# New perspectives in the analysis of Stark width regularities and systematic trends

**Z.** Majlinger<sup>1,2,3</sup> and I. Traparić<sup>4</sup>

 Astronomical Observatory, Volgina 7, 11060 Belgrade, Serbia
 <sup>2</sup> ES "I. G. Kovačić", Kralja Tomislava 18, 51326 Vrbovsko, Croatia
 <sup>3</sup> ES "V. Gortan", Prilaz Vladimira Gortana 2, 51000 Rijeka, Croatia
 <sup>4</sup> Institute of Physics Belgrade, Pregrevica 118, 11080, Belgrade, Serbia (E-mail: traparic@ipb.ac.rs)

Received: September 29, 2023; Accepted: October 20, 2023

Abstract. Regularites and systematic trends among the sample of Stark widths obtained by using modified semiempirical method from the STARK-B database were analysed. Two different approaches are independently used – multiple regression method combined with simple cluster analysis, and random forest (RF) machine learning algorithm. Predicted values of Stark widths calculated with estimate formulae obtained from multiple regression method, and those values predicted by using RF algorithm, were compared with already known corresponding experimental Stark widths published elsewhere. Results of this analysis indicate that both of these methods can mostly predict new Stark width values within the acceptable range of accuracy.

**Key words:** line profile — Stark broadening -- atomic data — machine learning

# 1. Introduction

Stark broadening theory plays the important role in investigation of high temperature dense plasma, where the collisional processes between the charged particles contribute significantly to the spectral line broadening. From technological perspective, Stark widths and shifts of spectral lines in the spectra of neutral atoms and ions are of interest for a number of problems - for example, analysis and modelling of laboratory, laser produced, fusion or technological plasmas, accurate spectroscopic diagnostics and modelling, etc. Applications of Stark broadening theory are also various in research of astrophysical plasma as well for example, for interpretation, synthesis and analysis of stellar spectral lines, determination of chemical abundances of elements from equivalent widths of absorption lines, opacity calculations, estimation of the radiative transfer through the stellar atmospheres and subphotospheric layers, radiative acceleration considerations, nucleosynthesis research, etc.

Calculation of Stark broadening parameters sometimes can be difficult, and it can take a time, especially if quantum theory is applied. If conditions to apply less accurate but faster quasistatic, unified, semiempirical or semiclassical methods are not satisfied, quick and simple estimates could become important, especially if we do not need a great accuracy, or there is no time for more complicated and more accurate calculations, or if we have a great number of Stark broadening parameters to calculate in very small period of time. This is very common case, for example, if astrophysical spectra are investigated. According to (Wiese, Konjevic, 1982), regularities and systematic trends (RST) can be found among the Stark widths of atomic spectral lines, which can simplify the way of obtaining these estimates. This is especially significant when some atomic data, necessary to perform more accurate Stark broadening methods of calculations, are missing. For example, the lack of atomic data, such as energy levels or transition probabilities is usually noticed in the spectral data for rareearth elements. Analysis based on RST is mostly the only way to determine Stark widths and shifts in sometimes very complex spectra of these elements, which become more and more important in spectral investigations of hot stars of spectral type A and B, and white dwarfs (Popović, Dimitrijević, 1998).

In this investigation, we focused on searching systematic trends among great amount of Stark widths from STARK-B database (Sahal-Bréchot et al., 2014b, 2015), obtained by modified semiempirical (MSE) method (Dimitrijević, Konjević, 1981) as a continuation of our previous work on determination of unknown MSE Stark widths and studying of RST among the MSE Stark broadening parameters (see, for example, Majlinger et al., 2015, 2017a,b, 2020a). Two different methods are used to analyse the sample – classical statistical regression method, which has already been used many times in previous investigations of regularities and systematic trends, and random forest (RF) algorithm from a group of machine learning methods, which become very popular methods more often used in these days whenever some classification or non-linear regression is needed to be performed. Unlikely to previous analyses of RST, here some new atomic parameters, which have not taken into consideration before, are included. We will shortly explain both of these methods and finally discuss and compare the obtained results.

