Machine learning model designed to predict the amount of CO_2 produced by a small pellet boiler

Alexander Backa[1*], *Juraj* Drga[1], *Lucia* Martvoňová[1], and *Mária* Polačiková[1]

[1]KET, SjF, Žilinská Univerzita v Žiline, Univerzitná 1, 010 26 Žilina.

Abstract. Emissions, including CO_2 emissions, are generated during the combustion process. Perfect combustion of biomass should not lead to the formation of CO, but all carbon should burn perfectly and change to CO_2 by the oxidation process. Under real conditions, complete combustion never occurs and part of the carbon is not burned at all or only imperfectly to form CO. The aim of the work was to create a prediction model of machine learning, which allows to predict in advance the amount of CO_2 generated during the combustion of wood pellets. This model uses machine learning regression methods. The most accurate model (Gaussian process) showed a root-mean-square error, RMSE = 0.55. The resulting mathematical model was subsequently verified on independent measurements, where the ability of the model to correctly predict the amount of CO_2 generated in % was demonstrated. The average deviation of the measured and predicted amount of CO_2 represented a difference of 0.53 %, which is 8.8 % of the total measured range (3.08 - 9.2). Such a model can be modified and used in the prediction of other combustion parameters.

1 Introduction

A number of households also include a heat source for heating, in which energy is converted into heat, which is used to ensure thermal comfort. Mankind used mainly biomass, coal and later also natural gas as a fuel source. In recent decades, innovations have also taken place in the field of thermal technology, and less traditional sources have been used, which we also call alternatives. Today's users often demand comfort, which includes the operation of the combustion device itself, with those that are automated being preferred or without the need for constant control of the combustion mode, if at all. Such automated heat sources also include pellet boilers, where the operator provides refueling in the form of pellets, cleaning of the burner and ashtray and basic maintenance. The most used pellets are wood pellets, where cleanliness and low humidity are important. However, other types of pellets are increasingly being produced (e.g. from phytomass) or are enriched with additives such as bark, paper sludge, coffee and others.

When burning pellets in order to obtain energy, it is often necessary to maintain some combustion parameters in the required range of values so that undesirable conditions do not occur or, conversely, in order to achieve the necessary combustion outputs. These values can

---

* Corresponding author: alexander.backa@fstroj.uniza.sk

be influenced in time by setting the combustion device, the choice of fuel and also by changing the physic-chemical properties of the media that enter the combustion process [1]. Thanks to advanced diagnostic measurement methods, the combustion process can be extensively monitored [2]. In some applications it is necessary not only to monitor, but also to evaluate changes in quantities using prediction. Due to the complex nature of combustion, it is difficult to accurately predict the state of the investigated quantity, so creating a quality model is often difficult [3]. In most existing models, differential equations describing the combustion process are used in the calculation [4]. The disadvantage of most existing models is their complex computational technique and complex transformation with the necessary changes to the model [5]. An alternative to the existing mathematical models used is the use of artificial intelligence, for the purposes of prediction, which has been increasingly used in recent years in the research of combustion and combustion plants [6].

The study by Baruah et al. [7] describes the development of a model of biomass gasification based on an artificial neural network (ANN) in gasification plate gasifiers with a fixed bed. The prediction of the concentration of CH_4, CO, CO_2 and H_2 was calculated using a model based on ANN and the results show a good agreement with the experimental outputs by achieving a coefficient of determination - $R^2 > 0.98$ for CH_4, CO, CO_2 and H_2 [8].

One of the sub-areas of artificial intelligence is machine learning, in which it is possible to use classification or regression, where the created model is learned from a set of training data. In classification, the output is the classification of the training example into a selected set, and in regression, the output is a specific value corresponding to the training example [9]. To predicting some combustion parameters, which are based on other measured quantities of combustion, it is therefore appropriate to use regression techniques to obtain specific values. Using this technique, a model is also created in this article, where the amount of CO_2 emissions generated during the pellet combustion process is determined.

2 Procedures and methods

Input data for the regression model calculation were obtained from available measurements for 100% wood pellets (FIG. 1).

Fig. 1. 100% wood pellets

Wood pellets were made from spruce sawdust, which underwent a crushing process and subsequent pelletization. Basic parameters - physical and chemical composition, deformation temperature, moisture content and ash content of used wood pellets are given in Table 1. Fuel

properties were determined by thermogravimetric analyzer LECO TGA701, LECO CHN628 device for chemical analysis and also LECO AC 500 device for determination of higher heating value (FIG. 2).

