Journal of the Energy Institute xxx (2016) 1–9

Contents lists available at ScienceDirect

Journal of the Energy Institute

journal homepage: http://www.journals.elsevier.com/journal-of-the-energy-institute

Genetic programming based high performing correlations for prediction of higher heating value of coals of different ranks and from diverse geographies

Suhas B. Ghugare [a, b, c], Sanjeev S. Tambe [a, *]

[a] Artificial Intelligence Systems Group, Chemical Engineering and Process Development Division, CSIR-National Chemical Laboratory, Dr. Homi Bhabha Road, Pashan, Pune 411008, India
[b] Chemical Engineering Department, All India Shri Shivaji Memorial Society's College of Engineering (AISSMSCOE), Kennedy Road, Pune 411001, India
[c] Savitribai Phule Pune University (SPPU), Ganeshkhind Road, Pune 411007, India

ARTICLE INFO

Article history:
Received 14 December 2015
Received in revised form
23 February 2016
Accepted 1 March 2016
Available online xxx

Keywords:
Coal
Higher heating value
Genetic programming
Proximate analysis
Ultimate analysis

ABSTRACT

The *higher heating value* (HHV) is the most important indicator of a coal's potential energy yield. It is commonly used in the efficiency and optimal design calculations pertaining to the coal combustion and gasification processes. Since the experimental determination of coal's HHV is tedious and time-consuming, a number of proximate and/or ultimate analyses based correlations—which are mostly linear—have been proposed for its estimation. Owing to the fact that relationships between some of the constituents of the proximate/ultimate analyses and the HHV are nonlinear, the linear models make suboptimal predictions. Also, a majority of the currently available HHV models are restricted to the coals of specific ranks or particular geographical regions. Accordingly, in this study three proximate and ultimate analysis based nonlinear correlations have been developed for the prediction of HHV of coals by utilizing the computational intelligence (CI) based *genetic programming* (GP) formalism. Each of these correlations possesses following noteworthy characteristics: (i) the highest HHV prediction accuracy and generalization capability as compared to the existing models, (ii) wider applicability for coals of different ranks and from diverse geographies, and (iii) structurally lower complex than the other CI-based existing HHV models. It may also be noted that in this study, the GP technique has been used for the first time for developing coal-specific HHV models. Owing to the stated attractive features, the GP-based models proposed here possess a significant potential to replace the existing models for predicting the HHV of coals.

1. Introduction

The potential energy yield (total heat content) of a unit mass of coal is determined in terms of the *higher heating value* (HHV) (also known as *gross calorific value*). It is defined as the amount of heat evolved when a unit weight of the fuel is burnt completely and the combustion products cooled to a standard temperature of 298 K and at a standard pressure of 101.33 kPa [1]. The proximate and/or ultimate analyses of coals and their HHVs are strongly correlated. While the proximate analysis determines the individual content of *moisture*, *volatile matter*, *ash*, and *fixed carbon* in a coal, the ultimate analysis measures the amounts of various elements, namely, *carbon*, *hydrogen*, *nitrogen*, *sulphur*, and *oxygen*. Since HHV is a major indicator of coal's quality, it is used extensively in: (a) the efficiency and optimal design calculations of coal combustion and gasification equipment [2,3], (b) defining coal's rank (over much of the rank range), and (c) evaluating the pollution compliance of coal-based processes [4].

There exist mostly linear models that correlate the constituents of a coal's proximate and/or ultimate analysis to its HHV [5–9]. The early HHV prediction models were developed for the coals with specific ranks or from particular geographical regions; these were also based on limited amounts of data [5–16]. Owing to its importance in designing, operating and optimizing coal-based processes, attempts to develop HHV predicting models with ever increasing prediction accuracies still continue. In recent years, generalized correlations encompassing

* Corresponding author. Tel.: +91 20 2590 2156; fax: +91 20 2590 2621.
E-mail address: ss.tambe@ncl.res.in (S.S. Tambe).

http://dx.doi.org/10.1016/j.joei.2016.03.002
1743-9671/

Table 1
A representative list of the correlations for predicting HHV of coal.

Reference	Model	Eqn. no.	Material	HHV basis
Majumder et al. [12]	HHV (MJ/kg) = −0.03ASH − 0.11*M* + 0.33VM + 0.35FC	1	Coal	Wet, Proximate
Mesroghli et al. [20]	HHV (MJ/kg) = 37.777 − 0.647*M* − 0.387ASH − 0.089VM	2	Coal	Wet, Proximate
Kavšek et al. [15]	HHV (MJ/kg) = −3.57 + 0.31VM + 0.34FC	3	Coal	Wet, Proximate
Dulong [8]	HHV (kcal/kg) = 81C + 342.5(H − (O/8)) + 22.5S − 6(9H − *M*)	4	Coal	Wet, Ultimate
Mesroghli et al. [20]	HHV (MJ/kg) = −26.29 + 0.275ASH + 0.605C + 1.352H + 0.840N + 0. 321S	5	Coal	Wet, Ultimate
Mesroghli et al. [20]	HHV (MJ/kg) = 6.971 + 0.269C + 0.195N − 0.061ASH − 0.251O_{ex} + 1.08H_{ex} − 0.21*M*	6	Coal	Wet, Ultimate
Cordero et al. [16]	HHV (MJ/kg) = 0.3543FC + 0.1708VM	7	Biomass	Dry, Proximate
Parikh et al. [17]	HHV (MJ/kg) = 0.3536FC + 0.1559VM − 0.0078ASH	8	Biomass	Dry, Proximate
Ghugare et al. [18]	HHV (MJ/kg) = 0.365FC + 0.131VM + (1.397/FC) + (328.568VM/(10283.138 + 0.531FC^3ASH − 6.863FC^2ASH))	9	Biomass	Dry, Proximate