#### 2. Methods

## 2.1. Simple cluster and multiple regression analysis

Estimates of Stark widths can be divided into three main groups:

- approximations derived from the theory e.g., Cowley's formula (Cowley, 1971) or MSE formula (Dimitrijević, Konjević, 1987)
- formulae based on statistical analysis on a large number of existing Stark widths (see e.g. Purić, Šćepanović, 1999; Purić et al., 1978).

 formulae based on systematic trends noticed without statistical analysis on corresponding examples (Wiese, Konjevic, 1982).

Whether it is one type or another, the formula for estimating Stark widths for lines of multiple ionized atoms usually can be expressed as a non-linear function of atomic and plasma parameters:

$$\omega_E = f(\lambda, N_e, T, Z, E_{ion}, E_{upper}, E_{lower}) \tag{1}$$

Sometimes some of these parameters are included in the estimate implicitly, through the effective ionization potential  $\chi_j$  for level j:

$$\chi_j = E_{ion} - E_j, \ j = \text{upper, lower}$$
(2)

according to Purić, or effective principal quantum number of the upper  $(n_+)$  or lower  $(n_-)$  level, which has already been used, for example, in MSE theory of Dimitrijević and Konjević (Dimitrijević, Konjević, 1981):

$$n_+^2 = \frac{Z^2 E_H}{\chi_{upper}} \tag{3}$$

$$n_{-}^{2} = \frac{Z^{2}E_{H}}{\chi_{lower}} \tag{4}$$

Here Z - 1 is the charge of the ion,  $\omega_E$  is the estimated Stark width in Å,  $\lambda$  is the wavelength in Å,  $N_e$  is the perturber density in cm<sup>-3</sup>,  $E_H$  is the energy of the hydrogen atom (or Rydberg constant),  $E_{ion}$  is the ionization energy, and  $E_i$  is the energy of upper or lower levels in cm<sup>-1</sup> (j = upper, lower).

Immediately after the first article on Stark broadening (Holtsmark, 1919), simple approximate formulas derived from the theory began to appear. Cowley's formula (Cowley, 1971) is probably the best known among them and it is still commonly used in astrophysics. Cowley (1971) specified three different formulas, one for neutral emitters, one for electrically charged emitters (which humble Cowley contributes to Griem), and one for estimating widths for temperatures close to 10000 K. The authors use different variants of Cowley's formula in addition to the original ones from the article (Cowley, 1971), and the difference is in the neglect or addition of the lower effective principal quantum number as a number and in the values of the numerical constant in the formula (see e.g. Killian et al., 1991; Alwadie et al., 2020).

Jagoš Purić made a great effort in studying RST among the Stark width values. The first works on regularities were published by (Purić, Ćirković, 1973) and (Purić et al., 1978). Purić and his co-workers found the correlation between Stark width and difference between ionization energy and energy of the upper state (what he called the upper effective ionization potential) and a number of experimental and theoretical values of Stark widths, offering a set of different estimation formulae. In the following decades, a number of papers on this topic were published, where different correlation parameters were stated for different transitions, different charges and different homologous and isoelectronic

sequences (see e.g. Miller et al., 1980; Purić et al., 1978, 1993, 1997, 2008). This statistical research is also supported by some other authors (see e.g. Djeniže, 1999; Djeniže et al., 2001), with occasional attempts to generalize this approach for all different transitions, different elements and different charge values (Purić, Šćepanović, 1999; Scepanovic, Puric, 2013). Comparing the great amount of Stark width data from STARK-B database (Sahal-Bréchot et al., 2014b, 2015), Purić offered so-called "generalized" estimate (Purić, Šćepanović, 1999) which should be used, according to the authors, "to calculate Stark line widths of the multiply charged ion of different elements along the periodic table." These scientific articles evolve over time, so Purić and co-workers later give up searching for a universal formula for all lines and focus their statistical analysis only on individual homologous or isoelectronic series (Dojčinović et al., 2011, 2012, 2013a,b; Tapalaga et al., 2011, 2018; Jevtić et al., 2012; Trklja et al., 2019b,a). However, the possibility to apply all of these estimates to predict new unknown Stark widths should be furtherly discussed (Majlinger et al., 2017a,b, 2020b).

