



# Using a multiple regression model for predicting the soft X-ray laser gain coefficient from laser plasmas

G. Ghani-Moghadam<sup>1</sup>

Received: 20 January 2024 / Accepted: 13 September 2024  
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2024

## Abstract

The laser plasmas produced from the interaction of the high-power laser with the target surface are suitable sources for soft X-ray laser, which has many applications in industry and medicine. In order to create the maximum efficiency of the output soft X-ray laser, the parameters of the pump laser should be optimized. In a double-pulse pump laser, the intensity and width of the pre-pulse, the intensity and width of the main pulse and the delay time between the two pulses are effective parameters in the output X-ray production. In this research, by using a machine learning based on the multiple regression equation, the dependence of the gain coefficient on the pump pulse parameters are investigated and a statistical model for predicting the gain coefficient of soft x-ray laser based on the feature of the pump pulse is presented without the need to numerical solution of the complex hydrodynamic equations.

## 1 Introduction

Recently, the use of machine learning and its applications have become common in many fields of science [1–3]. The role of machine learning in physics has also been seen recently in several reports [4–11]. Regression analysis is a fundamental concept in the field of machine learning. It undergoes supervised learning where the algorithm is trained with both input features and output labels. It helps in establishing a relationship among the variables by estimating how one variable affects the other [12]. Regression in machine learning consists of mathematical methods that allow data scientists to predict a continuous outcome ( $y$ ) based on the value of one or more predictor variables ( $x$ ). Linear regression is probably the most popular form of regression analysis because of its ease-of-use in predicting and forecasting. In this paper, one of the fields of physics, namely X-ray lasers, has been investigated using this machine learning model.

Soft X-ray lasers have many applications in the field of industry and medicine like microscopy, holography and lithography [13]. Soft X-ray lasers are also used for interferometry and radiography of compressed fuels in the field of ICF [14]. Laser plasmas or plasmas produced through

high-power optical lasers are one of the sources of soft X-ray laser production that have been used in many experimental and theoretical researches [15–18]. The first laboratory experiment to produce a soft X-ray laser in 1985 goes back to the use of  $\text{Se}^{+24}$  as the active medium [18]. Then, significant progress was made theoretically and experimentally using various methods of X-ray laser pumping [14]. Among the various types of pumping methods, collisional excitation scheme has been expanded and the saturation gain at the wavelengths of soft X-ray laser in Ne-like zinc [19], Ne-like germanium [20] and Ne-like selenium [21] ions has been reported using this pumping method. In this method, first, an optical beam is created in the visible region focused on the target and the plasma medium. After that, population inversions of transition levels produce with the collision of free electrons that take the bound electrons to quasi-stable levels in Ni-like or Ne-like ions, and activity of soft X-ray laser in the form of enhanced spontaneous emission (ASE) occurs. Since the production efficiency of X-ray lasers from optical lasers was relatively low, a pre-pulse was successfully used to enhance the production efficiency of Ne-like soft X-ray lasers in several experiments [22, 23]. Research shows that the use of active plasma medium and the use of the collisional excitation approach with the use of transient pumping are the most beneficial in the production of soft X-rays [15]. From the physical point of view, the absorption efficiency of pump laser radiation is higher with the addition of a pre-pulse and less energy is spent on plasma expansion.

✉ G. Ghani-Moghadam  
ghanimoghadam@yahoo.com

<sup>1</sup> Hazrat-e Masoumeh University, P.O.BOX: 37195-1179, Qom, Iran

In this case, with the sudden increase of the electron temperature in the area where the density of Ne-like ions is high, a fundamental improvement in the X-ray laser efficiency is achieved. This rapid increase of electron temperature is possible through plasma radiation produced with a short (few picoseconds) and high intensity laser pulse. This pulse is so fast that it prevents processes such as vaporization, heat conduction, expansion and re-ionization, or reverse processes during pump.

In reference [11], it has been proposed the use of a multi-layer perceptron neural network for prediction of X-ray laser gain coefficient, which requires training the neural network with multiple data, therefore in this research it is proposed using of a multiple regression model that is presented an equation to predict gain coefficient.