Fig. 2. LECO CHN628 (on the left) and LECO AC 500 (on the right)

The performed measurements were realized on the LOKCA ÚSPOR boiler (FIG. 3), where a continuous measurement of parameters was performed.

To obtain the results, a heat source with a nominal heat output of 18 kW was used, which is primarily intended for burning wood pellets. This combustion device was equipped with an additionally mounted rotary burner the original. The retort burner was removed and replaced by a rotary burner with a nominal output of 25 kW.

The boiler contains a water exchanger with a volume of 55 l and a 200 l fuel tank in which a screw conveyor is located. The pellets are fed from above and transported to the rotary burner, where they are mixed with the combustion air, and by gradually burning and shifting, the burnt fuel is forced into the ashtray.

The control system was selected for a preset maximum boiler temperature of 70 °C and the fuel dosing to the rotary burner was adjusted accordingly. The heat produced from the heating system was dissipated by means of a heat exchanger to the cooling tower. The regulation was taken care of by the MAGA S. CONTROL device, which automatically controlled the combustion air supply, fuel dosing and maintained the maximum boiler temperature using the control electronics and our preset values.

Table 1 Basic characteristics of burned pellet samples

INPUT MEASUREMENT		
Proximate analysis (from TGA)	**Unit**	**100% wood pellets**
Moisture	%wt	3,86
Ash content	%wt	0,49
Volatile combustible matter	%wt	74,99
Fixed carbon	%wt	16,31
Ultimate analysis		
Carbon	%wt	48,78
Hydrogen	%wt	6,31
Nitrogen	%wt	1,25
Calorimetric analysis		
Higher heating value	MJ.kg-1	18,29
Lower heating value	MJ.kg-1	16,82
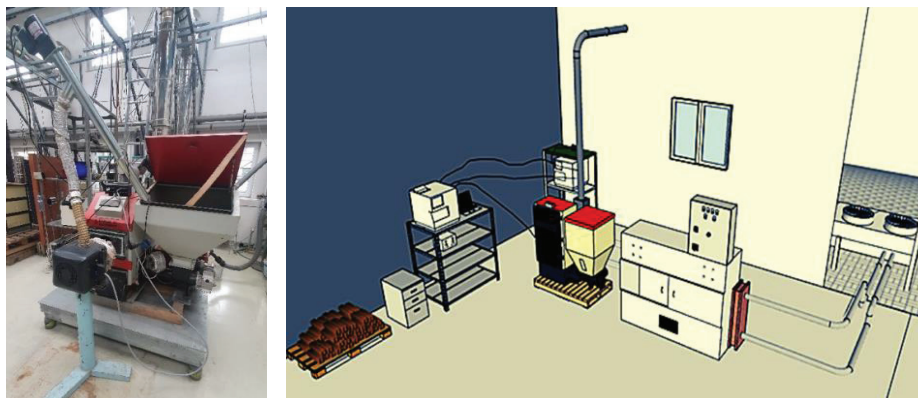
Fig. 3. boiler LOKCA ÚSPOR 18 with a rotary burner with an output of 25 kW (on the left) and boiler room scheme (on the right)

A total of 371 measurements were recorded in this way, which were used as inputs to the proposed model. The selected set of parameters - Table 2, which was selected for this model, can be changed depending on the need to improve the predictive quality of the model.

Table 2 Measured parameters during the combustion of 100% wood pellets

INPUT MEASUREMENT	
Quantity	**Unit**
Ambient air temperature	°C
Chimney temperature	°C
Chimney draft	Pa
O_2 in the flue gas	%
CO in the flue gas	ppm
Heat output	kW
CO_2 in the flue gas	%

To predict the amount of CO_2 produced, it would be appropriate to add parameters such as humidity and speed of the supplied air, abrasion resistance of the pellets and the amount of fuel supplied. When creating a comprehensive model for different types of fuels and combustion samples, such a model would also include characteristics that differ for different samples. It would be appropriate to add quantities such as chemical composition, moisture content and amount of ash to such a model. The selection of specific quantities was based on knowledge of the impact on CO_2 production and also on the basis of performed measurements and the availability of measuring devices for the combustion equipment and the combustion process. If significant influence quantities on the investigated parameter are neglected, it is necessary to keep them constant in time in new predictions or as changing as during the measurements from which the mathematical model was created. Only thanks to this can the prediction be correct and sufficiently accurate.