The final purpose of this research was to find new general estimates accurate enough to approximately predict the unknown values of Stark widths. Our assumption is that these new estimates should be related on existing estimates, e. g. Cowley's from Cowley (1971) and Purić's from Purić and Šćepanović (1999). However, after investigation of accuracy in prediction of uknown Stark widths by using of these two estimates, in the cases of MSE calculated electron-impact widths for Lu III and Zr IV spectral lines, it was obvious that they don't offer enough accurate approximation (Majlinger et al., 2017a, 2020b). At least in the case of Zr IV Stark widths, several possible reasons were suggested to explain this discrepancy (Majlinger et al., 2017b):

- numerical coefficients in estimations are not properly adjusted
- some important parameters are neglected in equation (1) but significantly contribute to the result, and
- temperature functions used in previous estimates could be incorrect

According to statistical analysis of Stark widths calculated for 143 transitions from 26 multiply charged ions of 17 elements using the modified semiempirical method, (for example, most of them are elaborated by (Dimitrijević, Konjević, 1981), and previous assumptions, new estimates of Stark widths were found. After providing simple cluster analysis (Aggarval, V., 2014) and multiple regression analysis (for example Chatterjee, Simonoff, 2014), we concluded that MSE Stark width sample has to be devided in three separate groups:

1. For a type I of transitions: nl-nl', L = l, L' = l' (for example,  $2s^1S - 2p^1P^o$ ,  $3s^3S - 3p^3P^o$ ,  $3p^1P^o - 3d^1D$ ,  $4s^3S - 4p^3P^o$ , etc), proper estimate is Cowley-like:

$$\omega_{E1} = 3.438 \cdot 10^{-24} N_e \lambda^2 \frac{n_+^4 + n_-^4}{Z^2 (2l_> - 1)^{-\frac{3}{4}}} f(T)$$
(5)

2. For a type II of transitions: nl-n'l', L = l, L' = l' (for example,  $2p^3P^o-3s^3S$ ,  $4p^2P^o-5d^2D$ ,  $4d^2D-5f^2F^o$ ,  $4d^2D-6f^2F^o$ , etc.), proper estimate is Purićlike:

$$\omega_{E2} = 0.808 \cdot 10^{-25} N_e \lambda^2 \frac{n_+^6 + n_-^6}{Z^2 (2l_> - 1)^{-\frac{1}{5}}} f(T) \tag{6}$$

Here f(T) is chosen temperature function (which will be explained later), while  $l_{>} = max(l_{upper}, l_{lower})$ , where  $l_{upper}$  and  $l_{lower}$  are orbital quantum numbers for upper and lower level respectively.

3. For all other types of these simplest transitions (type III), like nl-nl',  $L \neq l$ ,  $L' \neq l'$  and nl-n'l',  $L \neq l$ ,  $L' \neq l'$  (for example,  $3s^1P^o-3p^1D$ ,  $3s^4P^o-3p^4P$ ,  $3p^5D^o-3d^5F$ ,  $3d^1F^o-4p^1D$ , etc.), a well known general expression, valid also for the first two types, can be used to obtain width for particular lines within a multiplet from an average width as a whole:

$$\omega_{E3} = \left(\frac{\lambda_{E3}}{\lambda_0}\right)^2 \omega_0 \tag{7}$$

where  $\omega_{E3}$  and  $\omega_0$  are estimates of uknown Stark widths and a Stark width obtained with estimates (5) or (6), while  $\lambda_{E3}$  and  $\lambda_0$  are corresponding wavelengths respectively.

After optimizing the number of parameters in these estimates according to minimum description length properties (see, for example, Grünwald, 2004), and keeping in mind that all models are uncertain, idealizing reality (Wit et al., 2012) and that sample is not equal to population, we rounded exponents in (5) and (6) on the closest integer or rational number, to avoid physically meaningless results (for example,  $\lambda^{1.74}$  is replaced with  $\lambda^2$ ,  $Z^{1.95}$  with  $Z^2$ , etc.) and to approach enough to probable statistical model ideally concerning about population.

