



A comparison between multivariate linear model and maximum likelihood estimation for the prediction of elemental composition of coal using proximate analysis

Fuguo Liu

State Grid Shandong Electric Power Research Institute, 2000 Wangyue Road, Jinan, 250003, PR China



ARTICLE INFO

Keywords:

Coal
Proximate analysis
Ultimate analysis
Multivariate linear model
Maximum likelihood estimation

ABSTRACT

The elemental composition of coal is essential for analysing the overall process of energy conversion systems. Simultaneous information regarding the elemental composition of coal fed into a boiler is of increasing interest for plant operation. In this study, methods of estimating the elemental composition of coal using proximate analysis and a higher heating value were developed to meet requirements in boiler operation. A novel method was developed to formulate the multivariate linear model (MLM) for predicting the elemental composition of coal by solving a set of simultaneous equations with extensive correlations between coal components and an elemental composition constraint. The maximum likelihood estimation (MLE) approach for predicting carbon, hydrogen, oxygen, nitrogen, and sulphur contents was also developed based on a series of extensive correlations among coal components. The maximum likelihood estimator employs more correlations and considers the performance of prediction residuals for each correlation, offering a better prediction accuracy than the MLM, which is based on fewer correlations and neglects prediction residuals. A total of 743 data points was used to derive the MLM and MLE models, which were validated and verified by another set of data that included the same variety of coal types. The proposed methods can estimate the complete elemental composition of coal (C , H , O , N , and S) with acceptable prediction accuracy for engineering purposes. Another important finding is that average absolute error corresponding to the measured values of nitrogen is only 14.14%, although the predicted nitrogen contents did not follow the trends of the measured values.

1. Introduction

Equipment and burners used in coal handling and firing systems in power station boilers are designed to incorporate fuel properties such as calorific value, volatile matter, and ash fusion temperature. Notable variations in the properties of coal fed into boilers can have significant impacts on the performance and safety of a power plant [1–3]. Firing off-design or low-grade coal may result in furnace slagging, an increase in pollutant emissions, and a reduction in boiler efficiency [4–6]. Thus, it is vital for power plant operators and process control systems to be aware of the coal properties and to make adjustments to abate negative impacts due to variations.

Chemical composition, normally defined in terms of proximate analysis and ultimate analysis, is a critical property of coal. Proximate analysis only presents the mass percentages of volatile matter (VM), fixed carbon (FC), moisture (M), and ash (ASH). Ultimate analysis provides detailed elemental composition information (i.e., C , H , O , N , and S

content). Proximate analysis, which requires only common equipment and is much easier to conduct, can be used to provide a general evaluation of the quality of coal [7]. Thus, proximate analysis data, along with total sulphur content and higher heating value (HHV), are frequently available in power plant databases. Unfortunately, there is a lack of data on ultimate analysis due to the sophisticated equipment and highly skilled analysts required for the experimental measurement. However, the elemental composition of coal helps determine the stoichiometric air requirement, composition, and flow rate of the flue gas, which are critical for maintaining an appropriate proportion of air and coal in the combustion chamber, reducing heat loss as a result of excess air supply, and minimising the power consumed to convey air and flue gas throughout the furnace. Monitoring the elemental composition of coal in real time is a growing concern in power plant operation. However, online analysers such as instrumentation based on prompt gamma neutron activation analysis (PGNAA) are expensive and require a special installation location.

E-mail address: lephico@163.com.

One approach is to estimate coal properties that are unattainable due to limitations in existing instrumentation as a function of other easily obtainable properties. For nearly a century, attempts have been made to correlate a wide variety of coal properties with proximate or ultimate analyses, including higher and lower heating values [8–11], grindability [12,13], vitrinite maximum reflectance, and maceral composition [14, 15]. Several correlations for estimating elemental composition using proximate analysis have been published; however, these have focused on biomass fuels [16–19] and black liquor [20]. There has been limited published research pertaining to elemental coal prediction. Yao et al. [21] developed an artificial neural network-based model to predict the hydrogen content of coal from proximate analyses. Saptoro [22] formulated several empirical models to demonstrate the potential use of artificial neural networks (ANNs) and indicated that ANNs produced the best result among the tested methods and were a promising tool for predicting the elemental composition of coal. Yi et al. [23] proposed a multiple quadratic regression model, neglecting interaction terms among independent variables to predict the elemental composition of coal using proximate analysis. The range of data points considered for VM, FC, ASH, C, H, and O contents were 1.87–49.7%, 35.64–91.6%, 1.5–27.19%, 49.70–90.80%, 0.95–5.40%, and 1.02–24.40%, respectively. The data range was divided into four separate regions for anthracite, high-rank bituminous, subbituminous, and lignite. Piecewise functions were selected to establish the relationship between proximate and ultimate analyses. As a result, four sets of correlations were derived for predicting carbon, hydrogen, and oxygen composition. Yi et al. also reported that the estimation error for oxygen content increases greatly if one unified set of correlations is developed for the prediction of elemental composition using the entire reported data range. It is difficult to derive general correlations to accurately predict the elemental composition of coal, especially over a wide range of coal types.

There are also challenges in predicting comprehensive components for ultimate analysis. Predicting all components using correlations derived directly from the regression analysis model is problematic in terms of the total elemental composition due to an unsatisfied compositional constraint (total content equals 100%). Until now, correlations for carbon, hydrogen, and oxygen composition, rather than for all components, have been published irrespective of an indication of biomass or coal [16–19,22,23]. This ultimate analysis, neglecting nitrogen and sulphur, is incomplete and places restrictions on the practical application of these correlations.

The objective of this study is to develop generalised empirical models to predict comprehensive compositions for the ultimate analysis for a wide range of coal. The study is divided into: (i) developing a multivariate linear model (MLM) with deliberately selected inputs using a large number of coal samples, (ii) presenting a novel method to formulate a MLM by solving a set of simultaneous equations composed of both component correlations and an elemental composition constraint, (iii) predicting the elemental composition of coal based on maximum likelihood estimation (MLE), and (iv) comparing the results of the MLM with the maximum likelihood estimator due to its frequently used in existing literature [24,25], and finally comparing both results with published correlations.

2. Collected and experimental data

The data points considered in this study for developing and validating the proposed model were obtained either from literature [26–28] or by experimental determination. To enable extensive and universal applicability of the derived models, the dataset covers a wide variety of coal types, including anthracite, bituminous, and lignite, with 855 data points. 743 data points were applied to formulate the model. Another 112 data points were used for the validation and verification of the derived model. The range of reported data for VM, FC, HHV, C, H, O, N, and S are 1.71–61.80%, 38.20–98.29%, 25.726–37.177 MJ/kg, 63.54–97.22%, 0.54–6.85%, 0.06–26.02%, 0.03–9.25%, and

0.02–12.74%, respectively. The reported data are in dry and ash-free wt. %.

