FEATURE SELECTION FOR COAL HEATING LEVEL ESTIMATION IN THERMAL POWER PLANTS

by

Sanja M. VUJNOVIĆ [a*], ***Dejan B. CVETINOVIĆ*** [b], ***Vukman V. BAKIĆ*** [b], ***and Željko M. DJUROVIĆ*** [a]

[a] School of Electrical Engineering, University of Belgrade, Belgrade, Serbia
[b] VINCA Institute of Nuclear Sciences, National Institute of the Republic of Serbia, University of Belgrade, Belgrade, Serbia

Original scientific paper
https://doi.org/10.2298/TSCI240124124V

Several recently signed environmental agreements and protocols emphasize the global need to reduce GHG emissions, with a focus on limiting coal consumption due to high NO_x and CO_2 emissions. However, many countries, including those in the Western Balkans, rely heavily on coal for electricity generation. The outdated thermal power plant infrastructure in these regions poses a major challenge when it comes to meeting modern environmental standards while maintaining efficiency. This study is part of the more comprehensive research which aims to develop an expert system that utilizes existing measurements to estimate key parameters crucial for both energy production and pollution reduction. The focus is on Serbian thermal power plants, particularly plant Nikola Tesla unit B1. One of the critical parameters for optimizing thermal power plant control loops is the heating value of coal, which is challenging to measure in real time due to the coal's varying chemical compositions and caloric values. This paper examines 74 different parameters measured in 59 instances to estimate the hating value of coal at unit B1. Through detailed analysis and feature selection methods, including linear regression, this research aims to identify the most informative parameters for estimating the heating value of coal, which will improve the control system that enables more efficient and environmentally friendly power generation in coal fired thermal power plants.

Key words*: feature selection, linear regression, thermal power plants, coal heating value*

Introduction

For decades the scientific community has been aware of the need to reduce GHG emissions and their harmful impact on climate [1]. This has been formally and publicly acknowledged in 1992 at the UN Conference on Environment and Development, also known as Earth Summit, where the UN Framework Convention on Climate Change (UNFCCC) was established. Soon after, in 1997, the Kyoto Protocol was adopted which specified the obligations of the member states in reducing greenhouse gas emissions in the coming period. The Paris Agreement on Climate Change, signed in 2016 by the 195 members of the UNFCCC, presents an action plan to limit global warming. Countries have committed to keeping the rise in average global temperatures well below 2 °C compared to pre-industrial levels [2].

* Corresponding author, e-mail: svujnovic@etf.bg.ac.rs

As a result of the adopted obligations, the most industrially developed countries of the world have introduced numerous programs to monitor, reduce and limit GHG emissions. Considering that CO_2 is the most significant GHG, the biggest challenge of the Paris Agreement is in reducing the consumption of fossil fuels, primarily coal, because it has the highest emission factor, *i.e.* the highest emission of CO_2 per unit of energy used.

Out of all fossil fuels, coal reserves are by far the largest and the most evenly distributed in the world. Their mass exploitation ensures a stable and relatively low price on the international market. That is why neither the most developed countries of the world nor developing countries can give up coal in the structure of primary energy sources to ensure a stable supply of their consumers at affordable prices. For that reason, thermal power plants (TTP) are still the main producer of electricity in many countries of the world, including the Western Balkan region with 54% of all produced power [3]. Additionally, most of the plants which are currently in use have been producing energy for decades, using technology which was state of the art in the period in which they were built, but which is now mostly outdated, both in terms of performance as well as in terms of satisfying new requirements regarding sustainable production which has low impact on the environment. In order for these plants to adhere to recent pollution policies and recommendations made by the European Commission regarding the limiting of NO_x, SO_2 and CO_2 gases, they often need to sacrifice their efficiency and energy production [4]. Furthermore, some of the Eastern European plants [5] are so outdated that the only solution is to shut the plant down entirely and to build a modern one which can adhere to current standards. The aim of this research is to help the energy production systems to reduce harmful emissions using the measurements and technology which is available on site, with minimal investments necessary, concentrating on specific problems facing Serbian thermal power plants, specifically TPP *Nikola Tesla* unit B1 – TENT B1. This will be done by creating an expert system which uses existing measurements to estimate relevant parameters crucial to power production and pollution reduction. Using these estimates the existing control loops in TPP can be augmented to take into account additional criteria related to pollution.

Steam boilers in TENT are designed for the domestic lignite from the coal mine Kolubara as the main fuel with lower heating value of 6.700 MJ/kg, moisture content of 47.8%, ash content of 19%, and sulfur content of 0.5% [6]. However, in today's energy sector, as a consequence of the current unfavorable geopolitical events, there are fewer and fewer power systems that have a guaranteed coal supply from specific mines. As it happens, there is a necessity on almost daily basis to mix coal that comes from different locations, and therefore, with different chemical compositions and caloric value. Therefore, the authors' experience is that dramatic changes in the quality of coal occur in small time intervals, which adversely affects the quality of regulation, and also the concentration of gases released into the atmosphere. This combination of circumstances causes special challenges for control systems, and the first and logical assumption of such solutions is dynamic knowledge of coal parameters [7].

Since the heating value of coal is one of the most important parameters which needs to be taken into account when considering optimization of TPP control loops and minimizing pollution, it is only natural to implement the system which measures it directly. This, however, is very challenging in practice. The measurement of heating value of coal is not possible in real time because the extraction of the coal, its processing and testing is a process which requires a lot of time. The second-best solution, therefore, is to attempt to estimate its value. Unfortunately, estimation of this parameter based on all the available measurements is an extremely non-linear multivariate problem which is not clearly solved in [8, 9].

The research presented in this paper deals with the problem of estimating the heating value of coal at the TENT B1 TPP in Obrenovac, Serbia, taking into account 74 different parameters measured at 59 different time instances. An attempt is made to perform detailed analysis of these parameters (*i.e.* features or predictors as they are called in the literature) and to select a subset of them which are most informative for estimating the heating value of coal. Linear regression will be used to test the selected features, and several different methods for feature selection will be tested.

Case study

Coal fired TPP are main producers of electricity in Serbia and they are mostly designed according to knowledge and technology that was available in the 1960's and 1970's. Steam boilers use lignite (low quality coal) as a fuel and are designed with subcritical steam parameters. The design efficiency levels of such boilers are 87-88% for coal of medium quality, lower heating value of 6700 kJ/kg. Any decrease in the efficiency of the boiler means that its heat losses have increased, that is, less electricity has been produced than the amount that was possible from the energy available in the coal. To illustrate, tab. 1 shows the heat losses in the boilers and, therefore, the loss in electricity production in units with a nominally installed power of 300 MW, 350 MW, and 670 MW.

Table 1. Annual losses when reducing the efficiency of the boiler by 1%

Losses	Boiler with nominal power [MW]		
	300	350	670
Coal fed into the boiler, but not used [10^3 tonne per year]	28	33	63
Lost heat [10^6 MJ per year]	210	247.5	472.5
Less electricity produced [GWh per year]	20.6	24.3	46.3
Monetary equivalent of non-produced electricity (calculated with a price of 0.045 €/kWh) [million € per year]	0.930	1.090	2.080

Although the heating value of coal primarily affects the efficiency of electricity production, for the optimization of the combustion process it is necessary that the other parameters of coal are in nominal values. Specifically, most of the boilers in Serbian power plants are designed for lignite with moisture content of 47.8%, ash content of 19%, sulfur content of 0.5%, ash melting point of 1345 °C, and so on. Variations in these parameters lead to disturbances in the combustion process, as described in tab. 2. It is evident that the most common causes of reduced boiler efficiency are either inadequate coal quality or inadequate functioning of the combustion devices [10]. These causes lead to significant changes in several critical parameters of the combustion process, primarily an increase in the two main heat losses of the boiler:

– loss of heat carried by flue gases into the atmosphere due to the increase in temperature of the flue gases at the exit from the boiler and
– heat loss due to the removal of solid unburned materials from the boiler through slag and ash [11].