From interpolation of analysed data, new temperature function is suggested:

$$f(T) = \frac{1-\beta}{\sqrt{T}} + \beta \frac{\ln T}{\sqrt{T}}$$
(8)

where  $\beta$  is the linear function of temperature defined as:

$$\beta = AT + B \tag{9}$$

Numerical constants A and B are estimated to be  $A = 9.62 \cdot 10^{-7}$  and  $B = -4.167 \cdot 10^{-2}$  from the values of lower temperature limit for all considered Stark widths. Lower temperature limit for most of considered Stark widths lies in a range 15000 – 70000 K which corresponds to range of distance between perturbing and perturbed levels used in all considered Stark width calculations around cca 7500 – 38500 cm<sup>-1</sup>. It is easy to see that relation  $0 \le \beta \le 1$  is mostly valid for such choice of A and B, and that temperature function approximately

simulates both lower and upper temperature limit conditions when  $\beta$  reach to its limits, which is in a good agreement with some previous analyses of behavior of Stark width values for highest and lowest value of temperature (for example Sahal-Bréchot et al., 2014a).

Correlation between new estimates of full Stark width at half maximum (FWHM) obtained by using relations (5), (6) and (7), and existing MSE values for transition type I, type II and type III with corresponding regression lines are displayed in Figs. 1-3 respectively. To calculate correlation parameters for each estimate, we used the general symbol  $\omega_{EST}$  instead of  $\omega_{E1}$ ,  $\omega_{E2}$ , and  $\omega_{E3}$ . Correlation coefficients corresponding to each estimate WEST are presented in Table 1. In the most idealistic scenario, for log-log regression equation  $\log \omega_{EST} = C_1 + C_2 \log \omega_{MSE}$ , should be valid C1 = 0, C2 = 1 and therefore  $\omega_{EST} = \omega_{MSE}$ . As the additional attemption to confirm a validation of this method, predicted Stark widths with estimates from above are compared with corresponding experimental values from references (Konjević et al., 1984, 2002). Result of this comparison is presented in Fig. 4.

**Table 1.** Correlation parameters for log-log regression equation  $\log \omega_{EST} = C_1 + C_2 \log \omega_{MSE}$ , between results of estimates (5), (6) and (7) respectively, and MSE values of FWHM Stark widths from analysed sample.

Transition type	$C_1$	$C_2$	$ErrC_1$	$ErrC_2$	St. dev.	$R_{corr}$
Ι	-1.45E-5	0.9126	0.0325	0.0336	0.13	97.76
II	28E-5	1.0266	0.0473	0.0262	0.19	99.32
III	-0.0334	0.84	0.025	0.029	0.064	97.23

#### 2.2. Machine learning methods and RF algorithm

As machine learning represents a very popular tool for different types of problems encountered in science, here it was applied on the study of regularities of Stark broadening. Machine learning model based on Random Forest algorithm was developed and described in detail in reference (Tapalaga et al., 2022), so here it would be briefly described for the sake of completeness. Before developing the model, we needed to develop and create a database for training and testing of the future models. This database was created as a combination of two databases, namely NIST atomic database (Kramida et al., 2022), from which we took atomic parameters of interest for every transition and Stark B database (Sahal-Bréchot et al., 2015) from which we took Stark width and plasma parameters for each calculated width. After the completion of this database, it contained around 53 000 lines. Features selected for this research were: Plasma electron density, electron temperature, atomic number, charge of the emitter, energies of both upper and lower levels, total angular momentum of both upper and lower levels, principal and orbital quantum numbers for initial and final of corresponding transitions.

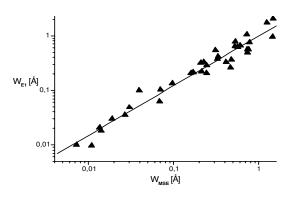


Figure 1. Log-log correlation between FWHM Stark width values obtained by using estimates from multiple regression analysis ( $\omega_{E1}$ ) and MSE values ( $\omega_{MSE}$ ), with corresponding regression line.