In the experimental determination, coal samples were collected from a large number of fossil fuel power stations in China for proximate analysis, ultimate analysis, and determination of calorific values in a laboratory. The sampling and preparation of coal were conducted in accordance with ISO 18283-2006 (E). The proximate contents of moisture, volatile matter, and ash were determined according to ISO 589:2008, ISO 562:2010, and ISO 1171:2010, respectively. The fixed carbon content of the test samples was calculated using the difference. The gross calorific value was determined using the bomb calorimetric method, and the calculation of net calorific value was in compliance with ISO 1928:2009. The ultimate analysis of coal samples was conducted according to ISO 17247:2013.

3. Methodology

In few previous studies, multiple regression models were usually developed to predict the C, H, and O content in coal or biomass fuel as a linear or quadratic function of FC and VM, and ash content (ASH). The ultimate and proximate components of these correlations are normally on a wt.% dry basis, which requires determination of ash content and suggests that the dataset used to derive the model contains errors in ash measurement. The ash and moisture content of coal are strongly affected by external conditions during the mining process, transportation, and storage, which can result in additional errors in the estimation of elemental composition. This is verified by Nhuchhen [16], who found that correlations derived using the ash term for biomass fuel have larger estimation errors than correlations without the ash term. Nhuchhen inferred that this must be due to additional errors associated with ash content, which is further supported by reported results that the measured value of ash content in biomass is significantly affected by the ashing temperature maintained during the experimental measurement [18,29].

In this study, to eliminate the impact of ash and moisture measurements on the accuracy of estimating elemental composition, ultimate and proximate analysis data were selected based on dry and ash-free (daf) wt.%. The daf basis is the most stable means of representing the proportions of C, H, O, N, and S, which helps improve model precision. Additionally, HHV has been used as an input for predicting coal elemental composition in conjunction with proximate analysis data because HHV has a close relationship with elemental composition [8–12] and can be easily obtained from power plant databases. Thus, the model for predicting the elemental composition of coal is presented as

$$Y = f(FC, VM, HHV) \quad (1)$$

where Y represents the elemental content of C, H, O, N, or S. FC, VM, and HHV represent fixed carbon, volatile matter, and higher heating value, respectively, expressed in mass percentages or MJ/kg on a dry and ash-free basis.

In this study, an MLM and an MLE approach are applied to formulate the model, as shown in Eq. (1).

3.1. Extensive correlation between compositional components of coal

Correlations for elemental composition from the regression model, although extremely useful for estimating the content of a single elemental component, cannot be used to compute all elemental components because they are not constantly in compliance with the compositional constraint for ultimate analysis, which can be expressed as

$$C + H + O + N + S = 100 \quad (2)$$

Previous studies have focused on determining correlations for elemental compositions as functions of FC and VM; in these cases, care

was seldom taken to establish interrelationships between the elemental components of coal, except in those given by Neavel [30]. An interrelationship derived between elemental components specifies the constraint of elemental composition and can be used as a member of the set of simultaneous equations to determine the elemental composition. Based on this consideration, any correlation between ultimate analysis, proximate analysis, and HHV may be helpful in obtaining empirical expressions of the elemental composition of coal. Thus, extensive correlation was investigated using the presented dataset, which includes the correlation of elemental composition with proximate analysis, correlation between elemental components, correlation of HHV with ultimate analysis, and correlation among different algebraic combinations of components.

It is assumed that x and y represent an individual component, HHV, or a combined parameter of components, and are expressed as

$$x = X(U, P, \text{HHV}) \quad (3)$$

$$y = Y(U, P, \text{HHV}) \quad (4)$$

where U and P represent the components of the ultimate and proximate analyses on a dry and ash-free wt.% basis, respectively, denoted as

$$U = (C, H, O, N, S) \quad (5)$$

$$P = (FC, VM) \quad (6)$$

The dependent variable y is predicted from the independent variable x using the polynomial model.

$$y = \hat{y} + \varepsilon \quad (7)$$

$$\hat{y} = \sum_{i=0}^n a_i x^i \quad (8)$$

where \hat{y} is the predicted value of y , ε is the prediction residual, n is the order of the polynomial, and a_i is the model constant that can be estimated based on the principle of least squares.

3.2. Multivariate linear model based on compositional constraint

When the value of ε in Eq. (7) is sufficiently small to be omitted, combining the equation with Eq. (8) yields

$$y = \sum_{i=0}^n a_i x^i \quad (9)$$

A series of correlation forms are found using 743 coal samples; four of them are selected to formulate equations, as shown in Eq. (9), and incorporated with the constraint of elemental composition [Eq. (2)] for closure. This formed a set of simultaneous equations for determining the content of C, H, O, N, and S expressed as a function of P and HHV. They are multivariate linear functions in case the order of the polynomial in Eq. (8) equals 1, i.e., $n = 1$.

3.3. Maximum likelihood estimator method

Prediction of the ultimate analysis of coal using the maximum likelihood approach is based on a series of correlations presented as Eqs. (7) and (8). Given m correlation forms, for correlation j ($j = 1, 2, \dots, m$), after the model constant a_i in Eq. (8) has been evaluated using the presented dataset, the prediction residual value of ε is calculated using Eq. (7) over each data point. All residuals pertaining to the dataset are summarised to extract the probability density distribution using the *hist* function in MATLAB. The probability density function of the prediction residuals for correlation j over the dataset is denoted as $\mathcal{O}_j(z_j)$, where z_j represents the prediction residual. From m correlation forms, probability density functions can be obtained as $\mathcal{O}_1(z_1), \mathcal{O}_2(z_2), \dots, \mathcal{O}_m(z_m)$.

For a targeted coal sample whose elemental composition is to be

estimated, at given observed values of elemental composition U , the probability can be calculated using the joint density as

$$L(z_1, z_2, \dots, z_m | U) = \prod_{j=1}^m \mathcal{O}_j(z_j | U) \quad (10)$$

The MLE approach is used to find an estimate \hat{U} that maximises the joint density L as

$$\hat{U} = \underset{U}{\operatorname{argmax}} L(z_1, z_2, \dots, z_m | U) \quad (11)$$

The estimated \hat{U} should also comply with the constraint presented in Eq. (2).

3.4. Prediction errors

We use four statistical parameters, mean absolute error (MAE), average absolute error (AAE), average biased error (ABE), and the coefficient of determination (R^2) to characterise the accuracy and performance of the predictions. They are defined as:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |P_i - M_i| \quad (12)$$

$$\text{AAE} = \frac{1}{n} \sum_{i=1}^n \frac{|P_i - M_i|}{M_i} \times 100 \quad (13)$$

$$\text{ABE} = \frac{1}{n} \sum_{i=1}^n \frac{P_i - M_i}{M_i} \times 100 \quad (14)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (P_i - M_i)^2}{\sum_{i=1}^n (P_i - \bar{M})^2} \quad (15)$$

where M , P , \bar{M} denote the measured, predicted, and an average of measured elemental compositions of coal. n is the number of sample data points for MLM and MLE modeling.