The influence of the heating value of coal on the efficiency of the boiler, as well as these two types of heat losses, can be seen in tab. 3. What is interesting to note is that there is an almost linear dependence of boiler efficiency and the heating value of coal.

Table 2. Disturbances in the combustion process and operation of TPP caused by changes in coal quality

Parameter	Deviation	Effect
Heating value	Too high	• Overheating and damage to the burner • Increased scaling
	Too low	• Increased coal consumption • Increased transport costs • Possible power plant failure
Ash content	Too high	• It can cause an increased participation of fine particles in the grinding product • Increased erosion of parts of the mill, pipes and burners • Increased scaling • Increased emission of solid particles
Volatiles content	Too low (<10%)	• Unstable flame • Increased consumption of liquid fuels
Moisture content	Too high	• Coal flow becomes problematic • Grindability of coal may be weakened • Reduction of combustion efficiency
Hardgrove index	Too low	• May affect grinding capacity
Initial melting temperature	Too low	• Increased scaling
Coal feeder size	Too high (>150 mm)	• Reduced combustion efficiency

Table 3. Boiler efficiency change with respect to the heating value of coal

Heating value of coal [kJkg^{-1}]	Boiler efficiency [%]	Loss in exhaust flue gases [%]	Loss in unburned solids [%]	All other losses [%]
9289	86.80	11.60	0.99	0.61
9211	86.43	11.90	1.04	0.63
8326	85.54	12.21	1.54	0.71
7992	85.51	12.44	1.34	0.71
7816	85.00	12.75	1.44	0.81
6700	83.64	← Approximately determined boiler efficiency for guaranteed heating value of coal		

It is useful to analyze which specific critical points affect the efficiency of the TPP block, with respect to the heating value of coal [12]. There are, in fact, three phenomena which are significant, and they are given and described in tab. 4.

Although in literature, as well as on the market, there are various methods and devices that measure the heating value of coal in-situ [13], this approach is often, if not impossible, certainly overly complicated, expensive or inappropriate. This is mainly due to the fact that heating value of coal would have to be measured on all conveyor belts to individual mills and the obtained values would be the characteristics of the particular samples which are taken. Also, each of the devices would require an evaluation time between 300 seconds and 900 seconds, which is already comparable with the dynamics of changes in the heating value of coal at plants where the authors of this paper have experience. Accordingly, the idea to design an algorithm that will estimate the heating value of coal based on available on-line measurements is of exceptional importance.

Table 4. Parameters depending on the fuel quality and which affect the efficiency of the unit

Number	A phenomenon that indicates a reduced efficiency of the block	The most likely cause
1.	Temperatures of flue gases at the boiler outlet are higher than the design and/or increased amount of water for steam temperature regulation	• Coal quality deteriorated - outside the guaranteed composition (absence of fuel homogenization system, improper operation of crushers) • Non-sealing of the boiler, as a result of which the burning time of fuel particles in the combustion chamber is extended, so the burning is often done in the initial part of the channel of the subsequent heating surfaces • Disturbance in heat transfer in the relationship flue gas – working medium due to the formation of deposits on the outside and/or inside of the pipes and the heating surfaces of the boiler
2.	Increased content of unburned solids in slag and ash	• The quality of coal deteriorated – outside the guaranteed composition and granulation • Improper functioning of the combustion device, primarily mills (coarser fineness of grinding and/or increased humidity of coal powder) • Leakage of the boiler, which causes an irregular fuel-air ratio at the burner levels, *i.e.* insufficient secondary air, so the fuel jet is insufficiently turbulized in the combustion chamber, *i.e.* insufficient speed of the *air mixture* and/or secondary air at the mouth of the burner
3.	Not achieving nominal steam parameters	• Improper functioning of certain parts of the boiler plant (water heater, evaporator, superheater, *etc*.) – most often in direct connection with the aforementioned causes

Multiple linear regression

The heating value of coal influences many of the parameters described in the previous section. It is therefore, clear that measuring several key variables can help to estimate the heating value of coal. Linear regression maps the relationship between input variables (also known as features or predictors, x) and output variables, y. Three different cases of linear regression, depending on the number of predictors: simple linear regression (one predictor x and one output y), multiple linear regression (several predictors and one output), and multivariate linear regression (several predictors and several outputs) [14]. The goal of this paper is to estimate the heating value of coal depending on the several measured quantities, so the multiple linear regression approach is going to be used.

Multiple linear regression in the case of n measurements and N predictors can be represented:

$$Y = X\beta + \epsilon \tag{1}$$

where $Y = [y_1\ y_2 \cdots y_n]^T$ is the vector of outputs, $\beta = [\beta_0\ \beta_1 \cdots \beta_N]^T$ – the vector of regression coefficients, the vector of random errors is denoted by $\epsilon = [\epsilon_1\ \epsilon_2 \ldots \epsilon_n]^T$, and X – the augmented sample matrix with N predictors:

$$X = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1N} \\ 1 & x_{21} & x_{22} & \cdots & x_{2N} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nN} \end{bmatrix} \tag{2}$$

This model can be used if the three assumptions are met:

– The process is indeed linear, *i.e.* it holds that

$$E\{y_i\} = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_N x_{iN}, \quad i = 1, \ldots, n$$

– $\text{var}(y_i) = \sigma^2$, $i = 1,...,$ *n i.e.* the variance of error terms is the same, and
– $\text{cov}(y_i, y_j) = 0$, for all $i \neq j$, *i.e.* error terms are uncorrelated.

The first assumption states that the model is linear and that there are no additional terms that are needed to estimate output variable y. Strictly speaking, this condition is not satisfied for the problem at hand in which there are surely several predictors which are connected to the output variable in a non-linear fashion. This will be studied in more detail in the future research, but for the purpose of rating the informativeness of the predictors it is assumed that this non-linear association can be modelled as linear.

Additionally, in order to obtain statistically significant results, and to avoid the problem of weak conditioning, the number of measurements needs to exceed the number of parameters of the model: $n > N + 1$. With all this in mind, the least squares estimate of regression coefficients, $\hat{\beta}$, is given:

$$\hat{\beta} = \left(X^T X\right)^{-1} X^T Y \tag{3}$$

This estimate minimizes the error sum of squares:

$$SSE = \sum_{i=1}^{n} \left(y_i - \hat{y}_i\right)^2 = \sum_{i=1}^{n} \left(y_i - x_i \hat{\beta}\right)^2 = \left(Y - X\hat{\beta}\right)^T \left(Y - X\hat{\beta}\right) \tag{4}$$

where $\hat{y}_i = x_i \hat{\beta}$ is the linear regression estimate of the output and $x_i = [1\ x_{i1}\ x_{i2} \cdots x_{iN}]$ is the augmented predictor vector of i^{th} measurement. Apart from SSE, other sum of squares values which can be useful when discussing linear regression are regression sum of squares (SSR) and total sum of squares (SSTO) given:

$$SSR = \sum_{i=1}^{n} \left(\bar{y} - \hat{y}_i\right)^2 \tag{5}$$

$$SSTO = \sum_{i=1}^{n} \left(y_i - \bar{y}\right)^2 = SSR + SSE \tag{6}$$

where

$$\bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i$$

is the sample mean of outputs y_i. These parameters indicate which amount of total variation of the output (SSTO) is due to variation around the estimated regression-line (SSE) and which is due to variation around the sample mean (SSR).

