6/388.9 nm with T_e as the color of markers. Here, the two ratios consist of the three most sensitive lines to T_e in the core, as shown in Fig. 7. The lower T_e region exists at the bottom and the higher T_e region exists at the upper right of the figure. In Fig. 13(b), the mixture of different T_e regions is less than that in Fig. 13(a). Thus, as is suggested in Fig. 10(a-c), the line selection of cases (i)-(iii) is more powerful to separate the contributions from ionizing and recombining components compared to case (iv)-(vi).

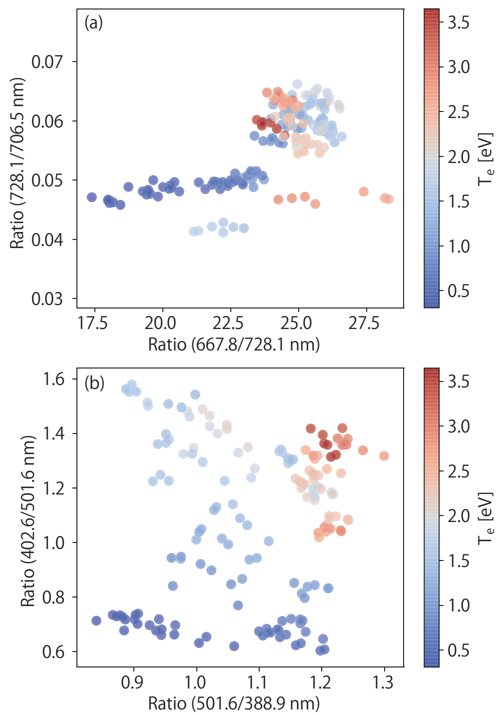

FIG. 13. Distributions of the line intensity ratios in the core region: (a) 728.1/706.5 nm vs. 667.8/728.1 nm and (b) 402.6/501.6 nm vs. 501.6/388.9 nm under various discharge conditions. The color of the marker represents TS-measured T_e values. This kind of plot is useful to find out if any overlaps of different T_e values exist in a line ratio pair.

In this section, we discussed the line selection, focusing on the data in the core region. However, as expected from Fig. 7, the best selection of lines can be different in the transition and periphery regions. Future work will focus on the development of robust tools for prediction that can be applied to various regions and other devices.

V. CONCLUSIONS

We showed that the machine learning methods can be used to predict the electron density (n_e) and temperature (T_e) from helium (He) line intensities if there is another reliable measurement method for n_e and T_e that can be used for training. Multiple regression analysis was applied to train the data set of optical emission spectroscopy (OES) data and n_e/T_e from Thomson scattering (TS) in the linear plasma device Magnum-PSI. The intensity of nine He I lines was used for the analysis: 728.1, 706.5, 501.6, 388.9, 667.8, 492.2, 447.1, 438.8, and 402.6 nm.

When using all the radial data set at once, the residual errors were large: ≈80% for n_e and ≈40% for T_e. Thus, the radial profile was separated into the core ($r < 3$ mm), transition ($3 < r < 7$ mm), and periphery ($r > 7$ mm) regions to identify the region, the data of which caused the large residual errors. It was found that the data at the periphery caused the large residual errors, probably because the population densities of, mainly, singlet states at the periphery are altered by absorption of photons transported from the core to the periphery. Based on the coefficients derived from the multiple regression analysis of the data in the core region, it was found that a satisfactory prediction of n_e and T_e requires, at least, the following five lines at 728.1, 706.5, 501.6, 388.9, and 402.6 nm. With this line selection, the error, originating from the mixture of ionizing and recombining components, is reduced.

This study demonstrated that machine learning can be a powerful tool for OES measurement to predict n_e and T_e in case there is another diagnostic for training a model. This method has an advantage, which does not require any complicated modeling for population distribution of He atoms, e.g. the effects of radiation trapping, neutral density/temperature, transport of metastable state atoms. In this study, we focused on the data from one device, which covers the n_e and T_e ranges of 10^{19}-10^{21} m^{-3} and 0.2-3 eV, respectively. In this study, we used the laser Thomson scattering system, which is reliable with small measurement errors, as an independent diagnostic to train the model. Even if the accuracy of an independent measurement system is low, we can make a regression analysis with enough training data unless those are biased. The trained model is expected to predict n_e and T_e with a better accuracy than the independent measurement system, since the line intensity is usually observed with a good signal to noise ratio. In the future, it is of interest to investigate the property of OES data from various experimental devices, including other linear devices as well as tokamak and helical fusion devices.

VI. ACKNOWLEDGEMENT

This work was supported in part by a Grant-in-Aid for Scientific Research 19H01874 and Fund for the Promotion of Joint International Research 17KK0132 from the Japan Society for the Promotion of Science (JSPS).

[1] M. Griener, E. Wolfrum, M. Cavedon, R. Dux, V. Rohde, M. Sochor, J. M. Munoz Burgos, O. Schmitz and U. Stroth: Review of Scientific Instruments **89** (2018) 10D102.
[2] M. Goto and K. Sawada: Journal of Quantitative Spectroscopy and Radiative Transfer **137** (2014) 23 .
[3] M. Agostini, P. Scarin, R. Milazzo, V. Cervaro and R. Ghiraldelli: Review of Scientific Instruments **91** (2020) 113503.
[4] S. Ma, J. Howard, B. D. Blackwell and N. Thapar: Review of Scientific Instruments **83** (2012) 033102.
[5] M. Goto: J. Qunantitative Spectroscopy and Radiative Transfer **76** (2003) 331.
[6] S. Sasaki, S. Takamura, S. Watanabe, S. Masuzaki, T. Kato and K. Kadota: Rev. Sci. Instrum. **67** (1996) 3521.
[7] R. F. Boivin, J. L. Kline and E. E. Scime: Physics of Plasmas **8** (2001) 5303.
[8] Y. Iida, S. Kado, A. Okamoto, S. Kajita, T. Shikama, D. Yamasaki and S. Tanaka: J. Plasma Fusion Research SERIES **7** (2006) 123.
[9] S. Kajita, N. Ohno, S. Takamura and T. Nakano: Physics of Plasmas **13** (2006) 013301.
[10] D. Nishijima and E. M. Hollmann: Plasma Physics and Controlled Fusion **49** (2007) 791.
[11] S. Kajita, G. Akkermans, K. Fujii, H. van der Meiden and M. C. M. van de Sanden: AIP Advances **10** (2020) 025225.
[12] T. Fujimoto, *Plasma spectroscopy* (Clarendon press, Oxford, 2004).
[13] F. B. Rosmej, N. Ohno, S. Takamura and S. Kajita: Contrib. Plasma Phys. **48** (2008) 243.
[14] S. Kajita, K. Suzuki, H. Tanaka and N. Ohno: Physics of Plasmas **25** (2018) 063303.
[15] Y. Iida, S. Kado and S. Tanaka: Physics of Plasmas **17** (2010) 123301.
[16] D. Nishijima, S. Kajita and G. R. Tynan: Review of Scientific Instruments **92** (2021) 023505.
[17] J. Rapp *et al.*: Fusion Engineering and Design **85** (2010) 1455 .
[18] H. J. van der Meiden *et al.*: Review of Scientific Instruments **83** (2012) 123505.
[19] S. Kajita, D. Nishijima, E. M. Hollmann and N. Ohno: Physics of Plasmas **16** (2009) 063303.
[20] S. Kajita and N. Ohno: Rev. Sci. Instrum. **82** (2011) 023501.
[21] F. Pedregosa *et al.*: the Journal of machine Learning research **12** (2011) 2825.