MAE has the same unit as the quantity of the predictions and shows the actual amount of error. AAE calculates the degree of closeness between the predicted and measured elemental compositions. ABE indicates the degree of overestimation or underestimation of the measured values. R^2 measures the degree of accuracy of the proposed correlations. A lower AAE indicates a higher prediction accuracy. A smaller absolute value of the ABE indicates a smaller bias in the prediction.

4. Results and discussion

4.1. Extensive correlations

A total of 743 data points was used to find correlations between the dependent variable (y) and the independent variable (x). Both variables are compositional components or algebraic operations of the components. Table 1 presents the expressions of x and y in the nine correlation forms. Fig. 1(a-i) show the scatter distribution plots of y with x for each correlation form over the dataset. The fitted polynomial and/or linear function plots [cf. Eq. (9)] with solid or dotted lines are presented in these figures; the calculated values of the coefficient of determination (R^2) for the correlations are listed in Table 1. These correlations have considerably higher R^2 values than existing correlations in literature for elemental composition using proximate analysis [16–18,23]. Statistical information on prediction residuals over the dataset, such as mean value (μ_j) and standard deviation (σ_j), is also included in Table 1 and is vital for establishing a maximum likelihood estimator to predict the elemental analysis of coal.

Correlations 1, 2, 3, and 8 in Table 1 reveal the relationship between different elemental components of coal. The scatter distribution plots for the correlation formed over the derived dataset are presented in Fig. 1

Table 1

Forms of correlations and prediction performance.

Correlations <i>j</i>	X	Y	$\varepsilon \sim N(\mu_j, \sigma_j^2)$		R^2
			μ_j	σ_j	
1	C	O	-0.002	1.564	0.919
2	C	H + O	-0.001	1.646	0.926
3	C	100-C-N	0.004	0.447	0.995
4	VM	H + O/8+(C-FC)/6	0	0.617	0.951
5	VM	$-0.1515C + 6.4040H + 1.2514O + 0.3627S$	0	4.137	0.890
6	VM	FC/C	0	0.029	0.926
7	C	FC-C + N/S	0.001	3.038	0.872
8	N/14	$3.3264 - 0.3964C/12 - 0.0274H - 0.53200/16 + S/32$	0	0.017	0.910
9	HHV	$0.3475C + 1.0062H - 0.0695(O-S)$	-0.002	0.680	0.858

(a–c, h). High R^2 values indicate good correlation and are likely due to the fundamental C–C, C–H, and C–O chemical bonds in coal that dominate the proportion of the elemental composition. Empirical relationships between carbon, hydrogen and oxygen contents of coals, correlations 1 and 2 in Table 1 for instance, can be very useful for combustion control in coal-fired power utilities [7].

Correlations 4, 5, 6, and 7 in Table 1 show the relationship of elemental composition with proximate components; these correlations

are dominant through the fundamental chemical bonds and also the degree of coalification. Fig. 1(d–g) show the scatter distribution plots of y with x in those correlations. The scatter points are more concentrated around the fitted polynomial or linear function plots, with high R^2 values in the range of 0.87–0.92. Empirical relationships between the elemental components and proximate analysis, such as correlations 4, 5 and 6 in Table 1, are of particular interest due to sophisticated equipment and highly skilled analysts required for the experimental elemental components measurement.

Several correlations for predicting HHV from the elemental analysis of different fuels have been reported [8,9]. Mathews [31] concluded that HHV predictions from elemental analysis-based correlation are accurate over a wide range of compositions, whereas prediction of other coal properties is restricted in applicability to only a select rank range. In this study, a correlation for the estimation of HHV from elemental analysis of coal is proposed as correlation 9 in Table 1. This correlation was derived from 175 data points from Chinese coal. The main advantage of the correlation is, based on using only elemental analysis data, a rapid, easy and economical estimation of the HHV value.

4.2. Multivariate linear model for predicting complete elemental composition

Correlations 2, 3, 4, and 9 in Table 1 are selected to yield the type of

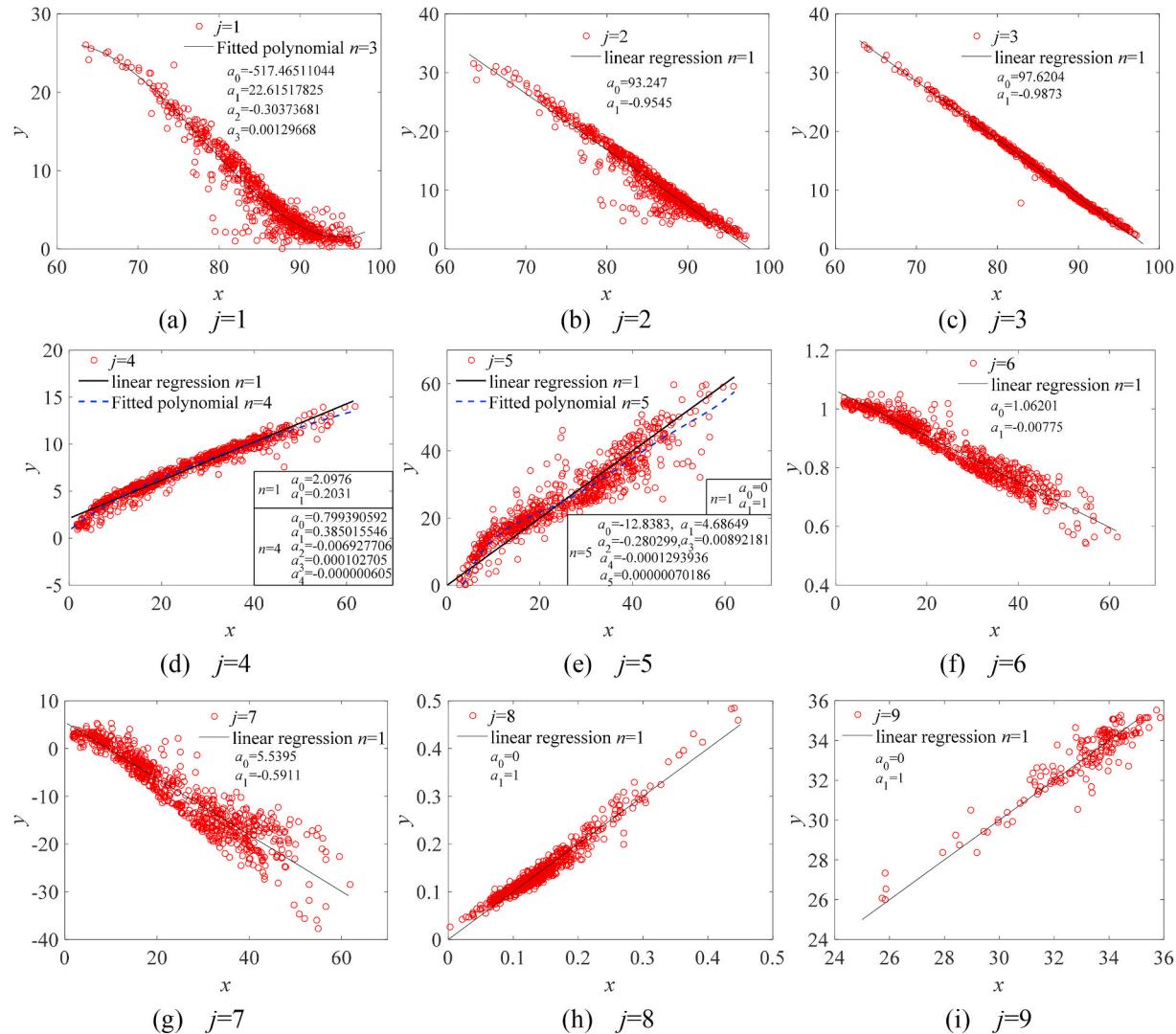


Fig. 1. Variation of y with x and regression functions.