Since the gain factor of the soft X-ray laser produced from laser plasmas depends on the characteristics of the pump laser [24–28], in this research, a model for predicting the gain coefficient has been presented using the multiple regression equation. Multi regression analysis is the process of estimating the relationship between a dependent variable (here gain coefficient) and several independent variables (pump pulse characteristics). In simpler words, it means fitting a function from a selected family of functions to the sampled data under some error function.

In the next section, a multiple regression model is explained. In the Sect. 3, the soft X-ray laser gain obtained from hydrodynamic calculations is discussed and the data used for regression analysis is shown. Then, in the Sect. 4, the results of regression analysis are shown and it can be seen that the regression equation predicts well the results of numerical simulations of hydrodynamic equations.

## 2 The multiple regression model

Regression analysis is one of the most basic tools in the area of machine learning used for prediction. Using regression you fit a function on the available data and try to predict the outcome for the future or hold-out data points.

In general, such a model along with random error is written as follows:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p + \varepsilon \quad (1)$$

Here,  $y$  is the predicted value of the dependent variable for any value of the independent variable ( $\mathbf{x}$ ),  $\beta_0$  is the value of  $y$  when  $\mathbf{x}$  is zero,  $\beta_1$  to  $\beta_p$  are the regression coefficients, meaning the expected change in  $y$  when  $\mathbf{x}$  increases and  $\varepsilon$  is the estimated error in the regression. Creating and estimating the regression coefficients in this model reveals the linear relationship between independent and dependent

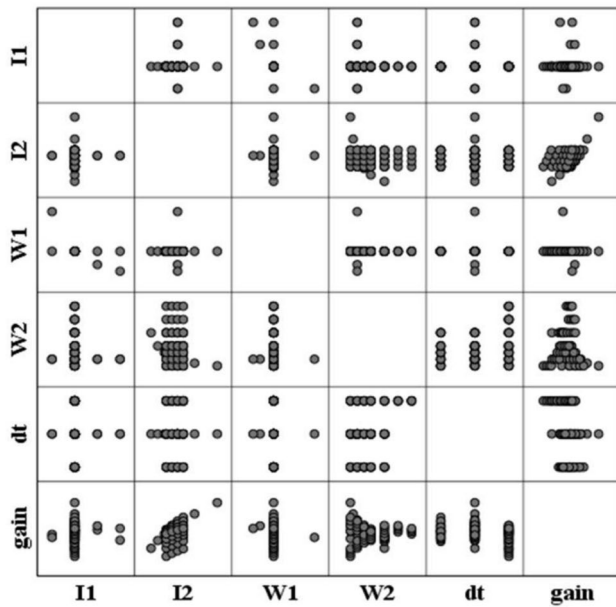
variables and provides the possibility of dependent variable prediction.

## 3 Properties of the pump laser

In reference [27], one-dimensional hydrodynamic simulation code MED103 [29, 30] has been used in order to investigate and study the plasma active medium during laser interaction with plasma. The laser plasma is produced using the transient pump method, which consists of two pump pulses with a delay time of several hundred ps. The first pulse, a long pulse with a width of several hundred ps, hits the solid target and produces plasma with the necessary degree of ionization. The second pulse, which is a short pulse with width of a few ps, heats the free electrons quickly in a shorter time than the plasma ionization time, up to several hundred eV. As a result, the necessary conditions for the pumping of laser active ions are prepared by collisional excitation in such a way that the gain of the soft X-ray laser is high. In the reference [27], the first, a pre-pulse with intensities in order of  $10^{13}$  W/cm<sup>2</sup> and widths of several hundred ps, and therefore with different energies and a wavelength of 800 nm, is irradiated to the germanium target surface with a thickness of 25  $\mu\text{m}$ . Then the main pulse is introduced with different intensities in order of  $10^{15}$  W/cm<sup>2</sup> and widths of several ps with the same wavelength and delay time of 100 to 200 picoseconds between the maxima of the two pulses. In this case, the maximum gain of the soft X-ray laser has been investigated at the wavelength of 19.6 nm. Therefore, the effective factors in producing the maximum gain of the soft x-ray laser are the intensity and width of the pre-pulse, the intensity and width of the main pulse and the delay time between the two pulses. These parameters are independent variables in regression model and the gain coefficient of the soft x-ray laser is dependent variable. They are used to create a regression model. In this case, by creating a regression relationship, it is possible to predict the output soft x-ray laser gain for each input of the effective parameters in the gain coefficient, without the need to solve complex hydrodynamic equations.