Linear regression quality metrics

It is clear from section *Case study* that the nature of the thermal power plant process is such that there are more descriptors than measurements, so standard linear regression approach needs to be modified. Specifically, there is a need to reduce the number of features by eliminating the ones that are not informative for coal energy level estimation and keep those that are. The problem is simply formulated: from the initial N predictors, $p < N$ predictors should be chosen which, according to certain criterion, will model the output variable in the best way, through a linear regression model:

$$Y = X_p \beta_p + \epsilon \tag{7}$$

This equation is similar to eq. (1), where Y is the output vector and ϵ is the error vector. The difference is that only p out of the original N predictors are used, so the vector of regression coefficients is now $\beta_p = [\beta_0 \beta_1 \cdots \beta_p]^T$, and the augmented sample matrix with p predictors becomes:

$$X_p = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix} \tag{8}$$

In order to assess which subset of p features is most informative, some metrics for assessing the quality of linear regression are needed. Perhaps the most intuitive metric which describes the quality of linear regression is the coefficient of determination, R^2, which quantifies which proportion of total variation of output can be attributed regression of descriptors, x:

$$R^2 = \frac{SSR}{SSTO} = \frac{\hat{\beta}^T X^T Y - n\bar{y}^2}{Y^T Y - n\bar{y}^2} \tag{9}$$

It is clear that R^2 has values between 0 and 1, where $R^2 = 1$ means that all measurements lie on the linear regression-line, while $R^2 = 0$ means that the estimated regression-line is horizontal. Colloquial interpretation of coefficient of determination is that *$R^2 \times 100\%$ of variation in output is explained by the variation in predictor*. Even though one should be careful when interpreting the value of this coefficient [15], generally it can be said that the higher the R^2 value, the stronger the linear relationship between the predictor (or the set of predictors) and the output. More specifically, while evaluating which subset of p predictors better describe the model, the R^2 test can be used to evaluate the model from eq. (7):

$$R_p^2 = \frac{\hat{\beta}_p{}^T X_p^T Y - n\bar{y}^2}{Y^T Y - n\bar{y}^2} \tag{10}$$

Another metric which can be found in the literature is the variance of the estimator. When p predictors are used, the variance of the predictor is defined:

$$s_p^2 = \frac{SSE}{n-p} = \frac{\left(Y - X_p \hat{\beta}_p\right)^T \left(Y - X_p \hat{\beta}_p\right)}{n-p} \tag{11}$$

The smaller the variance the better the prediction. This metric is especially useful when comparing the performance of linear regression models with different predictor sets.

There is a third metric that can be used to assess the performance of linear regression for specific purpose of finding the best subset of p predictors from an initial large number of parameters $N > p$. This measure takes into account both the variance of the estimator and its bias. The goal is to find such a model that will make a good compromise between these two components. It can be defined in two alternative ways:

$$C_p = p + (n-p)\frac{s_p^2 - s_N^2}{s_N^2} \tag{12}$$

or

$$C_p = \frac{SSE_p}{s_N^2} - (n - 2p) \tag{13}$$

The second definition is somewhat simpler numerically, however, the first gives the possibility for a physical interpretation of the results. If the bias is small for a particular model, then the second summation is close to zero and the measure C_p becomes close to the parameter p. Therefore, the diagram C_p, $p = 2, 3,..., N$ is usually sketched together with the diagram $C_p = p$.

Hypothesis testing

Hypothesis tests are performed in order to compare how certain subsets of predictors perform when it comes to multiple linear regression models. In regression analysis usually a so-called F-test is used which is a test that uses the F-distribution [14].

The first test that can be found in the literature is the overall regression test which forms hypothesis:

$$H_0 : \begin{bmatrix} \beta_1 & \beta_2 & \cdots & \beta_N \end{bmatrix} = 0_{1\times N} \tag{14}$$

In other words, the null hypothesis states that none of the predictors hold any information about the output. This test can be expressed with random variable:

$$F = \frac{\dfrac{SSR}{N}}{\dfrac{SSE}{n-N-1}} \tag{15}$$

This random variable is distributed as $F_{N,n-N-1}$ when H_0 is true, so the hypothesis is rejected when $F > F_{\alpha,N,n-N-1}$, where α is a selected threshold.

The second hypothesis test is the test on a subset of β. It takes into account the original set of N predictors, S_f. This set is divided into a subset of p predictors which will be used for regression, S_i, and the subset of $N - p$ predictors which will be rejected, S_d. Naturally it holds that $S_f = S_i + S_d$. The hypothesis in this case says that the predictors from the set S_d should be omitted because their influence on the output variable is non-existent. Therefore, the hypothesis is:

$$H_0 : \beta_d = 0_{1\times(N-p)} \tag{16}$$

and the test can be expressed with the random variable:

$$F = \frac{\dfrac{SSR_f - SSR_i}{N-p}}{\dfrac{SSE_f}{n-N-1}} \tag{17}$$

where SSR_f and SSR_i are regression sum of squares of the sets of predictors S_f and S_i, respectively, and SSE_f is the error sum of squares of the set S_f. The random variable from eq. (17) is distributed as $F_{N-p,n-N-1}$ when H_0 is true, so the hypothesis is rejected when $F > F_{\alpha,N-p,n-N-1}$, where α is a selected threshold.

Feature selection

In an attempt to include in the model all the signals which could directly or indirectly indicate the heating value of coal, the authors have oversized the initial set of predictor signals. This is usually done when it is not initially clear which set of predictors is informative for the problem at hand [16]. The initial set consists of 74 measurable variables that are available and archived in the existing data-logging system. Some of these variables should naturally be in the set of predictors, such as active electrical power, total fuel quantity, oil flow, flue gas tempera-

ture or steam flow in front of the turbine. However, as a precautionary measure so that some of the hidden connection between signals and output are not lost, the signals that at the first glance do not have a connection with heating value of coal are included in the initial predictor set. Thus, among the initial 74 predictors there were signals such as feed water temperature, total injection quantity, steam pressure in front of turbine, NO and O_2 concentration in flue gases, primary and secondary air-flow, and so on.

Regardless of the modern computational and memory capacity of the computers used to process this data, it turns out that such a large set of predictors is extremely demanding, especially considering the relatively modest number of laboratory measurements that were available. This problem is often indicated in literature. Namely, during fourteen months in 2021 and 2022, a laboratory check of the heating value of coal was performed approximately once a week. This procedure was carried out by taking samples from different physical locations in the coal delivery sector for one hour. These samples were mixed in order to obtain a homogeneous sample that will be sufficiently representative for the observed time interval. Thus, 59 samples and 59 measurements were obtained. Considering such a limited sample size, the multivariate model should not have too many predictors [17]. A large number of predictors in relation the number of experiments, or measurements, generates a technical problem of weak conditioning of the system of equations that needs to be solved. On the other hand, it is intuitively clear that this scenario of too many parameters inevitably leads to overtraining of the model.

Various predictor selection techniques are known in [18]. The best-known and most frequently used approaches are *All possible subsets* method and *Stepwise selection* [14]. However, applying these techniques to the initial set of 74 variables was almost impossible, especially for the *All possible subsets* method. Therefore, it has been decided that the first step will be a preliminary reduction of dimensionality of predictor set in a simple but intuitively quite justified way. Namely, the coefficient R^2 from eq. (10) is calculated between the output variable and each of the potential predictors. In that way 74 different values of this parameter were obtained, indicating the strength of linear connection between output variable and each of the predictors.

The highest R^2 index of 0.498 was obtained between the temperature of the steam before the superheater and the heating value of coal. A graphical representation of their correlation is given in fig. 1. Similarly, a high level of correlation was obtained between the active electrical power of the plant and the heating value of coal, fig. 2.