equation shown in Eq. (9). To conveniently solve equations, linear functions with coefficients presented in Fig. 1(b–d, i) are adopted in Eq. (9) rather than polynomial functions. The yielded equations and the compositional constraint of Eq. (2) form a set of simultaneous equations:

$$\begin{cases} H + O = 93.2473 - 0.9545C \\ 100 - C - N = 97.6204 - 0.9873C \\ H + \frac{O}{8} + \frac{C - FC}{6} = 2.0976 - 0.2031VM \\ 0.3475C + 1.0062H - 0.0695(O - S) = HHV \\ C + H + O + N + S = 100 \end{cases} \quad (16)$$

Symbolic operation in MATLAB was used to solve the equations for complete elemental composition as multivariate linear functions of proximate analyses and HHV:

$$C = 50.7368 - 0.5799FC - 0.7066VM + 2.8301 \text{ HHV} \quad (17)$$

$$H = -13.6696 + 0.2219FC + 0.2704VM - 0.1532 \text{ HHV} \quad (18)$$

$$O = 58.4886 + 0.3316FC + 0.4041VM - 2.5482 \text{ HHV} \quad (19)$$

$$N = 1.7352 + 0.0074FC + 0.0090VM - 0.0359 \text{ HHV} \quad (20)$$

$$S = 2.7089 + 0.0190FC + 0.0232VM - 0.0928 \text{ HHV} \quad (21)$$

The proximate analysis, ultimate analysis, and HHV of coal are on a

dry and ash-free basis in these equations. The merit of the proposed correlations is based on using only proximate analysis data and HHV, which are normally available and regularly updated in power plant databases; these can be used conveniently and economically to predict the complete elemental composition of coal, as mentioned in the prospect section of the literature [15,16,22].

For the verification and validation of the proposed correlations, predictions of C , H , O , N , and S contents were computed using the FC , VM , and HHV values of the 112 coal samples. The MAE, AAE, and ABE were calculated using the measured elemental compositions and are discussed further in sections 4.4 and 4.5.

4.3. Predicting elemental composition using maximum likelihood estimator

The maximum likelihood estimator is based on nine correlation forms presented in Tables 1 and i.e. $m = 9$ in Eq. (10). The probability density distribution of the prediction residuals for each correlation over the dataset is presented graphically in Fig. 2, which indicates that the histogram for each correlation is approximately consistent with a Gaussian distribution (normal distribution). The *normfit* function in MATLAB was used to fit the probability densities as a Gaussian function.

$$\phi_j(z_j) = \frac{1}{\sqrt{2\pi}\sigma_j} e^{-\frac{(z_j - \mu_j)^2}{2\sigma_j^2}} \quad (22)$$

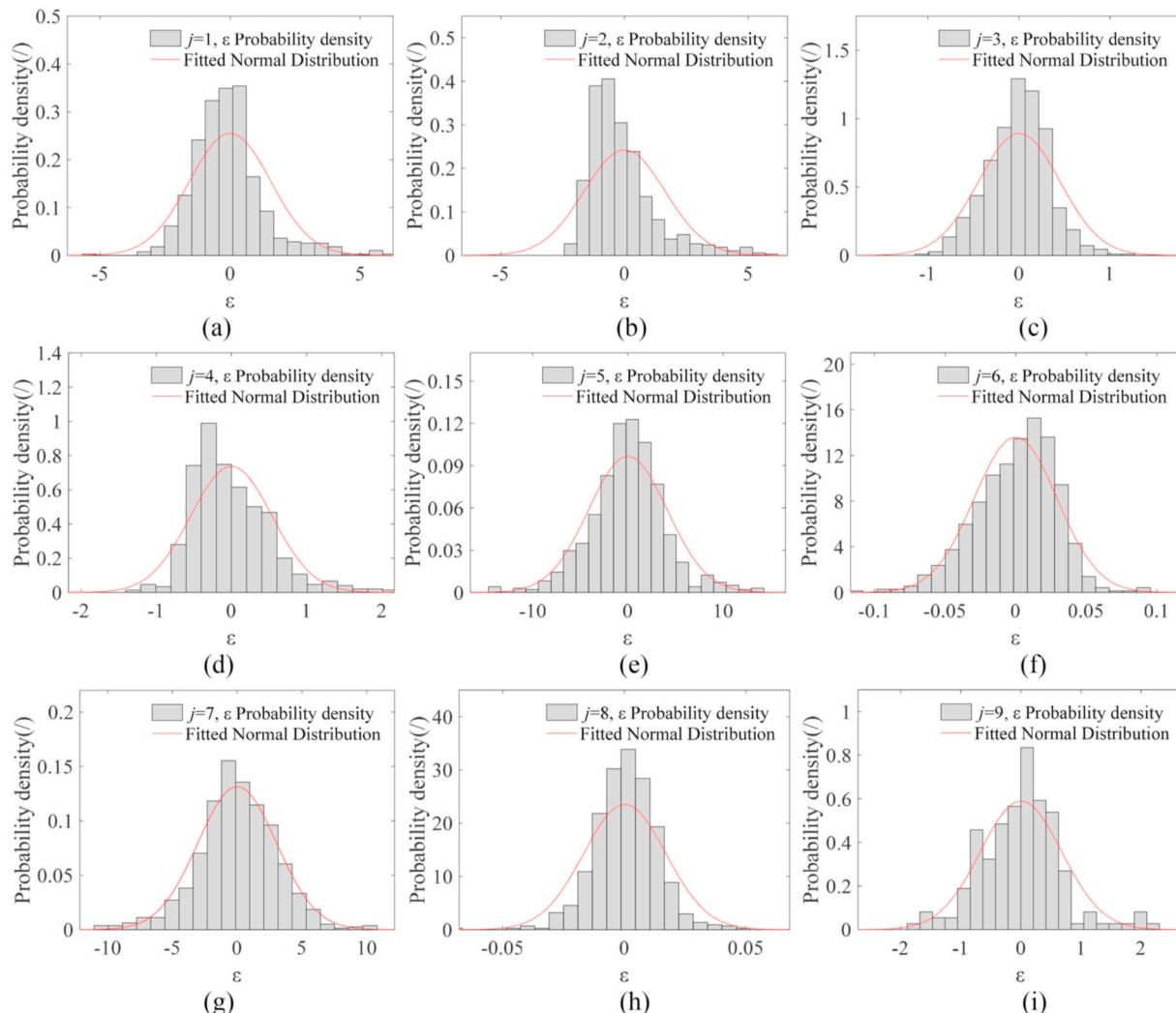


Fig. 2. Comparision of density distribution of prediction residuals with Normal distribution for each correlation.

where μ_j and σ_j represent the mean value and standard deviation of the prediction residuals for correlation j , respectively, as presented in Table 1. The fitted Gaussian function plots of the prediction residuals for each correlation are shown as solid red lines in Fig. 2.