## 4 Results and discussion

As stated in the previous section, in this research, the data of reference [27] is used to create a regression relationship. For this purpose and for the statistical analysis of these data, IBM SPSS Statistics software version 26 is used. Before the regression analysis, it is better to check the existence of a linear relationship between each of the independent variables and the dependent variable by using the scatter diagram and also calculating the correlation coefficient.



**Fig. 1** The matrix of the scatter diagram for independent and dependent variables ( $W_1$  and  $I_1$  are pre-pulse width and intensity;  $W_2$  and  $I_2$  are main pulse width and intensity;  $dt$  is delay time between the two pumped pulses and  $gain$  is the gain coefficient of soft X-ray laser)

As it is clear from the scatter diagram (Fig. 1), it seems that there is no linear relationship between the gain coefficient and the intensity and width of the pre-pulse. While a very good linear relationship can be established between the gain coefficient and the intensity of the main pulse, and also there is a relatively good relationship between the gain coefficient and the delay time parameter. Since one of the important conditions in estimating linear regression parameters is the normality of the residuals, the Durbin-Watson test is used to test the randomness and independence of the residuals. Then Pearson's correlation coefficient ( $R$ ) and coefficient of determination ( $R$  Square) have been investigated. The coefficient of correlation and the coefficient of determination are obtained 0.652 and 0.425, respectively. In fact, these coefficients indicate a linear correlation between the value of dependent variable and the value predicted by the model. If these coefficients be near to 1, the model has been able to predict the dependent variable well.

Table 1 shows the estimation of coefficients and characteristics related to their test. In B column, the real coefficients of the variables and the constant value in the model are given. Therefore, the regression equation is displayed as follow:

$$\begin{aligned} gain = & 300.521 - 9.206 I_1 + 15.266 I_2 \\ & - 0.158 W_1 + 1.560 W_2 - 0.341 dt \end{aligned} \quad (2)$$

**Table 1** Regression coefficients, standard errors, T-values and significant level (Sig.) for 5 independent variables

Predictors	Unstandardized coefficients		Standardized coefficients	T	Sig
	B	Std. Error	Beta		
(Constant)	300.521	52.182		5.759	.000
$I_1$	- 9.206	10.813	- .085	- .851	.397
$I_2$	15.266	2.170	.572	7.034	.000
$W_1$	- .158	.106	- .148	- 1.494	.139
$W_2$	1.560	2.860	.046	.545	.587
$dt$	- .341	.091	- .314	- 3.742	.000

$I_1$  and  $I_2$  are the intensities of the pre-pulse and main pulse of the pump laser beam,  $W_1$  and  $W_2$  are the width of the pre-pulse and the main pulse of the pump laser,  $dt$  is the delay time between the two pulses, and  $gain$  is the gain coefficient of the output soft X-ray laser. Also, comparing Eq. 1 and 2, it can be seen that the error term is omitted in Eq. 2, because error term is included to account for the difference between the actual and predicted results and is not part of the prediction itself. In fact it reflects the discrepancy between the actual and predicted values, which can only be observed after we know the true outcomes.

The real coefficients in B column are created according to the measurement unit of each variable. Therefore, it is not possible to distinguish the importance of the relevant variable in the regression model based on the magnitude of each coefficient. To determine the importance of each variable and their role in the regression model, one should pay attention to the column of standardized coefficients. Standardized data are used to calculate standardized coefficients. This means that each value of dependent and independent variables is subtracted from their mean and divided by the standard deviation. Then the regression model is fitted and the coefficients are calculated. Since the variables in this case don't have units, their magnitude isn't depended on the measurement unit.

Therefore we use the Standardize Beta Coefficients column to determine the importance of each predictor. Any variable with a larger Beta is more important in the regression model. This means the variables can be easily compared to each other with Beta. In this way, it is clear that the variable  $I_2$  with Beta=0.572 is the best variable for predicting the dependent variable in the regression model.