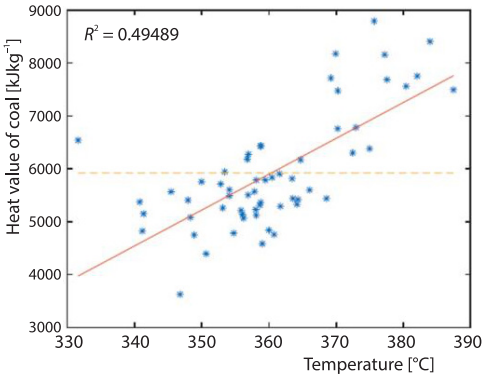

Figure 1. Variable 6: steam temperature before superheater

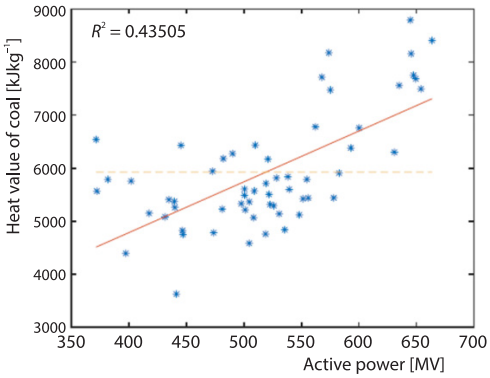

Figure 2. Variable 1: active power

On the other hand, for a large number of predictors, a very small value of R^2 was obtained. Such a weak correlation is generally expected for many signals. The correlation between heating value of coal with air temperature in front of Luv 1 is given in fig. 3, while the correlation with the amount of injection in the second stage is given in fig. 4.

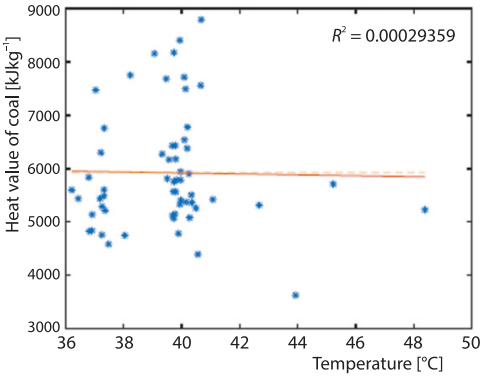

Figure 3. Variable 13: air temperature in front of Luv 1

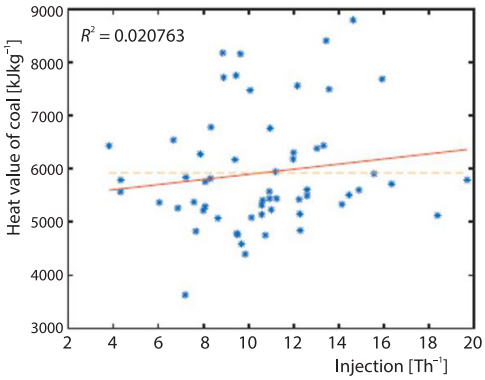

Figure 4. Variable 5: amount of injection in the second stage

Based on these diagrams and the obtained values of R^2, two interesting facts can be noted. The first is that the highest R^2 was not obtained for the expected predictors. Namely, if the losses in the TPP were ignored, as well as the variation in the enthalpy of the obtained steam in front of the turbine, a good indicator of the heating value of coal could be obtained as the quotient of the steam flow in front of the turbine and the sum of the fuel fed into the boiler. These quantities, however, did not stand out with the highest values of R^2 with regards to the heating value of coal. Also, based on figs. 3 and 4 it is noticeable that there are measurements that deviate significantly from the rest of the observations and can be considered, in this context, as outliers. The measurements in question were probably obtained under specific, unusual circumstances. These circumstances can be different, starting from extremely non-heating to extremely heating coal. It is possible that in some periods there were failures of certain mills, or plant operation with unusually high support of fuel oil, and the like. In any case, due attention needs to be paid to these outliers because they can significantly distort the picture of the correlation and determination of certain predictors and regression outputs.

In order to exclude from the further analysis all physical quantities that are likely not useful from the aspect of multivariate linear regression, the threshold $R_T^2 = 0.3$ was set. In this way, 20 potential predictors whose coefficients are higher than the adopted threshold were singled out. These signals are given in tab. 5.

This set of 20 predictors is sufficiently small to enter the predictor selection procedures. The main idea of this work, apart from the multivariate linear regression design itself, is to illustrate the predictor selection procedures in case of modelling the heating value of coal in TPP. On the other hand, as various techniques for the selection of input variables are known in the literature, it was interesting to make a short comparative analysis. These procedures are methodically described in [14] and the authors of this paper followed the instructions and advice that could be found there. The further is divided into two subsections, the first of which is devoted to the *All possible subsets* approach, and the second to the *Stepwise selection* technique.

Table 5. Initial set of 20 predictors

Symbol	Predictor description	Symbol	Predictor description
x_1	Active electrical power	x_{11}	NO concentration in the gas analyzer
x_2	Water temperature behind the economizer	x_{12}	O_2 concentration in the gas analyzer
x_3	Pressure of feed water in the bottle	x_{13}	Flue gas temperature in front of superheater 3 on the right side
x_4	Sum of steam injection	x_{14}	Flue gas temperature in front of superheater 3 on the left side
x_5	Steam temperature before superheater	x_{15}	Fuel oil consumption
x_6	Total air for the boiler	x_{16}	Steam flow in front of high pressure bypass
x_7	Air-flow for roasting	x_{17}	Steam pressure in the high pressure collector
x_8	Total fuel	x_{18}	Steam temperature in the high pressure collector
x_9	Flue gas temperature before Luv 1	x_{19}	Feed water flow to the boiler
x_{10}	Flue gas temperature before Luv 2	x_{20}	Feed water pressure behind the feed head

All possible subsets

The idea behind the *All possible subsets* approach is to first choose $p = 2$ of the initial N predictors (in this case $N = 20$). This can be done in $\binom{N}{2}$ ways, and for each of these

$$\frac{N(N-1)}{2}$$

ways a linear regression model is formed, as in eq. (7) and its efficiency is measured based on some criterion CRIT. The subset with the best value $CRIT_2$ is chosen. Then the procedure is repeated for $p = 3, 4, ..., N$. Finally, the diagram $CRIT_p$ is drawn and after its analysis which depends on the nature of the criterion function, a valid selection of the parameter p is obtained as well as a valid subset of p predictors which should be used for the multivariate linear regression model.

The choice of the criteria function used for predictor selection is extremely important. The literature usually gives the choice of three different criteria that can be used in such situations and which are presented in section *Linear regression quality metrices*. The first of them, R^2, is given in eq. (10). It is clear, from the nature of this parameter, that if the value of R_p^2 is plotted as the function of the number of predictors, p, the graph will be a monotonically increasing function and that the maximum of that function will be for $p = N$. However, it is also expected that at some point, with a further increase in the parameter p, the increase of the criteria R_p^2 becomes negligible. This feature is crucial for choosing the appropriate value of parameter p, but bearing in mind that the choice of this value is highly subjective because a quantitative measure of the moment when the increment of criteria becomes negligible is not unequivocally defined in the literature. While applying this method on the data available, the graph in fig. 5 was obtained. In order to make the selection of the parameter p as objective as possible, fig. 6 shows the increment of the criteria R_p^2 as a function of p.

Figure 6 suggests two possibilities for the choice of parameter p. If the main objective is to minimize the number of predictors as much as possible, a good choice would be $p = 9$, because for each $p > 9$, the increment of criteria is significantly smaller. On the other hand, looking at the tail of this curve, starting from $p = 13$ the increment of the criterion be-

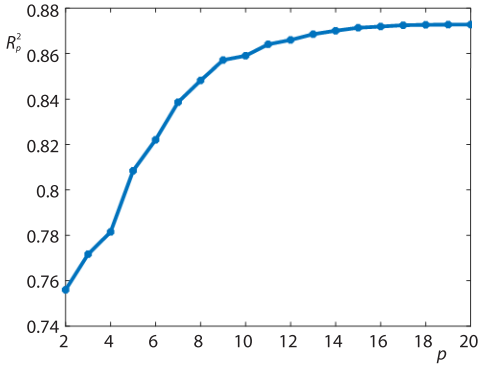

Figure 5. The R_p^2 for *All possible subsets* method

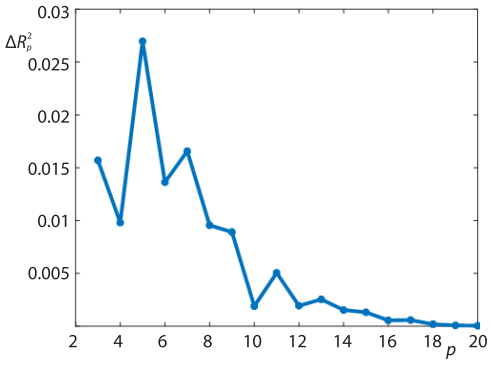

Figure 6. Increments of R_p^2

comes less than $2 \cdot 10^{-3}$. Therefore, in the next section of this paper it will be interesting to look specifically at how well the linear regression obtained in this way really corresponds to the set of acquired measurements.