Considering Eq. (22), the elemental composition U that maximises the likelihoods of $L(z_1, z_2, \dots, z_m|U)$ in Eq. (11) can be converted to search for the argument that minimises the negative logarithm of the likelihoods. The final maximum likelihood estimator model is expressed as

$$\left\{ \begin{array}{l} \hat{U} = \operatorname{argmin}_U \sum_j \left(\frac{y - \hat{y} - \mu_j}{\sigma_j} \right)^2 \\ \text{subject to } C + H + O + N + S = 100 \end{array} \right. \quad (23)$$

where \hat{U} ($\hat{U} = (\hat{C}, \hat{H}, \hat{O}, \hat{N}, \hat{S})$) are the estimated results of elemental composition. Eq. (23) uses the algebraic expressions of x and y presented in Table 1 for each correlation, and \hat{y} is calculated using Eq. (8), for which the polynomial coefficients are presented in Fig. 1.

The measured values of FC, VM, and HHV of the coal sample data are used as inputs for the maximum likelihood estimator to predict the elemental compositions. The MAE, AAE, and ABE were calculated over the dataset.

4.4. Validation and verification

Both the MLM and the MLE for the prediction of elemental composition were validated and verified using 112 data points. The MAE, AAE, and ABE were calculated using the dataset. The results show that the MLM has an MAE of 1.97%, 0.48%, 1.40%, 0.21%, 0.87%, an AAE of 2.41%, 9.91%, 19.86%, 14.76%, 88.33%, and an ABE of 0.26%, -0.52%, -1.27%, -3.63%, 62.17% corresponding to the measured values of C , H , O , N , and S , respectively. The MLE approach has a MAE of 1.44%, 0.35%, 1.16%, 0.20%, 0.81%, an AAE of 1.73%, 7.39%, 15.92%, 14.14%, 85.19%, and an ABE of 0.63%, -1.36%, -5.47%, -3.73%, 63.07% corresponding to the measured values of C , H , O , N , and S , respectively. It is clear that the MLE has a better prediction accuracy than the MLM irrespective of the use of MAE or AAE as a validation criterion. This must be because the MLE uses more correlations than the MLM, and the MLE considers the prediction residuals of each correlation, which are neglected by the MLM. However, the MLE has slightly larger positive values and smaller negative values for ABE than the multivariate linear correlations, indicating a slightly higher degree of overestimation or underestimation of elemental content than the multivariate linear correlations.

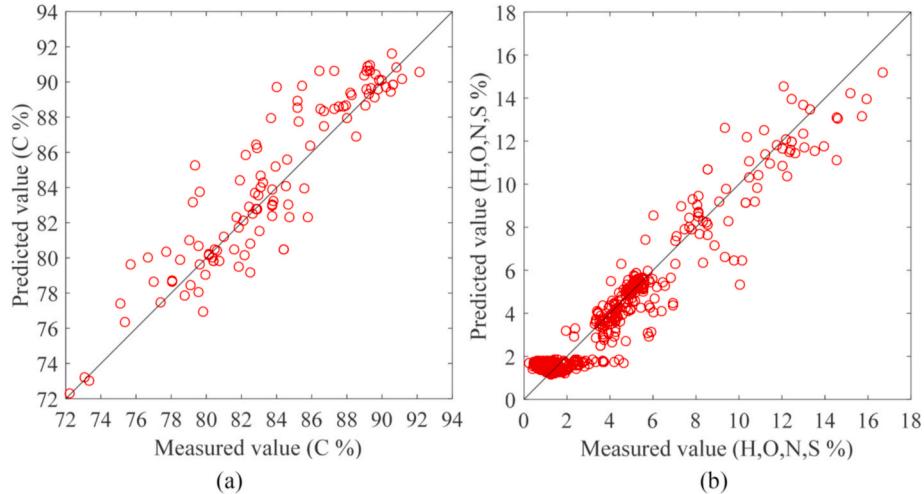


Fig. 3. Validation of maximum likelihood estimator for (a) Carbon, (b) Hydrogen, Oxygen, Nitrogen and Sulphur.

The measured and predicted carbon content values using the MLE are presented graphically in Fig. 3(a). Fig. 3(b) shows the deviation of the predicted and measured content values of H , O , N , and S . The predicted values are close to the main solid line in these plots, indicating good prediction performance. In Fig. 3(b), the measured content values between 0% and 4%, mostly attributed to data points for nitrogen and sulphur contents, and maintain a nearly straight line perpendicular to the y-axis (prediction axis). The scattered data points indicate that the predicted values of nitrogen and sulphur content have relatively small variations and might not capture the trends of the corresponding measured values. Nevertheless, the AAE for nitrogen is 14.14%, which shows that the nitrogen correlation has an acceptable prediction accuracy; the AAE for sulphur is 85.11%, which is higher as compared to that for nitrogen. The smaller AAE values for nitrogen are essentially attributable to marginal nitrogen contents in coal and a small variation range.

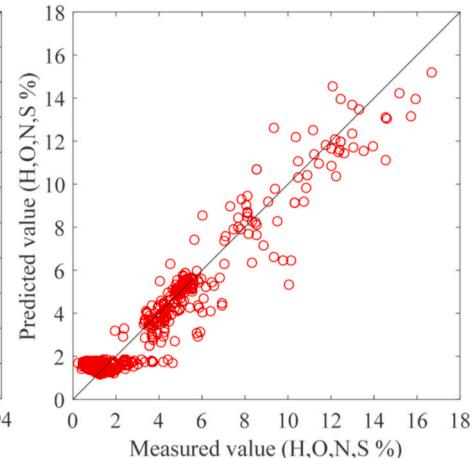
Fig. 4 illustrates how the average absolute error (AAE) of the sulphur correlation varies with the measured sulphur contents of coal samples; the AAE lines of 85.19% and 20% are shown as a green solid line and a blue dotted line, respectively. The AAE reaches the minimum at approximately 1.5% of the sulphur content value, and the AAE is less than 20% when the sulphur content is 1.2–2.3%. The MLM and the MLE have similar AAE characteristics about sulphur content.