Columns T and Sig are relevant to coefficient hypothesis test. The larger value of T has led to the weaker assumption that the coefficient is zero, and therefore the greater role of that variable in modeling. Table 1 shows the large T-values for the main pulse Intensity and the delay time therefore these two variables have the great role in the gain coefficient. This magnitude is also determined by the value of Sig. At

the error level of 0.05, if the value of Sig is smaller than 0.05, the null hypothesis, which indicates that the variable is ineffective in the model, is rejected. Therefore, for  $\text{sig} < 0.05$ , the regression model is suitable. A value of 0.05 is considered as the same type I error or significant level. According to Table 1, the coefficients of the intensity of the main pulse and the delay time between the two pulses have a Sig smaller than 0.05, so they are statistically significant and should be included in the model, while the intensity and width of the pre-pulse and the width of the main pulse have a Sig greater than 0.05 that are removed from the model and regression calculations should be performed again to reach the correct coefficients for significant variables.

Table 2 shows the coefficients of the regression model considering only two independent variables, the intensity of the main pulse and the delay time between the two pulses. In this case, more correct and accurate coefficients have been obtained for these two variables. However, due to the small difference of these coefficients with the previous coefficients, it can be concluded that the regression model presented in relation 2 is a successful model for predicting the gain coefficient using the pump pulse characteristics.

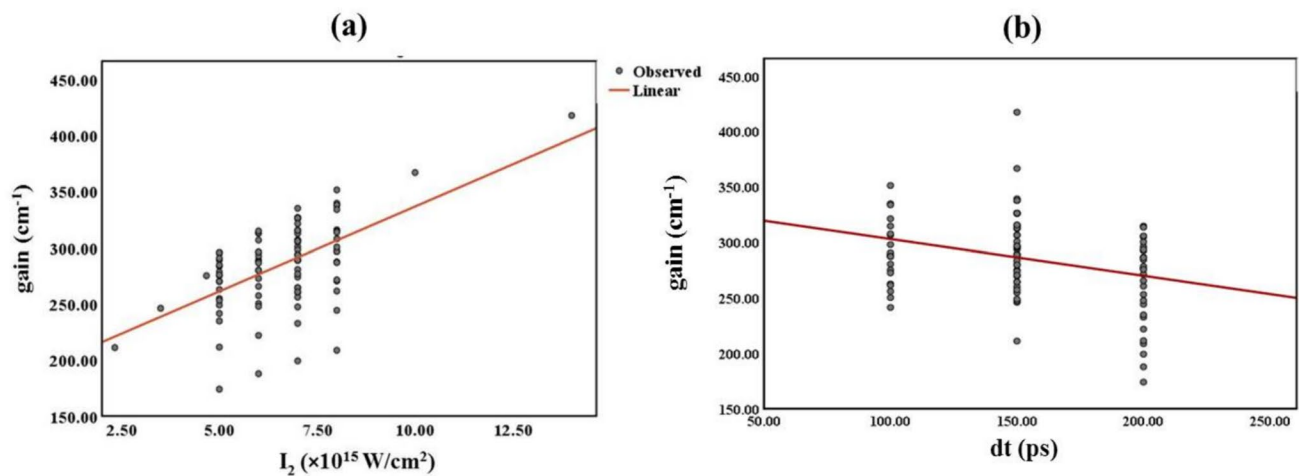
**Table 2** Regression coefficients, standard errors, T-values and significant level (Sig.) for 2 independent variables ( $I_2$  and dt)

Predictors	Unstandardized coefficients		Standardized coefficients	T	Sig
	B	Std. Error			
(Constant)	236.268	19.807		11.928	.000
$I_2$	14.997	2.141	.562	7.004	.000
dt	-.325	.087	-.299	-3.731	.000

Figure 2 shows the gain coefficient versus the main pulse intensity ( $I_2$ ) and the delay time between the two pulses (dt). The circles are the obtained values from the hydrodynamic code and the red line is the equation of the regression line. By comparing the two graphs in Fig. 2, it can be seen that the regression line is better for describing the data in the Fig. 2a. Since a previous study using a neural network revealed the significant role of the main pulse intensity in the gain coefficient [11], determining a linear relationship between efficiency and intensity can effectively contribute to predicting the gain coefficient. In this way, instead of solving complex hydrodynamic equations, this simple linear relationship can be employed for practical applications. However, according to the dispersion of the data compared to the regression line in Fig. 2, it should be noted that this regression line can be used in a narrow range of a small set of parameters. For a more accurate match, it is necessary to make corrections to the model, at least of the second order.

