



HAL
open science

Introducing Machine-Learning in Spectroscopy for Plasma Diagnostics and Predictions

M. Koubiti, M Kerebel

► **To cite this version:**

M. Koubiti, M Kerebel. Introducing Machine-Learning in Spectroscopy for Plasma Diagnostics and Predictions. Journal of Physics: Conference Series, 2023, 2439 (1), pp.012016. 10.1088/1742-6596/2439/1/012016 . hal-04529985

HAL Id: hal-04529985

<https://amu.hal.science/hal-04529985>

Submitted on 2 Apr 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

PAPER • OPEN ACCESS

Introducing Machine-Learning in Spectroscopy for Plasma Diagnostics and Predictions

To cite this article: M Koubiti and M Kerebel 2023 *J. Phys.: Conf. Ser.* **2439** 012016

View the [article online](#) for updates and enhancements.

You may also like

- [Rapid learning of bearing signal pattern using CfCs promoted by a self-attention mechanism](#)
Yanli Yang, Weisheng Pan and Huimin Zhou
- [Advances in reference materials and measurement techniques for greenhouse gas atmospheric observations](#)
Paul J Brewer, Jin Seog Kim, Sangil Lee et al.
- [Rydberg series of \$1^+_{g}\$ and \$1^0_{g}\$ states of the \$\text{Li}_2\$ molecule studied by the promotion model](#)
Hyun-Jin Kim and Chun-Woo Lee

PRIME
PACIFIC RIM MEETING
ON ELECTROCHEMICAL
AND SOLID STATE SCIENCE

HONOLULU, HI
Oct 6-11, 2024

Abstract submission deadline:
April 12, 2024

Learn more and submit!

Joint Meeting of
The Electrochemical Society
•
The Electrochemical Society of Japan
•
Korea Electrochemical Society

Introducing Machine-Learning in Spectroscopy for Plasma Diagnostics and Predictions

M Koubiti and M Kerebel

Aix-Marseille Université-CNRS, PIIM UMR 7345, F-13397 Marseille, France

Mohammed.koubiti@univ-amu.fr

Abstract. Artificial Intelligence (AI) and data science techniques are increasingly introduced in physics including plasma physics where Machine Learning (ML) is applied to emission spectroscopy for plasma parameter determination. Recently, the open-access python-based Sickit-Learn ML platform was used to analyze line intensities in the order to infer the plasma electron densities and temperatures for conditions relevant to tokamak divertors. In this paper, we discuss the application of deep-learning (DL) to synthetic line spectra for conditions of magnetic fusion plasmas with hydrogen-deuterium (H-D) mixtures. The idea will be illustrated through application of Artificial Neural Networks (ANN) to spectra of the Balmer- α line emitted by H-D mixtures, the aim being to obtain the isotopic ratios. The objective of our approach is to provide a new method to infer the hydrogen isotopic ratio sufficiently fast that can be exploited for real-time applications. We will demonstrate the proof-of-principle of our method through the application of a TensorFlow DL regression algorithm to theoretical line spectra generated with predetermined parameters.

1. Introduction

Future power plants based on thermonuclear fusion of hydrogen isotopes will certainly be operated with a mixture of deuterium and tritium (D-T) owing to the highest reaction cross section in comparison to those of other fusion reactions. However, because of its radioactivity the tritium content of any device will be limited for safety reasons. In addition to these constraints, it may be also necessary to control the tritium content to optimize the performances of the fusions reactions. Therefore, it will be very useful to map the tritium content inside the power plants for real-time applications such as plasma performance control. However, nowadays experimental devices devoted to the research on fusion by magnetic confinement are operated by either hydrogen, deuterium or with hydrogen-deuterium (H-D) mixtures. There are only few conducted or planned experiments with deuterium-tritium mixtures [1-3]. Usually, one uses the isotopic ratio to characterize the quantity of a given hydrogen isotope in a plasma. To our knowledge, there is no direct technique for the measurement of isotopic ratio but there are indirect techniques. The most used technique relies on the analysis of the Balmer- α line emission of hydrogen isotopes [4-6]. There exists another method called the Residual Gas Analysis (RGA) [7-8] but it is not routinely used. In hydrogen-deuterium (H-D) plasma, one can define the isotopic ratio using one of the two expressions:

$$\eta_H = n_H / (n_H + n_D) \quad \text{or} \quad \eta_D = n_D / (n_H + n_D) \quad (1)$$