4.5. Comparison with published correlations

Parikh et al. [18] and Nhuchhen et al. [16] proposed correlations to estimate the elemental composition of biomass using proximate analysis. Yi et al. [23] presented a series of correlations in different coal rank ranges to predict carbon, hydrogen, and oxygen compositions. Information on these correlations, including algebraic expressions, application ranges, and estimation errors, is presented in Table 2 along with information from the current study (CS).

Parikh, Nhuchhen, and Yi et al. presented correlations only for elemental carbon, hydrogen, and oxygen contents of coal or biomass fuel. The MAE of the Nhuchhen correlations corresponding to the three major elements are 2.58%, 0.41%, and 2.60%, respectively. The MLM in the CS has a MAE of 1.97%, 0.48%, and 1.40%; the maximum likelihood estimator has a MAE of 1.44%, 0.35%, and 1.16%, corresponding to the measured values of C , H , and O , respectively. Thus, both the multivariate linear correlations and the maximum likelihood estimator proposed in the CS have a better prediction accuracy than the correlations developed by Nhuchhen.

It can be observed from Table 2 that the Nhuchhen correlations are applicable to a wider range of biomass types than the Parikh



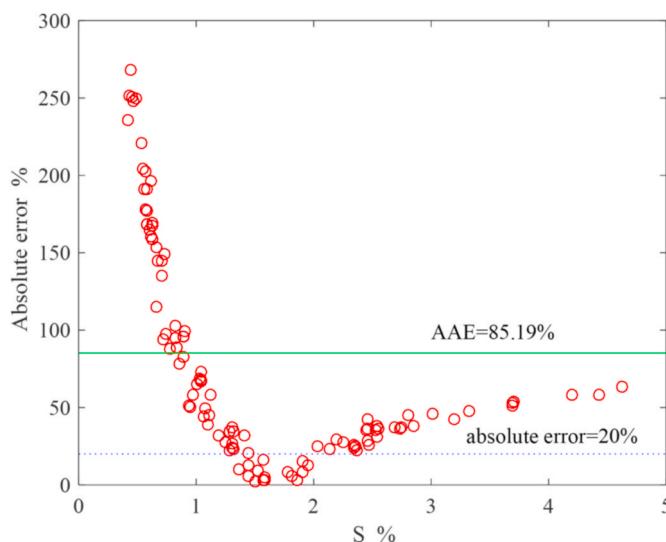


Fig. 4. Variation of absolute error corresponding to the measured values of sulphur.

correlations. As a result, the Nhuchhen correlations have larger AAE and ABE values than the Parikh correlations. However, the Nhuchhen correlations, including a wider range of data than the Parikh correlations, have a better prediction ability, especially for torrefied biomass fuels [16]. This clearly indicates that the applicable data range of the derived correlations and that the MAE, AAE, and ABE dictate prediction ability.

Yi et al. proposed different sets of correlations separately in each relatively small data range for anthracite, high-rank bituminous, sub-bituminous, and lignite; these correlations offer predictions with an AAE of 1.87%, 7.81%, and 12.16% corresponding to the measured values of C, H, and O, respectively, and indicate a slightly better prediction accuracy than one set of correlations [Eqs. (17–21)] in the data range considered in the CS. However, multiple sets of correlations introduce inconvenience in application.

Additionally, the AAE of the maximum likelihood estimator is

1.73%, 7.39%, and 15.92% for the measured values of C, H, and O, respectively, demonstrating better prediction accuracy than the Yi correlations for carbon and hydrogen contents. The advantage of the MLM and the maximum likelihood estimator proposed in the CS is the ability to predict comprehensive elemental composition for ultimate analysis based on proximate analysis data and HHV.

4.6. Model uncertainties

For further verification and to determine the uncertainties of the MLM and MLE for predicting the elemental composition of coal, the root mean square error (RMSE) was calculated using the presented dataset. Table 3 presents the calculated RMSE values for the MLM correlations and the MLE approach. The RMSE values of the back propagation neural network (BP-ANN) model presented by Sapto [22] were found to be higher than those of the MLE approach. The MLM correlations also have lower RMSE values than the BP-ANN model, except for the predicted hydrogen content. This must be due to the deliberately selected model inputs and composition basis of coal in the CS.

Fig. 5 shows the predicted and measured contents of C, H, O, N, and S for 112 coal samples, together with the 95% confidence interval. In Fig. 5, the plotted points are rearranged in ascending order of the measured element contents for clearer illustration. Most of the data points fall into the confidence intervals of the corresponding components in the ultimate analysis. Fig. 5(a–c) indicate that the predicted content values for C, H, and O can precisely capture the trends of the corresponding measured values. Fig. 5(d and e) show that the predicted N and S content values rarely follow the variations of the corresponding

Table 3

Comparison of RMSE of MLM correlations, MLE approach and BP-ANN.

Component	MLM RMSE	MLE RMSE	BP-ANN in Ref. [22] RMSE
C	2.72	1.91	4.99
H	0.63	0.47	0.44
O	1.83	1.41	2.34
N	0.27	0.24	/
S	1.03	0.95	/

Table 2

Comparison of the proposed correlations and maximum likelihood estimator with published correlations.

Fuel	Item	Correlations & Model	Ref.	MAE	AAE	ABE	Range
raw and torrefied biomass	C	$C = -35.9972 + 0.7698VM + 1.3269FC + 0.3250ASH$	[16]	2.58	5.23	0.45	VM:13.30–93.60
	H	$H = 55.3678 - 0.4830VM - 0.5319FC - 0.5600ASH$		0.41	9.94	2.82	FC:0.67–82.74
	O	$O = 223.6805 - 1.7226VM - 2.2296FC - 2.2463ASH$		2.60	8.79	2.01	ASH:0.01–48.70 C:19.12–86.28 H:0.57–7.46 O:4.31–49.59 (in wt.% dry basis)
raw biomass	C	$C = 0.637FC + 0.455VM$	[18]	/	3.21	0.21	VM: 57.20–90.60
	H	$H = 0.052FC + 0.062VM$		/	4.79	-0.15	FC: 4.70–38.40
	O	$O = 0.304FC + 0.476VM$		/	3.40	0.49	C: 36.20–53.10 H: 4.36–8.30 O: 31.37–49.50 (in wt.% dry basis)
coal	C	$y = a_1 + a_2ASH + a_3ASH^2 + a_4VM + a_5VM^2 + a_6FC + a_7FC^2$ (where y stands for the elements of C, H, and O, a_i depending on the data range divided for Anthracite, High-rank bituminous, subbituminous and Lignite)	[23]	/	1.87	-0.05	VM: 1.87–49.7
	H			/	7.81	-1.67	FC: 35.64–91.60
	O			/	12.31	-2.25	ASH: 1.5–27.19 C: 49.70–90.80 H: 0.95–5.40 O: 1.02–24.40 (in wt.% dry basis)
coal	C	$C = 50.7368 - 0.5799FC - 0.7066VM + 2.8301HHV$	CS	1.97	2.41	0.26	VM: 7.02–48.20
	H	$H = -13.6696 + 0.2219FC + 0.2704VM - 0.1532 HHV$		0.48	9.91	-0.52	FC: 51.80–92.98
	O	$O = 58.4886 + 0.3316FC + 0.4041VM - 2.5482 HHV$		1.40	19.86	-1.27	C: 72.23–92.16
	N	$N = 1.7352 + 0.0074FC + 0.0090VM - 0.0359 HHV$		0.21	14.76	-3.63	H: 3.30–6.85
	S	$S = 2.7089 + 0.0190FC + 0.0232VM - 0.0928 HHV$		0.87	88.33	62.17	O: 1.94–21.12
coal	C	MLE	CS	1.44	1.73	0.63	N: 0.78–2.04
	H			0.35	7.39	-1.36	S: 0.25–4.62 (in wt.% dry and ash free basis)
	O			1.16	15.92	-5.47	
	N			0.20	14.14	-3.73	HHV: 27.47–35.77 (HHV: Higher heat value on dry and ash free basis, MJ/kg)
	S			0.81	85.19	63.07	