There are two questions that need answering, at this point. The first question is which predictors are chosen for fixed p? It is shown that for $p = 9$, the best choice of predictors is: active electrical power (x_1), total fuel flow (x_8), steam flow in front of high pressure bypass (x_{16}), flue gas temperature in front of Luv 1 (x_9), flue gas temperature in front of superheater 3 right (x_{13}), flue gas temperature in front of superheater 3 left (x_{14}), steam pressure in the high pressure collector (x_{17}), steam temperature in the high pressure collector (x_{18}) and pressure of the feed water behind the feedhead (x_{20}). One has the impression that the choice of variables is good, that is, very logical and close to the choice that an expert would make from the point of view of the influence of some physical quantities on others.

The second, very important question is whether, by increasing the parameter p, the set of predictors selected for $p - 1$ only expands by one more variable, because in this way the number of possibilities for searching for the best subset of class p could be significantly reduced. Unfortunately, the experimental results show that it can easily be the case that the difference between the best subset of class $p - 1$ and class p is significantly greater than 1. In the case of these experimental results, when jumping from $p = 12$ to $p = 13$ the difference between the predictors amounted to as many as five variables.

Apart from determination coefficient, the other metric which is often used in the *All possible subsets* approach is the variance of the estimator, s_p^2, as defined in eq. (11). Unlike the previous criterion, the variance of the estimator, should be as small as possible. Theoretically, with increasing parameter p, the variance of s_p^2 should decrease and the minimum of this criterion should be reached for $p = \mathrm{N}$. However, it often happens in practice that the minimum of this criterion occurs for $p < N$. The rationale for such a result is very simple. Namely, if you look at the expression for s_p^2, you can see that it is defined through a quotient in which the numerator is the error sum of squares (SSE) and the denominator is the degree of freedom ($n - p$). Both of these expressions decrease with increasing parameter p, so it may happen that by including some new variables, the numerator decreases less than the denominator, so the total expression for the estimated variable increases. This effect is clearly a consequence of the specificity and structure of the set of measurements, rather than the nature of those measurements themselves. That is why some authors suggest that the optimal p^* be chosen either the p for which the minimum criterion is obtained or the smallest p that satisfies the condition $s_p^2 < s_N^2$, *i.e.*:

$$p^* = \min_p \left(s_p^2 < s_N^2 \right) \tag{18}$$

For the experimental measurements obtained in this research the diagram of variance of the estimator as a function of number of predictors, p, is shown in fig. 7. Based on this diagram, and following the recommendations mentioned in the previous paragraph, a good choice of parameter p could be either $p = 7$ because it is the smallest value of p for which the criterion s_p^2 is smaller than s_{20}^2, or $p = 11$ because it provides the minimum of criterion s_p^2.

Regardless of the fact that different optimal values were obtained for the parameter p depending on whether the coefficient of multiple determination R_p^2 or the variance of the estimator s_p^2 is used as a criterion, it is good to note that the choices of predictors themselves are almost identical. Namely, the set of optimal predictors for $p = 7$ with the criterion s_p^2 is a subset of the set of predictors for $p = 9$ and the criterion R_p^2. The only difference is in the flue gas temperature in front of superheater 3 on the right and the steam temperature in the high pressure collector.

The third metric is the C_p criterion as defined in eqs. (12) and (13). The literature suggests that one should look for such a p for which C_p is small and at the same time close to the line $C_p = p$. More precisely, one should look for a minimum p for which $C_p < p$.

Figure 8 shows results obtained regarding the criterion C_p. This diagram is interesting because it really illustrates the possibility of a trade-off between bias and the variance of the $\hat{y}_i$ estimate. If the intention is to keep the variance as small as possible, one should look for the point that is closest to the line $C_p = p$. This is the line drawn in the diagram with a dashed line. On the other hand, in order for the bias to be as small as possible, it is necessary to choose the smallest possible p. A compromise is achieved by choosing the smallest value of the parameter p for which $C_p < p$. In this case it is $p = 7$. What is also interesting is that the selected set of predictors is identical to the set of predictors obtained when applying the s_p^2 criteria.

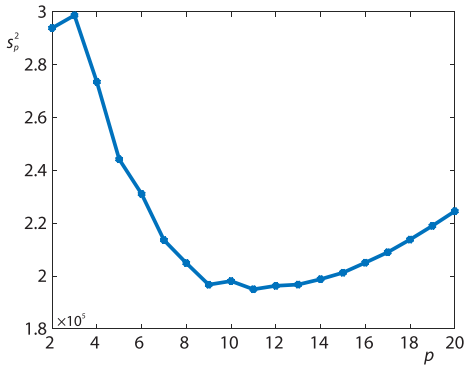

Figure 7. The s_p^2 for *All possible subsets* method

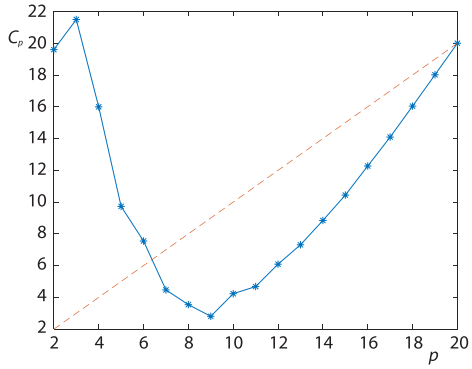

Figure 8. The C_p for *All possible subsets* method

Stepwise selection

In the literature, the problem of feature (or predictor) selection is often discussed because it is a very interesting issue. It is an optimization problem that is theoretically easily solvable because the number of possible solutions is finite. However, that number of finite solutions is very often so large that the obtained theoretical solutions are unusable. Hence, several approaches are suggested to find good enough suboptimal solutions that are numerically feasible. The *Stepwise selection* method is a good example of such an approach. It can be implemented in two forms. One form is forward selection, and the other is backward elimination. In order to

implement the forward selection procedure, firstly the set of predictors is selected for which the maximum value of the parameter F is obtained:

$$F = \frac{MSR}{MSE} = \frac{\dfrac{SSR}{q}}{\dfrac{SSE}{n-q-1}} \tag{19}$$

where MSR and MSE are the mean squares of regression and error, respectively, SSR and SSE are defined earlier and q is the number of predictors which participate in the model and in the first step we adopt $q = 1$. Also n is the number of measurements (or experiments) and in this case $n = 59$. In this way the first variable which will enter the predictor set is chosen. If this variable is denoted as x_1, the next variable which will be included in the predictor set is the variable x_j which maximizes the parameter F defined:

$$F = \frac{\dfrac{SSR\left(x_1, x_j\right) - SSR\left(x_1\right)}{SSE\left(x_1, x_j\right)}}{n-q-1} \tag{20}$$

Now $SSR(x_1, x_j)$ is the sum of regression squares of the second order model in which predictors are x_1 and x_j, and $SSR(x_1)$ is the sum of regression squares of the second order model in which the only predictor is x_1. A similar logic applies to the denominator, that is, the sum of squared errors $SSE(x_1, x_j)$. Since there are now two predictors, we adopt $q = 2$. If in this step we choose variable x_2 as a second predictor, in the next step we choose the variable x_k which will maximize the value of the parameter F defined:

$$F = \frac{\dfrac{SSR\left(x_1, x_2, x_k\right) - SSR\left(x_1, x_2\right)}{SSE\left(x_1, x_2, x_k\right)}}{n-q-1} \tag{21}$$

in which $q = 3$, and so forth. This procedure is repeated until the maximum value of parameter F in a given step becomes less than a threshold. Figure 9 shows how the maximum values of parameter F changes with each step and which variables are included in the set of predictors. Thus, from this diagram it can be seen that the first variable included in the set of predictors is the variable x_5 – the steam temperature before the superheater, then the variable x_8 – the amount of coal, and the last is the variable x_2 – the temperature of the feed water after the economizer.