M: Moisture (%), FC: Fixed carbon (%), VM: Volatile matter (%), ASH: Ash (%), C: Carbon (%), H: Hydrogen (%), H_{ex}: Hydrogen (%) exclusive of that in moisture, O: Oxygen (%), O_{ex}: Oxygen (%) exclusive of that in moisture, N: Nitrogen (%), S: Sulphur (%).

several ranks of coal as also unified ones for various types of biomass fuels (including coal) have been proposed [17,18] —a good review of which can be found in Mathews et al. [19]. These efforts have resulted in classical regression based models [4–13] as also computational intelligence (CI) based models. The latter type of models are based on *artificial neural networks* (ANNs) [14,15,18,20–23], *co-active neuro-fuzzy adaptive networks* (CANFIS) [21], *alternating conditional expectation* (ACE) [23] and *support vector regression* (SVR) [23,24] formalisms. Commonly, these models employ weight percentages of the constituents of the coal's proximate and/or ultimate analyses as inputs. A comprehensive list of the regression-based correlations predicting the HHV of coals is provided in Table 1; some of these correlations are applicable also to solid biofuels.

In Table 1, it is seen that a majority of the HHV models are linear. Based on extensive experimental data, Patel et al. [14] and Tan et al. [24] have developed CI-based inherently nonlinear coal HHV models possessing better prediction and generalization abilities than the linear models; this improved performance of the nonlinear models indicates that the relations between the HHV and some constituents of the coal's proximate and ultimate analyses are indeed nonlinear. It may be noted that the stated models were developed using data pertaining mostly to coals from India, USA and China. Thus, these models are not applicable for a large variety of coals mined globally. Consequently, there exists a clear need to develop coal-specific, comprehensive, and widely applicable nonlinear models possessing high HHV prediction accuracies and generalization capabilities as export/import of coal from one region to another is a common international trade practice. Accordingly, this study reports the development and performance evaluation of *genetic programming* (GP) based models for the prediction of HHV of coals. The GP-based models proposed in this study use constituents of the proximate or ultimate analyses of coals as inputs. The significance of the GP formalism in the present study is that depending upon the nature of the dependencies between the constituents of the proximate/ultimate analysis and the corresponding HHV, the technique by itself chooses an appropriate linear or a nonlinear model that optimally fits the example data. The highlights of this study are: (i) genetic programming has been used for the first time in the development of models for the prediction of coal HHV, (ii) a large set consisting of the proximate and ultimate analyses, and the corresponding HHVs of 7682 coal samples from 33 countries and belonging to different ranks, has been utilized in constructing the three GP-based models, (iii) the prediction accuracy and generalization capability of each of the three GP-based models have been found to be superior to the currently available high performing HHV models, and (iv) all the GP-based models are nonlinear and possess much lower complexity compared to the corresponding ANN, SVR, CANFIS, and ACE based HHV prediction models.

This paper is structured as follows. Section 2 provides an overview of the GP formalism. Section 3 titled “Results and discussion” first describes the development of three GP-based HHV predicting models followed by presentation of the results of a comparison of the prediction and generalization performance of the GP-based and their currently available high performing competitor models. Finally, section 4 provides the concluding remarks wherein key findings of this study are summarized.

2. Methods

2.1. Genetic programming (GP)

Genetic programming [25] is an evolutionary, population-based, and stochastic search and optimization methodology based on the Darwinian principles of natural selection and genetic propagation of characteristics. It was originally proposed to generate automatically computer codes performing pre-defined tasks. The other significant application of GP, namely, “symbolic regression” is of interest to this study.

The objective of the GP-based symbolic regression (GPSR) is to search and optimize the form and associated parameters of a linear/nonlinear function given by,

$$y = f(\mathbf{x}, \mathbf{c}) \tag{10}$$

which optimally fits an example input–output dataset consisting of N number of input–output patterns, where, $\mathbf{x} = [x_1, x_2, \ldots, x_I]^T$ refers to an I-dimensional vector of the model inputs (predictor/independent variables), y denotes the scalar model output (dependent variable, HHV), and $\mathbf{c} = [c_1, c_2, \ldots, c_J]^T$ represents a J-dimensional vector of function parameters. The GPSR possesses following characteristics: (a) given an example dataset consisting of inputs and outputs, the technique searches and optimizes an appropriate linear/nonlinear form (structure) and all the parameters of a function (model) that fits the dataset optimally, (b) it does not make any assumptions regarding the form and associated parameters of the data fitting function, (c) invariably, GPSR obtains models of lower complexity compared to the models secured using other CI-based methodologies, such as ANNs and SVR, and (d) due to their lower complexity, GPSR-based models are easier to understand and deploy in a practical setting than the corresponding ANN and SVR models.

Please cite this article in press as: S.B. Ghugare, S.S. Tambe, Genetic programming based high performing correlations for prediction of higher heating value of coals of different ranks and from diverse geographies, Journal of the Energy Institute (2016), http://dx.doi.org/10.1016/j.joei.2016.03.002