Here n_{H} and n_{D} stand for the neutral hydrogen and deuterium densities respectively. The ratios η_{H} and η_{D} satisfy the following condition:

$$\eta_{\text{H}} + \eta_{\text{D}} = 1 \quad (2)$$

In tokamaks, the Balmer- α lines $\text{H}\alpha$, $\text{D}\alpha$ emitted by hydrogen isotopes (H, D) are among the most intense lines and their relative intensities depend on their relative concentrations. These lines are observed in the divertor and Scrape-off-layer regions where several neutral populations co-exist thanks to various recycling mechanisms [9-10]. The typical parameters of these regions are: $n_e \approx 10^{14} \text{ cm}^{-3}$, $T_e \approx \text{few eV}$. For such conditions, the $\text{H}\alpha/\text{D}\alpha$ lines are dominated by Doppler-Zeeman effects as Stark broadening is negligible [11]. Note that such plasma conditions correspond to an ionizing regime where higher members of the Balmer series (such as $\text{H}\beta$, $\text{H}\gamma$, $\text{H}\delta$ and so on) are weak. However, these high- n Balmer lines are observed under recombining regimes ($T_e \leq 1 - 2 \text{ eV}$) and owing to their high sensitivity to Stark broadening they are used for electron density and temperature diagnostics [11]. Fitting experimental spectra of $\text{H}\alpha/\text{D}\alpha$ lines allows the evaluation of both the temperatures and fractions of the neutral populations and the isotopic ratio. Therefore, it is possible to map the content of a given hydrogen isotope using this technique to measurements along different lines of sight. To fit such spectra, at least two neutral populations are supposed to co-exist for each isotope, each population having its own velocity distribution function (VDF). As can be easily understood, this spectroscopic technique relies on the fit of whole line profiles and cannot be used for real-time applications such as plasma controls. In nuclear fusion reactors operated with D-T mixtures, a real-time control of the tritium content must be necessary to ensure good performances and meet safety conditions. This justifies the development of fast methods allowing the determination of the isotopic ratio. The aim of this paper is to demonstrate the proof-of-principle of a fast method combining artificial intelligence with plasma spectroscopy to determine the hydrogen isotopic ratio in tokamaks. We pursue two objectives: first help controlling plasmas through the monitoring of the hydrogen isotope content and second make predictions for future operations from available experimental data. Instead of fitting whole $\text{H}\alpha/\text{D}\alpha$ line spectra, we propose to use only few line features as input for a deep-learning algorithm to determine hydrogen isotopic ratios.

The recent years have seen an increasing research activity combining machine-learning (ML), plasma physics and plasma spectroscopy. For instance, a regression algorithm taken from the Sickit-Learn ML platform [12] was used for the analysis of helium lines to determine the electron density and temperature in the linear divertor simulator Magnum-PSI [13]. The assessment of the results was realized using independent measurements made with a Thomson Scattering system. Other applications of ML can be found in [14-18]. The introduction of ML in plasma spectroscopy is aimed to help getting faster the same information as the physical models with a good accuracy to allow real-time control and monitoring. This paper is organized as follows. In section 2, we describe the physical model on which is based the standard spectroscopic method used to obtain the isotopic ratio. In section 3, we briefly present ML especially deep-learning (DL) with a focus on Artificial Neural Networks (ANN) and its application to the Balmer- α emission line to demonstrate our method's proof-of-principle. Section 4 is devoted to discussion and conclusion.

2. The spectroscopic method and the physical model

For simplicity, we consider plasmas where the dominant gas is deuterium, hydrogen being present with a concentration of up to 25%. We aim here to determine the isotopic ratio η_{H} or η_{D} for plasmas with H-D mixtures. The Balmer- α line is emitted by hydrogen and deuterium neutrals that are present in the tokamak divertor region. Plasma-material interactions are at the origin of the presence of neutrals: hydrogen and deuterium neutrals are produced through chemical and physical sputtering