Subsequently, the parameters were entered into the Matlab program, where using the regression module - Regression Learner App, a model was built with the input parameters listed in Table 1, including the amount of CO_2 emissions, which were considered as the output (determined) quantity. The prediction model was developed to determine a specific value depending on the selected parameters, while the evaluation of success depends on the performance of tests in k-fold cross-validation using calculated parameters - coefficient of determination (R^2), mean square deviation (RMSE) and also root mean square error (MSE) [10]. The input data set is divided into subsets. One subset serves as a test set, the remaining subsets serve as training sets. The classifier trains the model on a trained set and uses the test set to test the accuracy and performance of this model. This process is repeated several times with different subsets. In multiple cross-validation, a value of 10 was chosen, which is most often used because it is the most objective [11].

$$R^2 = 1 - \sum_{i}^{m} \frac{(Mi - Pi)^2}{(Mi - Si)^2} \; (-) \tag{1}$$

$$RMSE = \sqrt{\frac{1}{m} \sum_{i}^{m} (Mi - Pi)^2} \; (-) \tag{2}$$

$$MSE = \frac{1}{m} \sum_{i}^{m} (Mi - Pi)^2 \; (-) \tag{3}$$

Where M_i is the measured value, P_i is the predicted value, S_i is the mean value and m is the number of data points. These parameters are used to compare individual models, but they also evaluate the success of the descriptive function, which is based on a specific data set. RMSE is used for the purpose of comparing the observed values with the values predicted by the model. A value of 0 would mean a perfect match of the data, so a lower RMSE value is generally better. The coefficient of determination R^2 shows how the regression prediction approaches the actual data. The value of R^2 normally ranges from 0 to 1, and if this value is closer to the number 1, the regression predictions are better expressed using the proposed regression function. The value of R^2 is a measure of how well the measured results are replicated by the model [12].

For control purposes, it is appropriate to predict the new data on an independent measurement using the created model. The model created in this way can then be verified and used in predicting the selected quantity in independent measurements, thanks to which it is possible to determine the approximate value of the sought quantity without the need for its measurement.

The regression models used in the calculation in this research were linear regression, regression decision trees, supporting vector machines, Gaussian process regression models, regression tree sets, and neural networks. After testing all types of machine learning, the one with the lowest RMSE was chosen.

3 Regression model of machine learning CO_2 emissions

After reading the measured data, it was necessary to choose a validation method, while for a specific application, cross-validation with 10 subsets was selected. This type of validation appeared to be the most appropriate when performing various experiments with changing validation settings.