measured values, indicating that nitrogen and sulphur contents are potentially dominated by random factors. This is also supported by results presented in Fig. 3(d–e).

5. Conclusions

To establish MLM of the elemental composition of coal using proximate analysis and HHV, a novel method was developed by solving a set of equations including correlations between the coal properties and the constraint of elemental composition. Compared with several published correlations estimating only the C, H, and O contents of coal or biomass fuel based on proximate analysis, the correlations presented in this study are capable of predicting complete elemental compositions for ultimate analysis, which could potentially be applied in performance monitoring and combustion tuning of boilers.

A model based on MLE was also established for predicting the

carbon, hydrogen, oxygen, nitrogen, and sulphur contents of coal. Information on the probability density of prediction residuals, which has been used as weighting to consider the prediction performance of correlations, was extracted over the derived data points. Hence, the MLE approach has a better prediction accuracy than the multivariate linear correlations, which neglect prediction residuals.

A total of 743 data points from published literature were used to derive the MLM and MLE models, which were validated and verified using an additional 112 experimentally measured data points representing a variety of coal types. The MLE and MLM results show that the predicted carbon, hydrogen, and oxygen contents correctly capture the trends of the corresponding measured values, with an AAE of 1.73%, 7.39%, and 15.92%, corresponding to the measured values of C, H, and O, respectively. The predicted N contents rarely follow the variations of the corresponding measured values, but still offer acceptable prediction accuracy owing to the marginal and relatively stable content of nitrogen

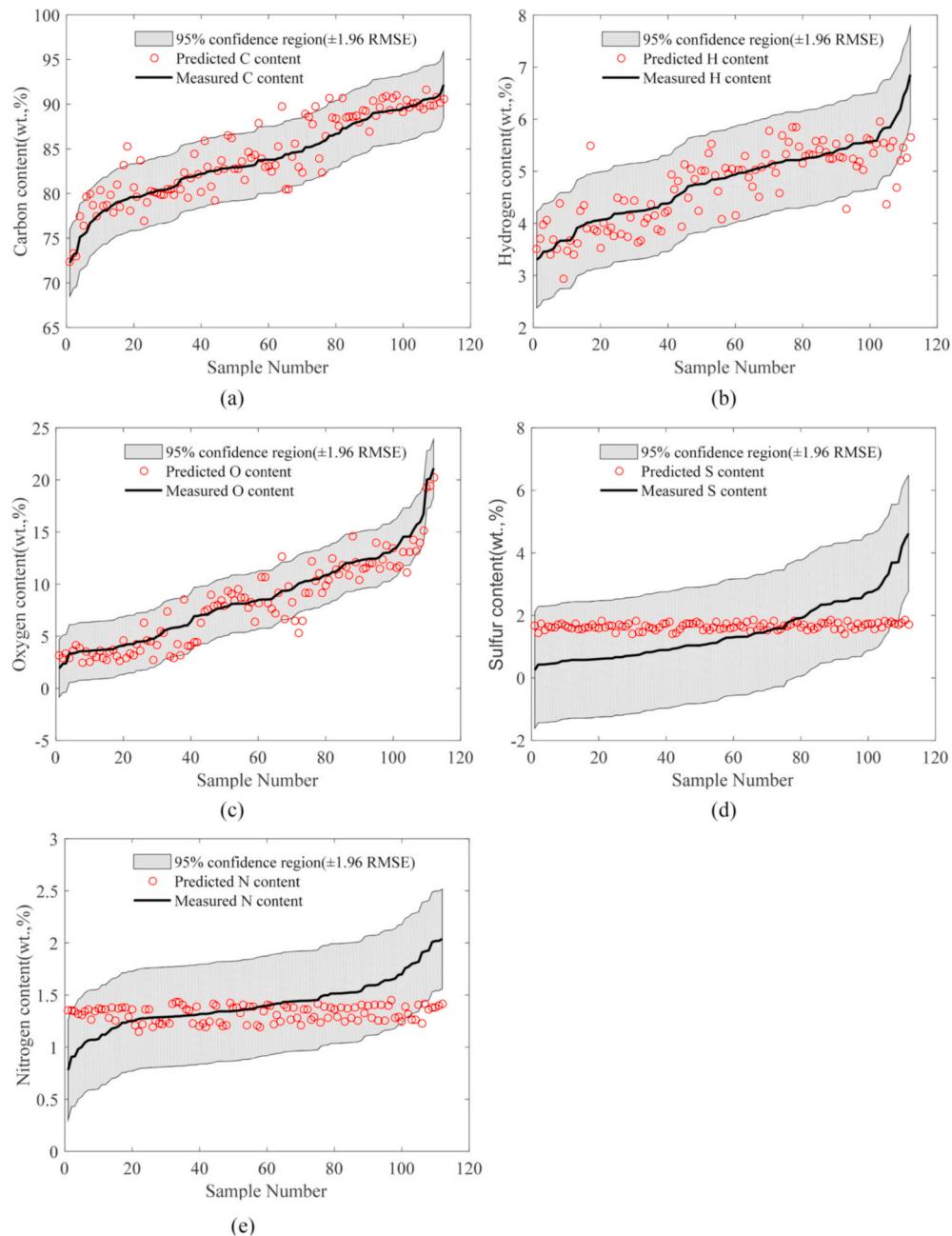


Fig. 5. The predicted and measured content values of (a) Carbon, (b) Hydrogen, (c) Oxygen, (d) Nitrogen, (e) Sulphur.

in coal.

The clear advantage of the MLM correlations and the MLE approach in the CS is the ability to predict comprehensive elemental compositions for ultimate analysis over a wide data range with better prediction accuracy than the existing correlations. To emphasise, the proposed models should be used with caution to predict sulphur content, especially for coal samples with high or low sulphur contents. However, the average absolute error (AAE) can be less than 20% for coals with a sulphur content in the range of 1.2–2.3%.