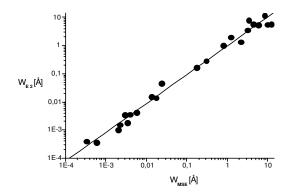


Figure 2. Log-log correlation between FWHM Stark width values obtained by using estimates from multiple regression analysis ( $\omega_{E2}$ ) and MSE values ( $\omega_{MSE}$ ), with corresponding regression line.

Additionaly, ionization energy and quantity called upper level effective potential after (Purić, Šćepanović, 1999) were taken into a set of input parameters, which provides a label data comparison. Data outliers were removed as data having higher energy of lower level than upper level. Finally, the analysis was constrained to the following plasma parameters:  $N \leq 10^{17} \text{ cm}^{-3}$ ,  $T_e \leq 150\ 000$ K and  $E_{upper} \leq 500\ \text{eV}$ . This restriction left us with around 32 000 available transitions for further analysis. To choose the best model and corresponding parameters, GridSearchCV (Grid Search Cross Validation) technique was applied. Here for every set of model parameters, model is trained and tested on

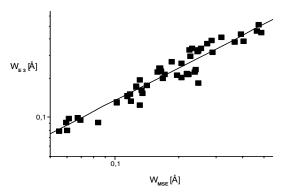


Figure 3. Log-log correlation between FWHM Stark width values obtained by using estimates from multiple regression analysis ( $\omega_{E3}$ ) and MSE values ( $\omega_{MSE}$ ), with corresponding regression line.

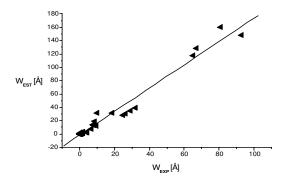


Figure 4. Linear correlation between FWHM Stark width values obtained by using estimates from multiple regression analysis ( $\omega_{EST}$ ) and corresponding experimental values ( $\omega_{EXP}$ ), with corresponding regression line.

given dataset, and best performance is reported. Along with best performance, algorithm reports with which parameters has been obtained. Three models were tested: Decision Tree, Random Forest and Gradient Boosting Regressor. Performances of the model are reported in table 2.

It can be seen that the best results were obtained with Random Forest for the following parameters: maximal depth of the tree equal to 10, minimal samples at one leaf set to 3 and number of estimators equal to 200. As in the case of multiple regression method, Stark widths predicted with using RF algorithm were compared with corresponding experimental widths from the same refer-

Model	Parameters	$R^2$ score
Decision tree	$max_depth = 5$	0.9
	$max_depth = 10$	
Random Forest	$\min\_samples\_leaf = 3$	0.97
	$n_{\text{-}}estimators = 200$	
	$max_depth = 10$	
Gradient Boosting Regressor	$\min_{\text{samples_leaf}} = 2$	0.96
	$n_{\text{estimators}} = 200$	

**Table 2.** Comparison of preformances for three learning machine models used in analysis.

ences as before (Konjević et al., 1984, 2002). Results of this comparison are shown in Fig. 5 and Fig. 6. As we can see on figure 5, RF model preforms well, except of few points that are estimated badly. Also, from the figure 6 it can be concluded that RF method performs better in visible part than in the ultra violet or infrared part of the spectrum.

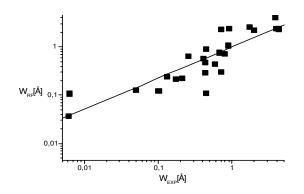
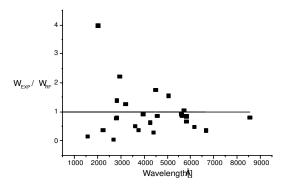


Figure 5. Log-log correlation between FWHM Stark width values obtained by using RF algorithm ( $\omega_{RF}$ ) and corresponding experimental values ( $\omega_{EXP}$ ), with corresponding regression line.

To improve the model and to test whether we could reduce the number of features in the dataset while keeping the accuracy of the model, permutation importance test was performed. This method permutes each feature randomly within dataset, and calculate the decrease in performance of the already trained model. Greater the decrease, more important is the feature. Results of this analysis are given in table 3.



**Figure 6.** Dependence of ratio between experimental FWHM Stark widths ( $\omega_{EXP}$ ) and corresponding values obtained by using RF algorithm ( $\omega_{RF}$ ) on wavelength of spectral lines for which Stark widths are calculated.