Subsequently, testing of individual types of regression models was started, while the lowest value of the mean square deviation, RMSE = 0.035, was shown by the regression model of the Gaussian process. Figure 4 shows the selected regression model, where the measured and prediction data are shown. The "x" axis represents the measurement sequence and the "y" axis the value of CO_2 emissions in %, while it is possible to graphically observe an excellent agreement, which is also confirmed by the coefficient of determination R^2 in this case with 1.00 and the root mean square error with MSE = 0.001. Figure 5 shows the deviations of the predicted values from the actual values entering the model using the prediction function. Matlab code was subsequently generated for this model. Using this code, it is possible to further predict (also) states other than those that were inserted into the model by entering all input quantities with the required values. In this way, it is possible to predict new model situations.
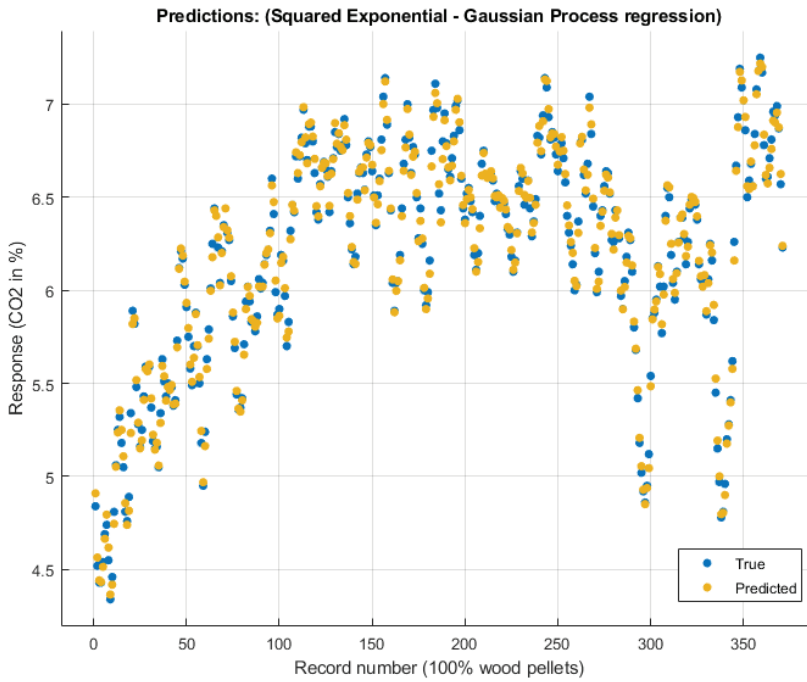
Fig. 4. Comparison of CO_2 emissions (%) of predicted data (yellow) from measured data (blue)
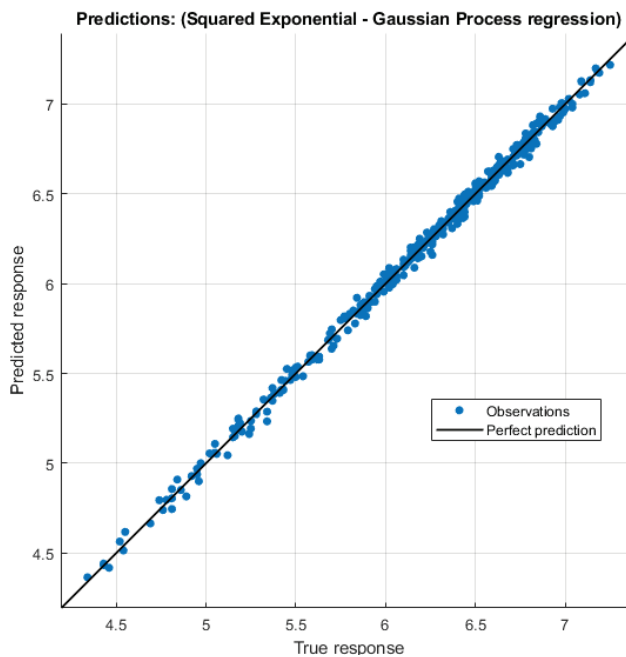
Fig. 5. Prediction accuracy indicator

To verify the functionality of this model, new special measurements (180 special measurements) were performed using the same type of fuel and also the combustion equipment. These new data were not the subject of the originally created model.

When predicting new data, it is necessary to keep all settings and conditions that entered the model the same as in the measurements for this model, where only some specific parameters were selected. A new 180 measured records were inserted into the created model for the prediction of CO_2 emissions and a value in % was determined for each measurement. This value was then compared with that predicted by the proposed prediction model, compiled from the originally measured data. In this way, the new error RMSE = 0.55 with the value MSE = 0.31 was determined, as well as the average deviation of the predicted data from the measured values, namely 0.53 %. Such a result shows that the proposed model is functional and can be used to predict the amount of CO_2 generated in the flue gas. A graphical comparison of the predicted and measured data is shown in FIG. 6.
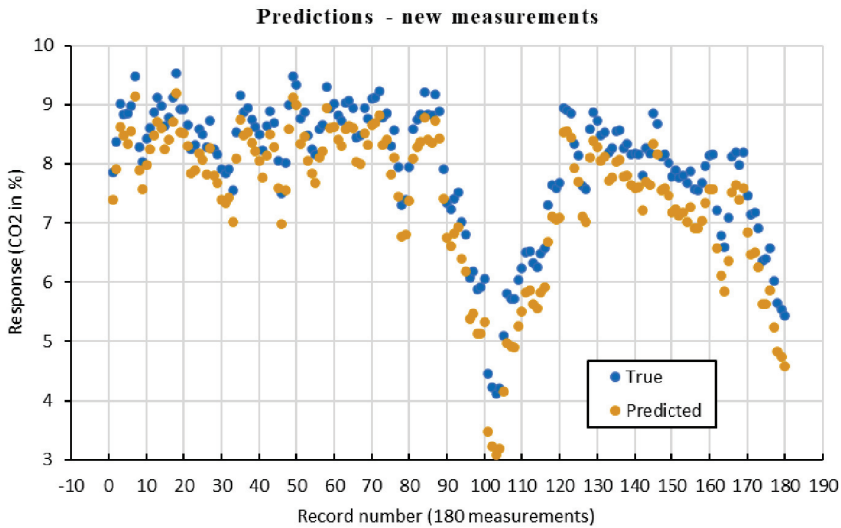
Fig. 6. Prediction of a new model situation for 180 measured records using the proposed model