## 5 Conclusion

Soft X-ray lasers are a type of coherent light source that operate in the soft X-ray region of the electromagnetic spectrum, typically ranging from about 10 to 100 nm in wavelength. These lasers are known for their high brightness, short wavelength, and ability to produce ultrafast pulses, making them invaluable tools for a wide range of scientific and technological applications. The amplification of soft X-ray lasers involves a process where the intensity of the emitted X-ray light is increased through stimulated emission. This is typically achieved using a high-energy pump source, such as a high-power optical laser, to excite the medium and create a population inversion. The amplified X-ray beam produced through



**Fig. 2** Gain coefficient vs (a): intensity of the main pulse and (b): the delay time between the two pulses. (Circles (●) are observed values and the straight line is the linear regression function with one independent variable)

this process is highly coherent and intense, making it suitable for applications that require high-resolution imaging and ultra-fast time-resolved measurements. Therefore, it is important to optimize the conditions of the pump laser to achieve an enhanced X-ray laser. With the expansion of machine learning in various sciences, including physics, in this research, a model is presented to predict the gain coefficient of soft X-ray laser output from laser plasmas by using the multiple regression analysis. In fact, with the characteristics of a two-pulse pump laser, including the intensity and width of the pre-pulse, the intensity and width of the main pulse, and the delay time between the two pulses, it is possible to predict gain of soft X-ray laser from laser plasmas by using this model directly and without solving the hydrodynamic equations numerically. For this purpose, by using values obtained from the hydrodynamic code for training, a linear relationship between the x-ray gain coefficient as the dependent variable and the characteristics of pump laser as independent variables was derived. With consideration of Sig and T values was conclude that the model yields satisfactory linear relationship between the gain coefficient and two of the independent variables, which are main pulse intensity and delay time. Although the presented model performs well within a specific range of parameters, it should be noted that for broader applications or higher accuracy, further improvements and corrections are necessary. One possible approach to enhance the model's accuracy is by adding second-order or higher-order corrections. These modifications can improve the model's performance across a wider range of parameters and increase its alignment with real-world data.

**Author contributions** This manuscript has one author named G. Ghani-Moghadam who wrote and prepared the entire manuscript.

**Funding** No funding was received to assist with the preparation of this manuscript.

**Data availability** The data that support the findings of this study are available from the author upon reasonable request.