Credit author statement

LIU Fuguo: Conceptualization, Methodology, Software, Data curation, Writing – original draft preparation, Visualization, Investigation, Writing- Reviewing and Editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

- [1] Francisco Rodriguez, E. Tova, Vicente Cortés, Luis Canadas, OPTICOM: advanced automatic monitoring system of local combustion conditions for improving boiler performance in PC power plants[J], Fuel 81 (5) (2002) 637–645.
- [2] M.A. Field, Measurements of the effect of rank on combustion rates of pulverized coal[J], Combust. Flame 14 (2) (1970) 237–248.
- [3] C.K. Man, J.R. Gibbins, J.G. Witkamp, J. Zhang, Coal characterisation for NO_x prediction in air-staged combustion of pulverised coals[J], Fuel 84 (17) (2005) 2190–2195.
- [4] P.H. Young, L.J. Eun, K.H. Hee, Sangbin Park, Se Hyun Baek, Insoo Ye, Thermal resistance by slagging and its relationship with ash properties for six coal blends in a commercial coal-fired boiler[J], Fuel 235 (2019) 1377–1386.
- [5] S.H. Baek, H.Y. Park, S.H. Ko, The effect of the coal blending method in a coal fired boiler on carbon in ash and NO_x emission[J], Fuel 128 (14) (2014) 62–70.
- [6] R.I. Backreedy, J.M. Jones, L. Ma, M. Pourkashanian, A. Williams, A. Arenillas, et al., Prediction of unburned carbon and NO_x in a tangentially fired power station using single coals and blends[J], Fuel 84 (17) (2005) 2196–2203.
- [7] A. Saptoro, H.M. Yao, M.O. Tadé, H.B. Vuthaluru, Prediction of coal hydrogen content for combustion control in power utility using neural network approach[J], Chemometr. Intell. Lab. Syst. 94 (2) (2008) 149–159.
- [8] J. Parikh, S.A. Channiwala, G.K. Ghosal, A correlation for calculating HHV from proximate analysis of solid fuels[J], Fuel 84 (5) (2005) 487–494.
- [9] S.A. Channiwala, P.P. Parikh, A unified correlation for estimating HHV of solid, liquid and gaseous fuels[J], Fuel 81 (8) (2002) 1051–1063.
- [10] A.K. Majumder, R. Jain, P. Banerjee, J.P. Barnwal, Development of a new proximate analysis based correlation to predict calorific value of coal [J], Fuel 87 (13–14) (2008) 3077–3081.
- [11] S. Mesroghli, E. Jorjani, S.C. Chelgani, Estimation of gross calorific value based on coal analysis using regression and artificial neural networks[J], Int. J. Coal Geol. 79 (1–2) (2009) 49–54.
- [12] S.C. Chelgani, J.C. Hower, E. Jorjani, Sh Mesroghli, A.H. Bagherieh, Prediction of coal grindability based on petrography, proximate and ultimate analysis using multiple regression and artificial neural network models[J], Fuel Process. Technol. 89 (1) (2008) 13–20.
- [13] S. Ural, M. Akyildiz, Studies of the relationship between mineral matter and grinding properties for low-rank coals[J], Int. J. Coal Geol. 60 (1) (2004) 81–84.
- [14] M. Khandelwal, T.N. Singh, Prediction of macerals contents of Indian coals from proximate and ultimate analyses using artificial neural networks[J], Fuel 89 (5) (2010) 1101–1109.
- [15] J.R. Smith, J.W. Smith, A relationship between the carbon and hydrogen content of coals and their vitrinite reflectance[J], Int. J. Coal Geol. 70 (1–3) (2007) 79–86.
- [16] Nuhuchhen, D. Ram, Prediction of carbon, hydrogen, and oxygen compositions of raw and torrefied biomass using proximate analysis[J], Fuel 180 (2016) 348–356.
- [17] J. Shen, S. Zhu, X. Liu, H. Zhang, J. Tan, The prediction of elemental composition of biomass on proximate analysis, Energy Convers. Manag. 51 (2010) 983–987.
- [18] J. Parikh, S.A. Channiwala, G.K. Ghosal, A correlation for calculating elemental composition from proximate analysis of biomass materials[J], Fuel 86 (12) (2007) 1710–1719.
- [19] R. Saidur, E.A. Abdelaziz, A. Demirbas, M.S. Hossain, S. Mekhilef, A review on biomass as a fuel for boilers [J], Renew. Sustain. Energy Rev. 15 (2011) 2262–2289.
- [20] E.K. Vakkilainen, Estimation of elemental composition from proximate analysis of black liquor[J], Pap. Ja Puu 82 (7) (2000) 450–454.
- [21] H.M. Yao, H.B. Vuthaluru, M.O. Tadé, D. Djukanovic, Artificial neural network-based prediction of hydrogen content of coal in power station boilers[J], Fuel 84 (12–13) (2005) 1535–1542.
- [22] A. Saptoro, H.B. Vuthaluru, M.O. Tadé, A comparative study of prediction of elemental composition of coal using empirical modelling[J], IFAC Proc. Vol. 39 (2) (2006) 747–752.
- [23] L. Yi, J. Feng, Y.H. Qin, Yu-Hong Qin, Wen-Ying Li, Prediction of elemental composition of coal using proximate analysis[J], Fuel 193 (2017) 315–321.
- [24] Yuchen Si, Yingjie Yang, Myra Martel, Lifeng Zhang, Shelley Kirychuk, Characterization of electrical current and liquid droplets deposition area in a capillary electrospray[J], Results Eng. 9 (2021) 100206.
- [25] Younes El khchine, Mohammed Sriti, Performance evaluation of wind turbines for energy production in Morocco's coastal regions, J. Results Eng. 10 (2021) 100215.
- [26] P. Chen, Properties. Classification and Utilization of China Coal, Chemical Industry Press, Beijing, 2007.
- [27] Changwu Cao, The discussion of characteristics for receiving coal in power plants of shandong[J], Shandong Elec. Power 3 (1997) 46–56.
- [28] TPRI, Present Situation and Prospect of Large Thermal Power Units in China[R], Xi'an, Xi'an Thermal Power Research Institute, 1993.
- [29] J. Zhang, Z.L. Fan, X.F. Lin, Y.Q. Xu, Effect of ashing temperature on the characteristics of biomass ash[J], J. Fuel Chem. Technol. 32 (5) (2004) 547–551.
- [30] R.C. Neavel, S.E. Smith, E.J. Hippo, N.M. Robert, Interrelationships between coal compositional parameters[J], Fuel 65 (3) (1986) 312–320.
- [31] J.P. Mathews, V. Krishnamoorthy, E. Louw, H. Aime, N. Tchapda, Fidel Castro-Marcano, Vamsi Karri, et al., A review of the correlations of coal properties with elemental composition[J], Fuel Process. Technol. 121 (2014) 104–113.