When using such a model in practice, it is necessary to realize that the model will always show a certain deviation from the real measured quantity. However, this deviation can be eliminated by adding other quantities that have a significant effect on the parameter sought. It is also possible to maintain the "learning" of the regression model by adding new measurements.

4 Conclusion

An interesting concept for further direction would be to link the information obtained from a specific model with specific combustion phenomena and their application to create a comprehensive combustion model using artificial intelligence tools. It would also be possible to implement such a model directly in the combustion plant, which would be improved through further measurements and based on the model could correctly select the necessary parameters to achieve the optimal value of power or other parameters (e. g. emissions). From an environmental point of view, the prediction models could also be used for monitoring and evaluation of produced emissions at a specific time, using combustion plants connected via a network to a central register, which would obtain predicted values of discharges of certain emission values based on sent settings of individual combustion plants.

In this paper, a regression model was built, which allows the calculation of the selected parameter - CO_2 emissions, using the input measured parameters. The usability of such an application lies in the possibility of the correct setting of the combustion device, as well as the adoption of appropriate measures for efficient and high-quality combustion in order to achieve the required amount of CO_2 production. It is also possible to predict the values of a particular quantity using the model without the need for its continuous measurement. In a similar way, it is possible to build a model that would allow the prediction of other parameters.

The success of the proposed model was later determined by independent measurements on the same boiler using also wood biomass pellets. It was found that the predicted values were different on average by 0.53 % compared to those measured. This value represents an error of 8.8 % of the total measured interval (3.08 - 9.19).

References

1. JANDAČKA, J. – MIČIETA, M. – HOLUBČÍK. M. – NOSEK, R., (2016): *Inovácie na zefektívnenie procesu spaľovania biomasy.* Žilina: EDIS - vydavateľské centrum ŽU, 265 s. ISBN 978-80-554-1236-8.

2. HERNÁNDEZ, R. – BALLESTER, J. Flame imaging as a diagnostic tool for industrial combustion. Combustion and Flame. (2008). 155. s. 509-528.

3. HAN, Z. – LI, J. – HOSSAIN, M. – XU, CH. Prediction of combustion state through a semi-supervised learning model and flame imaging. Fuel. (2021). 289 s.

4. ZHENG, Z. et. al. Progress in the Application of Machine Learning in Combustion Studies. ES Energy & Environment. (2020).

5. MORARI, M. – LEE, J. H. Model predictive control: past, present and future. Computers & Chemical Engineering. (1999). 23(4-5). s. 667-682.

6. ELMAZ, F. – YÜCEL, Ö. – MUTLU, A. I. Predictive modeling of biomass gasification with machine learning-based regression methods. Energy. (2020). 191, ISSN: 0360-544.

7. BARUAH, D., BARUAH, D.C., HAZARIKA, M.K., (2017). Artificial neural network based modeling of biomass gasification in fixed bed downdraft gasifiers. Biomass and Bioenergy 98, s. 264–271.

8. MUTLU, A.Y., YUCEL, O., (2018). An artificial intelligence based approach to predicting syngas composition for downdraft biomass gasification. Energy 165, s. 895–901.

9. ZHOU, Z.-H. A brief introduction to weakly supervised learning. National Science Review. (2018). 5(1). s. 44-53.

10. BINOD, J. R. - RAVI, M. R. Development of numerical model for analysis of biomass based pottery furnace. [autorcafé], (2019), s.l.: Summer Research Fellowship Programme of India's Science Academies.

11. AMBROISE, C. - MCLACHLAN G.J. Selection bias in gene extraction on the basis of microarray gene-expression data. (2002). Proc Natl Acad Sci, 99 (10), pp. 6562-6566

12. DRAPER, N. R. - SMITH, H. Applied Regression Analysis. (1998). Wiley-Interscience. ISBN 978-0-471-17082-2