## Declarations

**Conflict of interest** The authors declare no competing interests.

**Ethical approval** This declaration is not applicable.

## References

- M.I. Jordan, T.M. Mitchell, Machine learning: Trends, perspectives, and Prospects. *Science* **349**, 255–260 (2015)
- S. Badillo et al., An introduction to machine learning. *Clin. Pharmacol. Ther.* **107**(4), 871–885 (2020)
- I.H. Sarker, Machine learning: Algorithms, real-world applications and research directions. *SN Computer Science* **2**, 160 (2021)
- M. Narhi et al., Machine learning analysis of extreme events in optical fiber modulation instability. *Nature Commun.* **9**, 1–11 (2018)
- L. Michaeli, A. Bahabad, Genetic algorithm driven spectral shaping of super-continuum radiation in a photonic crystal fiber. *J. Opt.* **20**, 055501 (2018)
- R. Maulik, N.A. Garland, J.W. Burby, X.Z. Tang, P. Balaprakash, Neural network representability of fully ionized plasma fluid model closures. *Phys. Plasmas* **27**(7), 072106 (2020)
- A.M. El-Mahalawy, K.H. El-Safty, Classical and quantum regression analysis for the optoelectronic performance of NTCDAP-Si UV photodiode. *Optik* **246**, 167793 (2021)
- R.N.U. Mahesh, A. Nelleri, Deep convolutional neural network for binary regression of three-dimensional objects using information retrieved from digital Fresnel holograms. *Appl. Phys. B* **128**, 157 (2022)
- A.N. Bakhtiyari, Y. Wu, D. Qi, H. Zheng, Modeling temporal and spatial evolutions of laser-induced plasma characteristics by using machine learning algorithms. *Optik* **272**, 170297 (2023)
- Y. Ding, A. Hu, J. Chen et al., A comparative study of classification models for laser-induced breakdown spectroscopy of Astragalus origin. *Appl. Phys. B* **129**, 125 (2023)
- G. Ghani-Moghadam, Prediction of soft X-ray laser gain value generated from laser plasmas by using a multilayer perceptron neural network. *Opt. Quant. Electron.* **55**, 683 (2023)
- S. Chatterjee, A.S. Hadi, *Regression analysis by example*, 5th edn. (Wiley, 2012)
- H. Daido, Review of soft x-ray laser researches and developments. *Rep. Prog. Phys.* **65**, 1513–1576 (2002)
- S. Suckewer, P. Jaegle, X-Ray laser: past, present, and future. *Laser Phys. Lett.* **6**, 411–436 (2009)
- G.J. Tallents, The physics of soft x-ray lasers pumped by electron collisions in laser plasmas. *J. Phys. D Appl. Phys.* **36**, R259 (2003)
- B. Rus et al., Multi-millijoule, “deeply saturated x-ray laser at 21.2 nm for applications in plasma physics.” *Plasma Phys. Control. Fusion* **44**, B207–B223 (2002)
- G. Ghani-Moghadam, A.H. Farahbod, Investigation of self-filtering unstable resonator for soft X-ray lasers. *Opt. Commun.* **371**, 154 (2016)
- D.L. Matthews et al., Demonstration of a soft x-ray amplifier. *Phys. Rev. Lett.* **54**, 110 (1985)
- P. Jaegle et al., High gain-production efficiency and large brightness X-UV laser at Palaiseau. *AIP Conf. Proc.* **332**, 25 (1995)
- A. Carillon et al., Saturated and near-diffraction-limited operation of an XUV laser at 23.6 nm. *Phys. Rev. Lett.* **68**, 2917 (1992)
- J.A. Koch et al., Observation of gain-narrowing and saturation behavior in Se x-ray laser line profiles. *Phys. Rev. Lett.* **68**, 3291 (1992)
- J. Nilsen et al., Prepulse technique for producing low-Z Ne-like x-ray lasers. *Phys. Rev. A* **48**, 4682 (1993)
- G.F. Cairns et al., Using low and high prepulses to enhance the J=0–1 transition at 19.6 nm in the Ne-like germanium XUV laser. *Opt. Commun.* **123**, 777 (1996)
- S. Le Pape, Ph. Zeitoun, Modeling of the influence of the driving laser wavelength on the beam quality of transiently pumped X-ray lasers. *Opt. Commun.* **219**, 323 (2003)
- M. Kim, K.A. Janulewicz, J. Lee, Pulse buildup from noise and intrinsic polarization of plasma-based x-ray lasers. *Phys. Rev. A* **84**, 013834 (2011)
- E. Khalilzadeh, A. Chakhmachi, J. Yazdanpanah, The dependency of the stochastic heating to the pulse shape in intense laser-plasma interaction. *Optik* **218**, 165249 (2020)
- G. Ghani-Moghadam et al., Parametric study of plasma active medium and gain saturation region in a Ne-like soft X-ray laser. *Contrib. Plasma Phys.* **61**, e202100042 (2021)

28. V.E. Guseva, A.N. Nechay, A.A. Perekalov et al., Investigation of emission spectra of plasma generated by laser pulses on Xe gas-jet targets. *Appl. Phys. B* **129**, 155 (2023)
29. A. Djaoui, S.J. Rose, Calculation of the time-dependent excitation and ionization in a laser-produced plasma. *J. Phys. B At. Mol. Opt. Phys.* **25**, 2745–2762 (1992)
30. A. Djaoui, “A user guide for the laser-plasma simulation code: MED103”, PAL-TR-96–099 (1996).

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.