Feature	Importance score
Electron density $N_e$	$5.9 \pm 0.2$
Upper - level principal quantum number $n_i$	$3.50\pm0.12$
Charge of emitter $q$	$1.04\pm0.04$
Upper-level effective potential $\chi$	$0.37 \pm 0.04$
Emitter $Z$	$0.31\pm0.03$
Upper-level orbital quantum number $l_i$	$0.19\pm0.02$
Energy of upper level $E_{upper}$	$0.11 \pm 0.02$
Lower-level principal quantum number $n_f$	$0.11\pm0.01$
Energy of lower level $E_{lower}$	$0.048 \pm 0.005$
Lower-level orbital quantum number $l_f$	$0.029 \pm 0.004$
Lower-level total angular momentum quantum number $J_{upper}$	$0.029 \pm 0.007$
Upper-level total angular momentum quantum number $J_{lower}$	$0.021 \pm 0.005$
$T_e$	$0.017 \pm 0.003$
$E_i$	$0.002 \pm 0.001$

Table 3. Feature importance test results for RF model.

Results in table 3 indicate that electron density is most important feature as expected, while other important features are naturally emitter, its charge, principal and orbital quantum numbers of upper level and upper-level effective potential  $\chi$ . Other parameters were removed from analysis, as upper level is included in definition of  $\chi$ , and model was retrained. As expected, model gave very similar results as those reported in this work, which just confirms that model didn't got confused with some redundant data in initial run.

## 3. Discussion

For Type I, ratios between estimates and MSE values vary between 0.5 and 2.6, for type II between 0.4 and 1.7, and for type III between 0.7 and 2.0. so we can say that accuracy of our estimates according to MSE values are mostly between -50% and +160%. Including predicted accuracy of MSE results, which is  $\pm 50\%$ , we expect that global accuracy of our estimates, according to statistical sample we used, should lie between  $\pm 50\%$  and  $\pm 100\%$  e.g. comparable with the old Griem's semiempirical theory (Griem, 1968).

Despite of several exceptions, ratio between most of new calculated estimates and corresponding experimental Stark widths from references (Konjević et al., 1984, 2002), lies between 0.2 and 2 (see Tab. 3), which leads to conclusion that our estimates are usually accurate in a range of  $\pm 100\%$ , in accordance with our expectations. On the other hand, average value of this ratio for comparison of estimates with experimental Stark widths is  $1.38 \pm 0.11$ , resulting with accuracy in a range between  $\pm 30\%$  and  $\pm 50\%$ , which is even better than theoretically predicted accuracy for modified semiempirical theory by Dimitrijević and Konjević (1981).

Ratios between Stark widths predicted by using RF algorithm and corresponding experimental Stark widths taken from the same references mentioned above (Konjević et al., 1984, 2002), with the exception of two extreme values 0.06 and 4, lie between 0.16 and 2.23, but average ratio of these values is 0.96  $\pm$  0.16, leading to an accuracy of around  $\pm$ 20%, which is much better than accuracy of predicted results obtained by using classical statistical method. To express the accuracy for both methods, as it is usual in statistics, we used arithmetical mean as the average value of analysed data, while the standard deviation is used as a measure of data dispersion. As a final proof that both of presented methods could be valid, in Fig. 7 we presented results obtained from mutual comparison of Stark width values predicted with these two different approaches. Linear regression equation which expresses dependence between Stark widths predicted with RF method  $\omega_{RF}$  and those predicted by using estimates WEST obtained by using formulae (5), (6) and (7) is found to be  $\omega_{RF}$ =  $0.0523 + 1.0563 \omega_{EST}$  with correlation coefficient  $R_{CORR} = 91.05\%$ . Figure 7 and values of correlation parameters show that both of these two methods are equivalent, e. g. the results of the estimates with RF model and classical multiple regression statistical method are almost the same. Although it is feeded with results obtained by using semiclassical perturbation method (see for example Sahal-Bréchot et al., 2014a), RF algorithm is shown to be a good predictor, despite of a theoretical method used to calculate analysed Stark width data, because it gives results comparable with estimates based on set of calculations obtained by using modified semiempirical method.