mechanisms leading to the co-existence of at least two neutral populations. A cold neutral population resulting from dissociation of molecules desorbed from the divertor target having a Maxwellian velocity distribution function (VDF) with a temperature of a fraction of eV to few eVs. The second population originates from reflected atoms, i.e., H^+ or D^+ ions capture electrons when hitting the target and are released as H or D neutrals. Such a population can thermalize before emitting radiation in which case their VDF can be assumed to be a Maxwellian with a temperature of some tens of eVs. A third neutral population with a typical temperature of about 100 eV may exist due to charge exchanges between plasma ions and released neutrals. This third population gives rise to a broad and weak line profile and therefore is not considered in our calculations. The shape of the $H\alpha/D\alpha$ line depends on the angle θ between the magnetic field and the observation directions. We show on **figure 1** normalized profile of the $H\alpha/D\alpha$ lines calculated for a magnetic field of 2 T and two neutral populations with Maxwellian VDFs and temperatures 2 eV and 20 eV. The calculation was done assuming that the cold population represents 75% while the warm population represents only 25%.

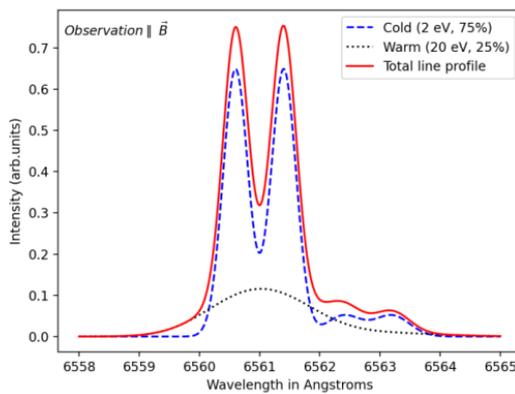


Figure 1. $D\alpha/H\alpha$ line profile calculated for an observation parallel to a magnetic field of 2 T for a gas mixture of 90%D with 10%H. The Total line profile (red solid line) is the sum of two contributions: neutral populations with Maxwellian VDF (blue dashed line: 2 eV) and (black dotted line: 20 eV).

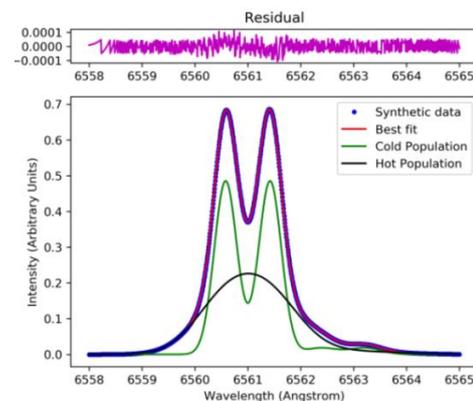


Figure 2. An example of the fit of a $D\alpha/H\alpha$ line profile calculated for an observation parallel to the magnetic field. A least-square minimization algorithm was used to infer various parameters. The known and deduced parameters are summarized in Table 1.

The standard method to determine isotopic ratios is based on the fit of experimental $H\alpha/D\alpha$ line spectra by modeling their theoretical profiles. To illustrate this, we consider the fit of a theoretical profile instead of an experimental one, the advantage being that all the parameters are known, making it easy to assess the optimization technique allowing the spectral fit. An example is shown on **figure 2**. As it can be seen there is a good agreement between the generated spectrum and the fitted profile. The parameters of the best fit are summarized with the known ones in Table 1. The inferred parameters are very close to the real ones. It is this technique which is the standard method used for the hydrogen isotopic determination from the experimental $H\alpha/D\alpha$ line spectra measured in tokamaks.

Table 1. Comparison of the best fit parameters with the real parameters (known parameters) of the synthetic line spectrum. T_C and T_W stand for the temperatures of the cold and warm neutral populations respectively. f_C and f_W stand for their relative fractions in percentages.

Parameters	Known values	Inferred values	Errors
(T_C, f_C)	2 eV ; 55 %	2.015 eV ; 54.998 %	0.744 % ; 0.0042 %
(T_W, f_W)	15 eV ; 45 %	15.111 eV ; 45.002 %	0.742 % ; 0.0051 %
B	2 T	2.091 T	4.56 %
η_H	5 %	5.0007 %	0.01308 %