It is clear from fig. 9 that the maximum values of F are not monotonically decreasing as the number of predictors increases, as one might expect. Also, this diagram does not offer an obvious threshold for the maximum value of F beyond which it would not make sense to add new predictors. Just by analyzing the graph there is a significant drop after the third step in which the variable x_1 was added, which would suggest that the procedure should end with just three predictors. Intuitively it is clear that this would result in a bad model. However, by careful analysis of the obtained numerical values, it seems that a good choice for the threshold is $F_{tr} = 1$, which would result in the formation of a model with 11 predictors, where the last added variable would be x_{10} – the flue gas temperature before Luv 2. The set of predictors:

$$S = \left\{x_1, x_5, x_8, x_9, x_{10}, x_{13}, x_{14}, x_{15}, x_{17}, x_{18}, x_{20}\right\} \tag{22}$$

Another way to implement the *Stepwise selection* procedure is backward elimination. Backward elimination starts from the assumption that we have included all available input variables in the set of predictors, and then, step by step, we eliminate one variable at a time.

The variable which yields the smallest values of the parameter F is eliminated in each step. So, in the first step, we eliminate the variable for which the minimum of the following expression is obtained:

$$F = \frac{SSR\left(x_1, x_2, \ldots, x_N\right) - SSR\left(x_1, \ldots, x_{j-1}, x_{j+1}, \ldots, x_N\right)}{\frac{SSE\left(x_1, x_2, \ldots, x_N\right)}{n - N - 1}} \tag{23}$$

Here, the notation is the same as in the forward procedure. If we mark the variable thus obtained with x_N, in the next step we remove the variable x_k which minimizes the expression:

$$F = \frac{SSR\left(x_1, x_2, \ldots, x_{N-1}\right) - SSR\left(x_1, \ldots, x_{k-1}, x_{k+1}, \ldots, x_{N-1}\right)}{\frac{SSE\left(x_1, x_2, \ldots, x_{N-1}\right)}{n - \left(N-1\right) - 1}} \tag{24}$$

and so on. The procedure is repeated until the minimum obtained F value becomes so large that the further elimination of variables will lead to a significant loss of information.

Implementation of the backward elimination procedure results in the numerical values shown in fig. 10, where it is shown how the minimum value of the parameter F changed in each step, as well as the series of input variables which were eliminated from the predictor set in each step.

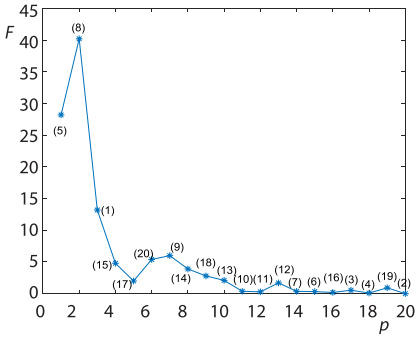

Figure 9. Maximum values of *F* for *Stepwise selection* method, forward selection procedure

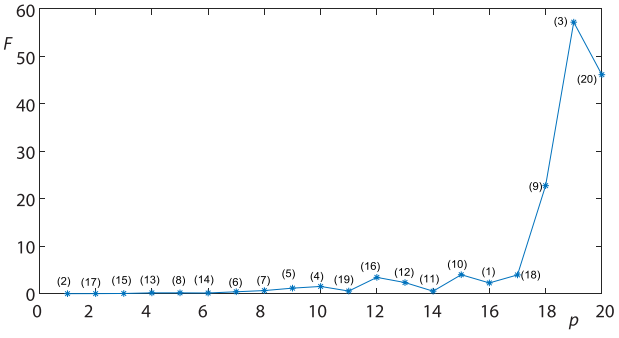

Figure 10. Minimum values of *F* for *Stepwise selection* method, backward elimination procedure

The interesting thing which can be seen from figs. 9 and 10 is that the variables were not eliminated in the reverse order in which they were added to the set of predictors. Only the variable x_2 – the temperature of the feed water behind the economizer was the last added in the forward selection procedure and the first eliminated in the backward elimination procedure. For example, the variables x_{15} and x_{17}, fuel oil consumption and water vapor pressure in the high pressure collector, respectively, were very quickly eliminated from the set of predictors as non-informative variables, but in the forward procedure they were introduced at the very beginning, as very informative. This disagreement indicates that this procedure is not optimal by any criteria, but at the same time it can illustrate the high complexity of the problem itself. Namely, a large number of variables among the initial 20 predictors are correlated among themselves, so it is not easy to recognize which variables among them are essential generators of change in the output variable.

Similarly as in the forward selection, based on fig. 10 it is not easy to recognize the threshold F_{tr} after which it becomes dangerous to eliminate the predictors because the loss of information will become significant. If we adopted the same threshold again, $F_{tr} = 1$, as we adopted in the forward procedure, the algorithm would form a set of predictors with 12 variables:

$$S = \{x_1, x_3, x_4, x_5, x_9, x_{10}, x_{11}, x_{12}, x_{16}, x_{18}, x_{19}, x_{20}\} \tag{25}$$

It is interesting to note that the intersection of the set of predictors generated by the forward and backward procedures consists of only five variables, which is not a lot considering that one set of predictors has 11 and the other 12 signals. It will be interesting to analyze the extent to which the regression model obtained on all these sets of predictors fits, that is, corresponds to the set of measurements on real experiments.

Results

In the previous section, a serious analysis was conducted about the choice of predictors. It is shown that different procedures and different metrics within those procedures lead to different choice of predictors. If we summarize the analysis in the previous section, it would read like this. In the *All possible subsets* procedure, if we use R_p^2 as a measure, and if we choose $p = 9$ as a good choice of the number of input variables, the corresponding set of predictors becomes:

$$S_1 = \{x_1, x_8, x_9, x_{13}, x_{14}, x_{16}, x_{17}, x_{18}, x_{20}\} \tag{26}$$

If we choose $p = 13$ as the number of input variables, the set of predictors would be

$$S_2 = \{x_1, x_3, x_4, x_5, x_8, x_9, x_{11}, x_{12}, x_{13}, x_{16}, x_{18}, x_{19}, x_{20}\} \tag{27}$$

When applying the same *All possible subsets* methodology, but if we choose the variance s_p^2 as a measure of quality, we also got two possible sets of predictors, one for $p = 7$, and the second for $p = 11$. The resulting sets are marked with S_3 and S_4, respectively:

$$S_3 = \{x_1, x_8, x_9, x_{14}, x_{16}, x_{17}, x_{20}\} \tag{28}$$

$$S_4 = \{x_1, x_3, x_8, x_9, x_{10}, x_{11}, x_{12}, x_{14}, x_{16}, x_{18}, x_{20}\} \tag{29}$$

When applying the coefficient C_p as a metric, the same set of predictors as shown in S_3 was obtained:

$$S_5 = S_3 = \{x_1, x_8, x_9, x_{14}, x_{16}, x_{17}, x_{20}\} \tag{30}$$

In the *Stepwise selection* method using the forward procedure, proved to be a good choice of predictors:

$$S_6 = \{x_1, x_5, x_8, x_9, x_{10}, x_{13}, x_{14}, x_{15}, x_{17}, x_{18}, x_{20}\} \tag{31}$$

and using backward elimination the following set is obtained:

$$S_7 = \{x_1, x_3, x_4, x_5, x_9, x_{10}, x_{11}, x_{12}, x_{16}, x_{18}, x_{19}, x_{20}\} \tag{32}$$

In order to understand whether there are variables which are recognized as informative predictors in each approach that is tested, as well as whether there are those that do not seem important in any scenario, the tab. 6 was created.