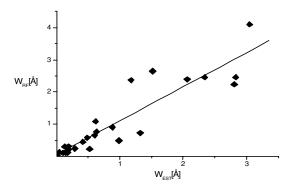


Figure 7. Linear correlation between FWHM Stark width values obtained by using estimates from multiple regression analysis ( $\omega_{EST}$ ) and corresponding values obtained by using RF algorithm ( $\omega_{RF}$ ), with corresponding regression line.

# 4. Conclusion

Both of the methods used in this study have some advantages and disadvantages. In general, the advantage of ML models is that they are faster and easier to perform with the proper knowledge of computer programming. On the other hand, any of ML algorithms is some kind of black box, e. g. we finally don't know how input and output parameters are connected. If we want to find out the relationship between Stark width values and atomic and plasma parameters presented in a form of simple formula, we have to continue to investigate regularities and systematic trends of Stark widths using the estimates as, for example, were obtained here (equations (5)-(7)). If we don't need to know this connection, using of some ML algorithm is probably the best solution. Results of predictions using RF model show that, if some general estimate really exists, according to previous vision of Jagoš Purić, it should be the function of 14 variables. In this case, number of input atomic and plasma parameters we used before in a group of equations (1) to find systematic trends among the Stark width value, should be enlarged. We proved that, with additional two parameters (upper and lower orbital quantum number) and considering transition type into analysis, strongly affect on result of estimate, as it is assumed, for example, in (Majlinger et al., 2017b).

However, it is very important to stress that the estimates obtained in this work should be valid under the assumption that they can be applied on simple type of spectra, as they have been analysed in this case (for example, where for all transitions in a whole spectrum parent term remains the same). For more complex spectra, these estimates should be improved, or some other methods are welcome to be used. Furthermore, although RF model shows very strong potential to be applied on RST analysis in future, it is tested only in the sample of Stark broadening parameters related to simple spectra described here, and in the case of Li I spectral lines (Tapalaga et al., 2022), so it should also be confirmed in a greater sample to make us sure that this method can be applied generally in prediction of new Stark widths despite of complexity of a spectrum we investigate. For the application of these methods to study regularities and systematic trends among the Stark broadening parameters of lines in more complex spectra, additional investigations are needed, and development of both of these method are neccessary. Created database used in this and previous study is published online and it is available for use. It can be found on the link https://github.com/ivantraparic/StarkBroadeningMLApproach.

Acknowledgements. The research was funded by the Ministry of Science, Technological Development and Inovations of the Republic of Serbia, Contract numbers: 451-03-68/2022-14/200024 and 451-03-68/2022-14/200146. We are very grateful to Milan S. Dimitrijević on useful suggestions and comments during the writing of this paper.

# References

Aggarval, C., V., R. C. 2014, Data Clustering (Taylor and Francis Group)

- Alwadie, N., Almodlej, A., Ben Nessib, N., Dimitrijević, M. 2020, Contrib. Astron. Obs. Skalnaté Pleso, 50, 86
- Chatterjee, S., Simonoff, J. S. 2014, *Handbook of Regression Analysis* (John Wiley and Sons)
- Cowley, C. 1971, The Observatory, 91, 139
- Dimitrijević, M., Konjević, N. 1987, J. Quant. Spectrosc. Radiat. Transfer, 20, 223
- Dimitrijević, M. S., Konjević, N. 1981, *Spectral Line Shapes* (Walter de Gruyter and Co.), 211
- Djeniže, S. 1999, J. Res. Phys., 28, 231
- Djeniže, S., Srećković, A., Kalezić, S. 2001, Serb. Astron. J., 164, 21
- Dojčinović, I. P., Tapalaga, I., Purić, J. 2011, Pub. Astron. Soc. Australia, 28, 281
- Dojčinović, I. P., Tapalaga, I., Purić, J. 2012, Mon. Not. R. Astron. Soc., 419, 904
- Dojčinović, I. P., Tapalaga, I., Purić, J. 2013a, Bull. Astron. Soc. India, 41, 281
- Dojčinović, I. P., Tapalaga, I., Purić, J. 2013b, Mon. Not. R. Astron. Soc., 429, 2400
- Griem, H. R. 1968, Phys. Rev., 165, 258
- Grünwald, P. D., A tutorial introduction to the minimum description length principle. 2004, ArXiv, math.ST/0406077
- Holtsmark, J. 1919, Annalen der Physik, 363, 577
- Jevtić, D., Dojčinović, I. P., Tapalaga, I., Purić, J. 2012, Bull. Astron. Soc. India, 40, 151
- Killian, J., Montenbruck, O., Nissen, P. 1991, Astron. Astrophys. Suppl., 88, 101