3. Application of Deep Learning to spectroscopic features of H α /D α line profiles

In Machine Learning, algorithms are used for classification, clustering and regression problems. Classification consists in the identification to which category an object belongs to. Clustering consists in the automatic grouping of similar objects into sets. Regression consists in predicting numerical values as a continuous function for an attribute associated with an object. For each of these issues, there are several algorithms e.g., Support Vector Regression, nearest-neighbors and random forest for regression. For plasma physics applications, we use a regression algorithm as we deal with numerical values of physical quantities. We have used a regression algorithm from the TensorFlow ML platform with the ADAM optimizer [19]. We have used six input features for the Deep-Learning algorithm: the magnetic field strength B , the temperatures T_C and T_W of the neutral populations, and three other features depicted on **figure 3**: a wavelength separation $\Delta\lambda_{H\alpha-\sigma^-(D)} = \lambda_{H\alpha} - \lambda_{\sigma^-(D)}$, and two line intensity ratios $\frac{I_{\text{Dip}(D)}}{I_{\sigma^-(D)}}$ and $\frac{I_{\sigma^+(H)}}{I_{\sigma^-(D)}}$. The target or label data is the fraction of hydrogen isotopic ratio η_H . A preprocessing program was used to extract these spectral features from theoretical profiles computed for H-D mixtures with H representing 1-25%, B in the range 1-5 T, $T_C = 2 \pm 0.2$ eV (20-80%) and $T_W = 20 \pm 2$ eV (80-20%). The extracted features are then transmitted to the regression algorithm. A total of 200 000 theoretical spectra were generated for training. In addition to the input and output neuron layers, the ANN we have used is composed of six hidden layers having each hundred neurons. The number of neurons of the input layer is equal to six in our case. The output layer contains a single neuron as the target is a single numerical quantity (the isotopic ratio).

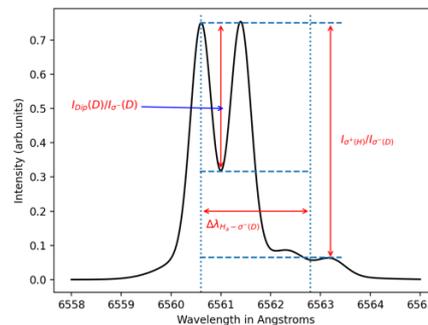


Figure 3. A scheme of Zeeman-Doppler H α /D α line profile indicating the spectroscopic characteristics which are part of the features data for the regression algorithm.

To illustrate the results of the application of the regression algorithm to our H α /D α line spectra, we show on **figure 4** the isotopic ratio (in %) predicted by the DL algorithm versus the real isotopic ratio. For a more realistic situation, we have added a gaussian noise to the synthesized H α /D α line profiles.

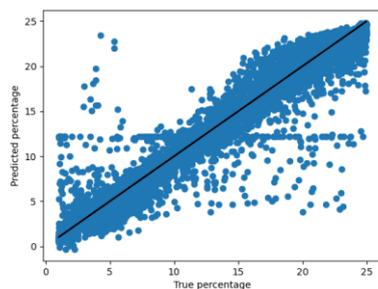


Figure 4. The predicted isotopic ratio as a function of the real (theoretical) value for a training set of 200 000 data points.

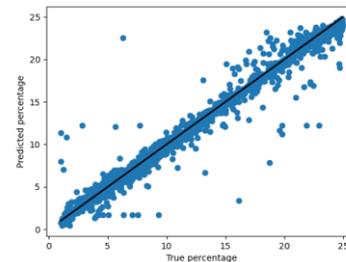


Figure 5. Same as figure 4 but for the test set composed of 20 000 data points.