It turns out that the variables x_1 – active electric power of the block, x_9 – temperature of the flue gas before Luv 1, as well as x_{20} – the pressure of the feed water behind the feed head are recognized as strong predictors in each of the input variable selection approaches. Also, as important predictors, we should mention x_8 – total fuel set point and x_{16} – steam flow in front of high pressure bypass. On the other hand, x_2 – feed water temperature behind economizer,

Table 6. The choice of predictors using different selection procedures

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}	x_{16}	x_{17}	x_{18}	x_{19}	x_{20}
S_1	■							■	■				■	■		■	■	■		■
S_2	■		■	■	■			■	■		■	■	■			■		■	■	■
S_3	■							■	■					■		■	■			■
S_4	■		■					■	■	■	■	■		■		■		■		■
S_5	■							■	■					■		■	■			■
S_6	■				■			■	■	■			■	■	■		■	■		■
S_7	■		■	■	■				■	■	■	■				■		■	■	■

x_6 – total air or x_7 – air-flow for roasting, were not recognized as informative predictors in any of the approaches used.

It is also interesting to see how well the outputs of the obtained linear models match with the experimentally obtained heating values of coal. These results are shown in the following graphs. Figure 11(a) shows the output of the linear model by adopting the predictor set S_1, in fig. 11(b) the predictor set S_2 is adopted and so on. Results for the set of predictors S_5 are excluded because they are identical to S_3. The figures also show the output of the model that is used in the TPP and which was created based on physical laws. In the roughest terms, this model estimates the heating value of coal as a quotient of the total energy obtained through the total amount of coal delivered to the boiler in a unit of time. At the same time, the total amount of energy is calculated through the flow of water vapor towards the turbine, its enthalpy, taking into account the energy loss through injection, the heat of the flue gases, the loss due to the non-sealing of the boiler and so on. In each of the figures shown, the caption shows the mean squared error value obtained through the output of our linear model (SSE) and physical model (SSEF).

Based on the obtained and presented results, two conclusions are indisputable. One is that, based on root MSE, all linear regression models perform better than the physical model. Another important conclusion is that, again based on the MSE, sets of predictors S_4 and S_7 stand out as the best. The S_4 was obtained by applying the *All possible subsets* technique and variance s_p^2 criteria, adopting a solution with 11 predictors, while the solution S_1 was obtained by applying the *Stepwise selection* technique in the form of backward elimination and choosing a set of 12 predictors. This interpretation on the best set of predictors, however, should be taken with a grain of salt. One reason for doubt is the dilemma: is the MSE of estimation really the *right* criterion for ranking? The other reason is that the chosen solutions S_4 and S_7 have 11 and 12 predictors, respectively, which is more or even significantly more than the sets of predictors S_3 and S_5 which have 7, S_1 with 9 or S_6 with 10 predictors. In other words, there are unfair elements in the omparison system itself, which leave room for certain doubts.

In order to verify and compare the obtained sets of predictors and the corresponding linear regression models, it is possible to perform two basic hypothesis tests mentioned in section *Hypothesis testing*. The first hypothesis test is the so-called overall regression test, which tests the hypothesis H_0 that none of the selected predictors predicts the output of the model. This hypothesis can be tested by introducing a random variable F from eq. (15).

Another hypothesis test that can be used in this context is a test on a subset of regression parameters and the hypothesis H_0 in this case says that the predictors from the subset of predictors should have been omitted because their influence on the output variable is non-existent. This hypothesis is tested using random variables F from eq. (17).

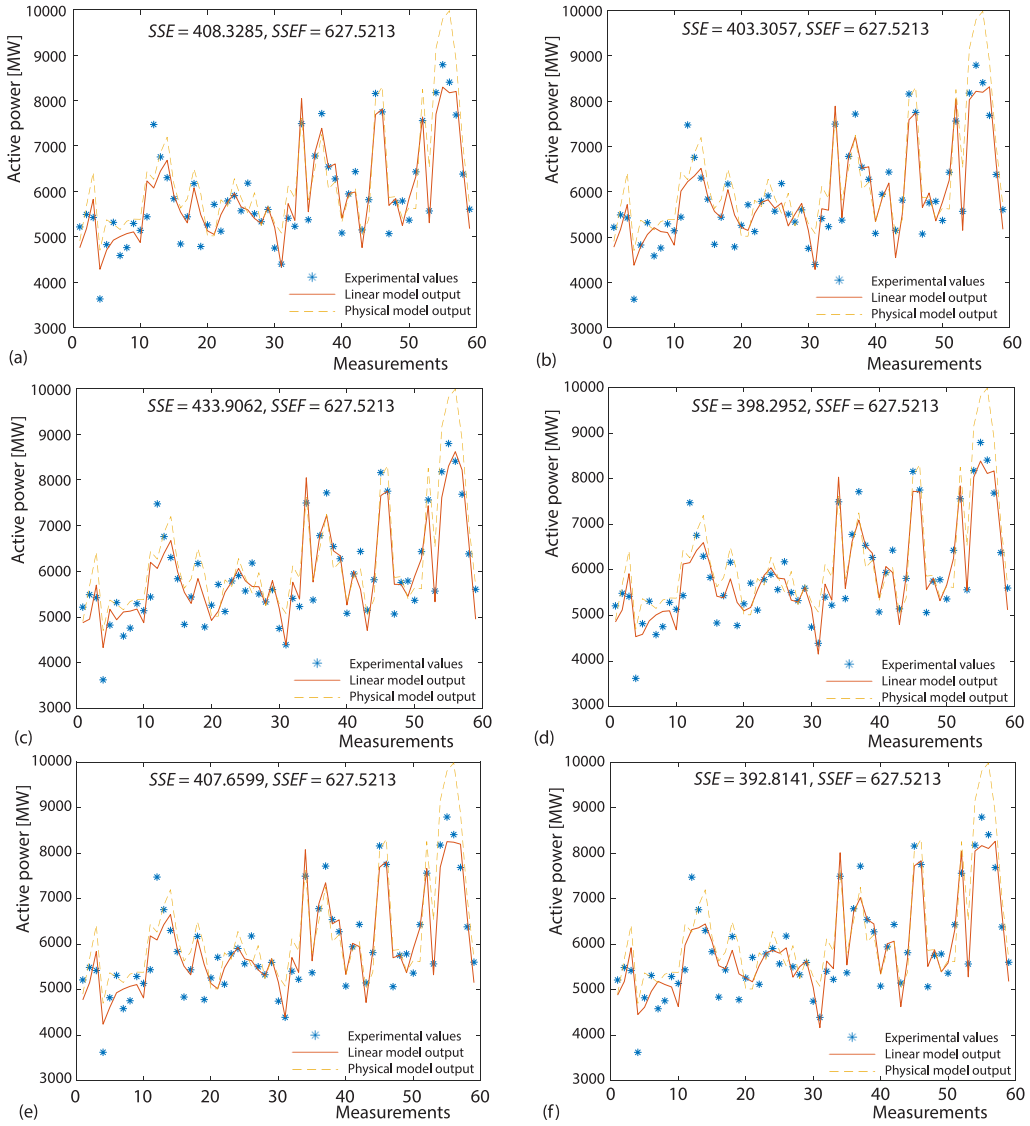

Figure 11. Performances of predictor sets; (a) S_1, (b) S_2, (c) S_3, (d) S_4, (e) S_6, and (f) S_7

These two hypothesis tests were conducted for all previously selected combinations of predictors and the results are shown in the tab. 7.