- Konjević, N., Dimitrijević, M. S., Wiese, W. L. 1984, J. Phys. Chem. Ref. Data, 13, 649
- Konjević, N., Lessage, A., Fuhr, J. R., Wiese, W. L. 2002, J. Phys. Chem. Ref. Data, 31, 819
- Kramida, A., Ralchenko, Y., Reader, J., NIST ASD Team. 2022, NIST Atomic Spectra Database (version 5.10), https://physics.nist.gov/asd, accessed: September 2023
- Majlinger, Z., Dimitrijević, M. S., Srećković, V. A. 2020a, Mon. Not. R. Astron. Soc., 496, 5584
- Majlinger, Z., Dimitrijević, M. S., Srećković, V. A. 2020b, Contrib. Astron. Obs. Skalnaté Pleso, 50, 164
- Majlinger, Z., Dimitrijević, M. S., Simić, Z. 2017a, Atoms, 5, 49
- Majlinger, Z., Simić, Z., Dimitrijević, M. S. 2015, Astron. Astrophys, 36, 671
- Majlinger, Z., Simić, Z., Dimitrijević, M. S. 2017b, Mon. Not. R. Astron. Soc., 470, 1911
- Miller, M. H., Lessage, A., Purić, J. 1980, Astrophys. J., 239, 410
- Popović, L. C., Dimitrijević, M. S. 1998, Contrib. Astron. Obs. Skalnaté Pleso, 27, 353
- Purić, J., Dimitrijević, M. S., Lakićević, I. S. 1978, Physics Letters A, 67, 189
- Purić, J., Dojčinović, I. P., Nikolić, M., et al. 2008, Astrophys. J., 680, 803
- Purić, J., Šćepanović, M. 1999, Astrophys. J., 521, 490
- Purić, J., Šćepanović, M., Milosavljević, V., Ćuk, M., XXIII ICPIG (Toulouse, France). 1997, V – 48
- Purić, J., Ćirković, L., XI ICPIG. 1973, 398
- Purić, J., Miller, H., Lessage, A. 1993, Astrophys. J., 416, 825
- Sahal-Bréchot, S., Dimitrijević, M., Ben Nessib, N. 2014a, Atoms, 2, 225
- Sahal-Bréchot, S., Dimitrijević, M. S., Moreau, N. 2015, STARK-B database, http://stark-b.obspm.fr, accessed: September 2023
- Sahal-Bréchot, S., Dimitrijević, M. S., Moreau, N., Ben Nessib, N., SF2A Conf. 2014b, 515
- Scepanovic, M., Puric, J. 2013, Rom. Rep. Phys., 65, 1275
- Tapalaga, I., Dojčinović, I. P., Purić, J. 2011, Mon. Not. R. Astron. Soc., 415, 503
- Tapalaga, I., Dojčinović, I. P., Purić, J. 2018, Mon. Not. R. Astron. Soc., 474, 5479
- Tapalaga, I., Traparić, I., Trklja Boca, N., et al. 2022, Neural Comput. Applic., **34**, 6349
- Trklja, N., Dojčinović, I. P., Tapalaga, I., Purić, J. 2019a, Atoms, 7, 99
- Trklja, N., Tapalaga, I., Dojčinović, I. P., Purić, J. 2019b, New Astron, 59, 54
- Wiese, W., Konjevic, N. 1982, J. Quant. Spectrosc. Radiat. Transfer, 28, 185
- Wit, E., van den Heuvel, E., Romeijn, J. 2012, Statistica Neerlandica, 66, 217