4. Discussion and conclusion

The isotopic ratios predicted by the deep-learning algorithm for the training set are scattered around the black solid line representing the $y = x$ function in **figure 4**. With respect to the expected values of the hydrogen isotopic ratio η_H , the mean and median errors are respectively 7.6% and 2.4%. Note that the addition of noise to the calculated profiles makes difficult the determination of the position of the peak of the $H\alpha$ line and hence introduces an error in the wavelength separation feature. The consequence of this is that the regression algorithm introduces in turn an uncertainty on the isotopic ratios as shown on **figure 4**. If one excludes outlier data, it can be reasonably considered that there is an overall good agreement between the predicted values of the isotopic ratio and the real values (true values in **figure 4**). It is worth noting that the reasons of such relatively high error values are not due to the DL regression algorithm but are due to the difficulty of the preprocessing program in the determination of some features especially when the spectrum minima cannot be defined accurately. As this point sets limitations to the use of the regression algorithm and affects the accuracy of the predictions, it will be tackled in the near future to improve the proposed technique. For the validation of the method, we have generated a total of 20 000 other theoretical spectra in a similar manner as for the generation of the training set. The results of the DL algorithm are shown on **figure 5** as the predicted values of the hydrogen isotopic ratio plotted as a function of the true values. For the test phase, the mean square error was estimated to be 7.9% while the median error value is 2.6%. One can see from the figure some outlier points for which the predicted values are far from the known values but for the majority of the data, the predicted values match well the known values. The algorithm was then applied to an experimental spectrum from the former Tore-Supra tokamak. The isotopic ratio of hydrogen was estimated to be 3.8% from the spectral fit while the value predicted by the DL algorithm was 4.4% which corresponds to a relative error of about 15%. The results presented here demonstrate that we can use simple spectral features of line spectra such as wavelength separation, dip-to-peak and peak-to-peak intensity ratios instead of fitting whole spectra to determine and predict isotopic ratios in H-D mixtures. Even though limited to H-D mixtures with hydrogen concentrations up to 25%, the results of our study can be extended to the opposite situation where deuterium represents up to 25% and to mixtures where hydrogen represents up to 50%. Our results aim to demonstrate the proof-of-principle of the use of DL combined with spectroscopy to determine the isotopic ratio in H-D mixtures. A more complete discussion can be found in [20]. Application of the presented method to experimental data with other input features such as the plasma pressure, plasma temperature and density as well as other quantities may allow to discover some hidden connections with the isotopic ratio, this can be easily seen from the correlations between the different features and the target data (isotopic ratio) in our case. Another perspective of this work is the extrapolation of the results to D-T mixtures as future DEMO and fusion nuclear power plants (reactors) will be operated with a mixture of deuterium and tritium. Determining the isotopic ratio D/D+T in a fast way will be an important advantage for the operators to control the plasma contents and to be in line with the recommendation of the safety authorities and manage the performances of the reactors.

References

- [1] Keilhacker M, Watkins M L and JET Team 1999 *J. Nucl. Mater.* **266-269** 1-13.
- [2] Hyun-Tae Kim K *et al* 2020 *Nucl. Fusion* **60** 066003.
- [3] Mailloux J *et al* 2022 *Nucl. Fusion* **62** 042026.
- [4] Neverov V S *et al* 2017 *Nucl. Fusion* **57** 016031.
- [5] Neverov V S *et al* 2019. *Nucl. Fusion* **59** 04601.
- [6] Koubiti M *et al* 2002 *Plasma Phys Control Fusion* **44** 261.
- [7] Drenik A *et al* 2017 *Phys. Scr.* **2017** 014021
- [8] Klepper C C *et al* 2010 *Rev Sci Instrum.* **81** 10E104.
- [9] Kubo H *et al* 2014. *Plasma Phys. Control. Fusion* **56** 025010.
- [10] Hey J, Chu C.C, Mertens Ph, Brezinsek S, Unterberg B 2004. *J. Phys. B* **37** 2543.

- [11] Koubiti M and Sheeba R S 2019. *Atoms* **7** 23.
- [12] Pedregosa F *et al* 2011. *The journal of Machine Learning Research* **12** 2825.
- [13] Kajita S *et al* 2021. *Plasma Phys. Control. Fusion* **63** 055018.
- [14] Gurung-López S *et al* 2022. *Monthly Notices of the Royal Astronomical Society* **510** 4525.
- [15] Cianciosa M *et al* 2020. *Journal of Quant. Spectrosc. and Radiat. Transfer* **240** 106671.
- [16] MacBride C D *et al* 2021. *Phil. Trans. R. Soc. A.* **379** 20200171.
- [17] Ren T, Modest M F, Fateev A, Sutton G, Zhao W, Rusu F 2019. *Applied Energy* **252** 113448.
- [18] Samuell C M *et al* 2021. *Review of Scientific Instruments* **92** 043520.
- [19] TensorFlow: <https://www.tensorflow.org/guide> (consulted on September 29, 2022).
- [20] Koubiti M and Kerebel M 2022. *Appl. Sci.* **12** 9891

Acknowledgments

This research was funded by A*Midex IDEX PIA AMX-19-IET-013 (Master's internship of M. Kerebel) as well as by the French Federation of Research on Magnetic Fusion (FR-FCM), Project 3IPP_AAPFR2022_Koubiti.