Table 7 shows the results of the hypothesis testing of the first and second tests. It is shown that each of the analyzed sets of predictors from S_1 to S_7 refuse the hypothesis H_0. Recall that the hypothesis H_0 assumes that the selected set of predictors does not model the output F_α, where α is the confidence level.

For the first test, the confidence level was chosen to belong to the set $\alpha \in \{0.001, 0.01, 0.05\}$. The row which corresponds to the set S_5 is omitted because this set is identical to the set of predictors in S_3, so the results are also identical. It can be seen that for each of the selected sets of predictors S_i, $i = 1, 2,...,7$, the hypothesis H_0 is rejected, and very con-

Table 7. Hypothesis testing results

	Test 1				Test 2			
	F	$F_{0.05}$	$F_{0.01}$	$F_{0.001}$	F	$F_{0.10}$	$F_{0.05}$	$F_{0.01}$
S_1	32.66	2.07	2.79	3.83	0.55	1.705	1.99	2.63
S_2	21.37	1.94	2.55	3.44	0.62	1.855	2.22	3.065
S_{23}	37.87	2.19	3.01	4.20	1.05	1.66	1.915	2.50
S_4	27.16	2.00	2.65	3.60	0.36	1.77	2.085	2.81
S_5								
S_6	28.90	2.03	2.71	3.71	0.57	1.735	2.035	2.715
S_7	25.16	1.97	2.60	3.51	0.23	1.81	2.145	2.925

vincingly. If we still need to somehow rank the persuasiveness of defeating this hypothesis, we should single out the set S_3 (or S_5) because the obtained value of the variable F is convincingly the highest compared to the threshold $F_{0.05}$.

The right half of tab. 7 shows the results obtained for the second hypothesis test. The hypothesis H_0 which is tested states that the omitted predictors in relation the initial, twenty-dimensional set of predictors do not carry any information about the output variable, that is, that their coefficients in the linear regression model should be equal to zero. The obtained results show that the value of the variable F in each of the mentioned sets of predictors is convincingly lower than the hypothesis rejection thresholds, and the hypothesis H_0 must be accepted. This is a desirable result because it indicates that each of these analyzed choices is meaningful and that the feature selection procedures, described in the previous section, are statistically justified. For the purpose of this hypothesis testing, $\alpha \in \{0.01, 0.05, 0.10\}$ were used as confidence levels. Again, regardless of the fact that all sets accepted the hypothesis H_0, if we were to rank the intensity of acceptance, we would single out the predictor set S_7 because in its case the value of the variable F is convincingly the smallest compared to the threshold $F_{0.10}$.

Conclusions

This paper detailed the development of a linear regression model to estimate the heating value of coal delivered to the mills of a TPP and emphasized the importance of this estimation for adjusting the regulatory structures within the boiler plant. The methodology was demonstrated using data from the TPP *Nikola Tesla* Unit B1 – TENT B1 in Obrenovac, Republic of Serbia.

Based on the definition of numerous potential predictors available in the DCS data-logging system, various techniques for selecting predictors and testing hypotheses about the acceptance or rejection of models were illustrated. The results show that even with only seven predictors, linear regression models provide an excellent estimate. Furthermore, with 12 predictors, the linear regression model outperforms the heating value estimator based on a physical model.

On the other hand, the analysis here also raised some interesting questions that were not answered in this paper. First, is the model obtained generic or does each thermal power plant have its own set of optimal predictors? Another important question is the possibility of implementing in the model some prior knowledge about the physical influence of the heating value of coal on certain physical quantities. The third question is whether the fact that there was a relatively small number of experimental measurements and that these were not divided into training and test groups affects the validity of the results obtained. Finally, when analyzing

the data obtained, it was found that there are measured values that deviate unusually from the rest of the population and are obviously outliers that occurred in some of the extreme situations (disappearance of the coal on the feeder, clogging of the hatch for the coal supply, filling of the mill*, etc.*). The question that arises is whether using one of the robust methods to pre-process the measurements would provide better results? These issues will be the subject of further research by the authors.

Acknowledgment

This work was financially supported by the Ministry of Science, Technological Development and Innovation of the Republic of Serbia under contract number: 451-03-65/2024-03/200103 and Science Fund of the Republic of Serbia, 1375/2023, NOx Reduction Based Thermal Power Plant Optimization – NOXTROT.

References

[1] Manisalidis, I., *et al.*, Environmental and Health Impacts of Air Pollution: A Review, Frontiers in Public *Health*, *8* (2020), 505570
[2] Kuyper, J., *et al*., The Evolution of the UNFCCC, *Annual Review of Environment and Resources*, *43* (2018), Oct., pp. 343-368
[3] ***, EURACOAL, European Association for Coal and Lignite, Coal Industry across the Europe 8th edition, 2024, ISSN 2034-5682
[4] Henderson, C., Upgrading and Efficiency Improvement in Coal-Fired Power Plants, CRC/221 IEA Clean Coal Centre, 2013, https://usea.org/sites/default/files/082013_Upgrading%20and%20efficiency%20improvement%20in%20coal-fired%20power%20plants_ccc221.pdf
[5] Wynn, G., Coghe, P., *Europe's Coal-Fired Power Plant: Rough Times Ahead*, Institute for Energy Economics and Financial Analysis, Detroit, O., USA, 2017
[6] Đoković, N., *et al*., Petrographical and Organic Geochemical Study of the Lignite from the Smederevsko Pomoravlje Field (Kostolac Basin, Serbia), *International Journal of Coal Geology*, *195* (2018), July, pp. 139-171
[7] Stupar, G., *et al*., Assessing the Impact of Primary Measures for NO_x reduction on the Thermal Power Plant Steam Boiler, *Applied Thermal Engineering*, *78* (2015), Mar., pp. 397-409
[8] Lu, S., Hogg B. W., Dynamic Nonlinear Modelling of Power Plant by Physical Principles and Neural Networks, *International Journal of Electrical Power and Energy Systems*, *22* (2000), 1, pp. 67-78
[9] Mesroghli, S., *et al*., Estimation of Gross Calorific Value Based on Coal Analysis Using Regression and Artificial Neural Networks, *International Journal of Coal Geology*, *79* (2009), 1-2, pp. 49-54
[10] Mladenović, M. R., *et al*., Criteria Selection for the Assessment of Serbian Lignites Tendency to form Deposits on Power Boilers Heat Transfer Surfaces, *Thermal Science*, *13* (2009), 4, pp. 61-78
[11] Hasler, D., Coal-Fired Power Plant Heat Rate Reduction, Sargent and Lundy, Report No. 13276-001, Chicago, Ill., USA, 2009
[12] Williams, A., *Combustion and gasification of coal*. Routledge, Abingdon, UK, 2018
[13] Wang, L., *et al.*, A Method for in-Situ Measurement of Calorific Value of Coal: A Numerical Study, *Thermochimica Acta, 703* (2021), 179011
[14] Rencher, A. C., Christensen, W. F., *Methods of Multivariate Analysis*, John Wiley and Sons Inc., New York, USA, 2012
[15] Hayes, A. F., *et al.*, Cautions Regarding the Interpretation of Regression Coefficients and Hypothesis Tests in Linear Models With Interactions, *Communication Methods and Measures*, *6* (2012), 1, pp. 1-11
[16] Zhou, J., Zhang, W., Coal Consumption Prediction in Thermal Power Units: A Feature Construction and Selection Method, *Energy*, *273* (2023), 126996
[17] Liu, Y., *et al*., Feature Selection Method Reducing Correlations among Features by Embedding Domain Knowledge, *Acta Materialia*, *238* (2022), 118195
[18] Chandrashekar, G., Sahin, F., A Survey On Feature Selection Methods, *Computers and Electrical Engineering*, *40* (2014), 1, pp. 16-28

Paper submitted: January 24, 2024
Paper revised: May 2, 2024
Paper accepted: May 15, 2024