Multi-gene genetic programming based predictive models for municipal solid waste gasification in a fluidized bed gasifier

Daya Shankar Pandey[a], Indranil Pan[b], Saptarshi Das[c], James J. Leahy[a], Witold Kwapinski[a*]

[a]*Carbolea Research Group, Department of Chemical and Environmental Science, University of Limerick, Ireland*
[b]*Energy, Environment, Modelling and Minerals (E^2M^2) Research Section, Department of Earth Science and Engineering, Imperial College London, Exhibition Road, London SW7 2AZ, United Kingdom*
[c]*School of Electronics and Computer Science, University of Southampton, Southampton SO17 1BJ, United Kingdom*
*Corresponding author:
Tel.: +353 61234935; fax: +353 61202568.
E-mail address: witold.kwapinski@ul.ie (W. Kwapinski).

Abstract

A multi-gene genetic programming technique is proposed as a new method to predict syngas yield production and the lower heating value for municipal solid waste gasification in a fluidized bed gasifier. The study shows that the predicted outputs of the municipal solid waste gasification process are in good agreement with the experimental dataset and also generalize well to validation (untrained) data. Published experimental datasets are used for model training and validation purposes. The results show the effectiveness of the genetic programming technique for solving complex nonlinear regression problems. The multi-gene genetic programming are also compared with a single-gene genetic programming model to show the relative merits and demerits of the technique. This study demonstrates that the genetic programming based data-driven modelling strategy can be a good candidate for developing models for other types of fuels as well.

***Keywords*:** municipal solid waste, genetic programming, gasification, fluidized bed gasifier

1. Introduction

The disposal of municipal solid waste is an ever-increasing problem in the European Union (EU) and other developing countries (Guerrero et al., 2013; Pires et al., 2011). Due to strict environmental standards, current solid waste management practices (landfills, inceneration) are under intense examination and innovative technologies are becoming attractive alternative

options (Pires et al., 2011). There are several alternatives to dispose municipal solid waste including thermal, biochemical and mechanical processes. Incineration has been extensively used in EU and other developed countries including Japan and Singapore for disposal and energy recovery from the wastes (Narayana, 2009). However, the flue gases from the waste incinerators contains high amount of particulate matter, NO_x, SO_x, dioxins and furans (Cheng and Hu, 2010). Apart from the high amount of emissions, incineration systems have high operating cost with relatively lower energy efficiency (Arena, 2012). One attractive thermal alternative to incineration is the municipal solid waste gasification. The gasification process can generate the electricity from the waste with an efficiency of 34% compared to incineration process, which has thermal efficiency around 20% (Murphy and McKeogh, 2004). It has been suggested that gasification is a viable technology for processing solid wastes, including municipal solid waste, while complying with present emission standards (Arena, 2012). This also offers an alternative solution to the landfilling option. Compared to other treatment processes, gasification technology is an attractive solution for the treatment of municipal solid waste while simultaneously minimizing pollution (Malkow, 2004; Xiao et al., 2007). The derived syngas from municipal solid waste gasification can be used to generate heat and electricity, which will help to offset the use of fossil fuels.

Gasification is the thermal conversion process of any carbonaceous fuel to a gaseous product with useable heating value. It is commonly performed with only a third of the oxygen necessary for complete combustion. Gasification includes pyrolysis, partial oxidation and hydrogenation whereas the dominant process is partial oxidation (Higman and Van der Burgt, 2011), resulting in gaseous products (hydrogen, carbon monoxide, carbon dioxide, water and other gaseous hydrocarbons), and a small amount of char, ash and condensable compounds (tars). Air, steam or oxygen can be used as a gasifying agent. For solid fuel combustion, gasification reactors can be categorised into three distinctive types: fixed bed (updraft and downdraft), fluidized bed and entrained flow gasifiers (Higman and Van der Burgt, 2011).

Biomass gasification is a complex thermochemical process (Puig-Arnavat et al., 2010). In the recent past, numerous researchers have tried to simulate a realistic gasification process and optimized the process analysis to make it more cost effective. Most of the fluidized bed (FB) biomass gasifier models fit reasonably well with the experiments selected for validation using various empirical correlations. However, there are very few measurements available for detailed

validation specifically for large scale gasifiers (Gómez-Barea and Leckner, 2010). Since conducting large scale gasification experiments are quite expensive and time consuming, modelling can be a viable alternative which saves both time and money. However, simulation of municipal solid waste processes are computationally expensive and fast meta-models are required to integrate these models into other systems level models which look at the whole value chain to conduct life cycle analysis, or other system level optimisation procedures. In general, mathematical models are exploited to investigate the influence of the main process parameters on calorific value and yield of the product gas. Irrespective of the type of reactors, several modelling techniques such as thermodynamic equilibrium models, kinetic rate models, Aspen Plus based models and artificial neural networks have been implemented for gasification systems (Puig-Arnavat et al., 2010). The artificial intelligence techniques such as artificial neural networks, genetic programming etc. demands less system information compared to equilibrium and kinetic based modelling, hence, these techniques can be useful for modelling FB gasifiers. In view of the complexity involved with the gasification process, a novel artificial intelligence paradigm known as genetic programming has been used to model the gasification system in the present study. The main objective of the present study is to show the application of the genetic programming approach in predicting syngas yield and heating value. To the best of the author's knowledge this is the first study using the multi-gene genetic programming technique to predict the lower heating value and yield of syngas produced from municipal solid waste.

In the recent past, artificial neural networks techniques have been extensively used by several researchers in the fields of pattern recognition, signal processing, function approximation, weather prediction and process simulations (Guo et al., 1997). Lately it has also received attention as a tool in renewable energy system prediction and modelling (Kalogirou, 2001). A back propagation neural network using the Levenberg–Marquardt (LM) algorithm has been applied to a hybrid upflow anaerobic sludge blanket reactor to predict the bio-degradation and bio-hydrogen production using distillery wastewater (Sridevi et al., 2014). A hybrid neural network model was developed for predicting the product yield and gas composition in an atmospheric steam blown fluidized bed gasifier. The authors tested four different kinds of biomass on a bench scale gasifier for training the hybrid neural network model. This study revealed that the feed forward neural network prediction was better than the traditional regression models (Guo et al., 2001). A feed forward neural network model was used to predict

the lower heating value of municipal solid waste from its chemical composition. It was concluded that the neural network model has better precision over the traditional model (Dong et al., 2003). A combined non-stoichiometric equilibrium approach with an artificial neural networks regression model was developed to predict product composition in an atmospheric air gasification fluidized bed reactor (Brown et al., 2006). A complete set of stoichiometric equations were formulated to explain the non-equilibrium behaviour for gas, tar, and char formation by reaction temperature difference. The artificial neural networks regression related temperature differences to fuel composition and operational variables. This first principle approach, illustrated with FB data, improves the accuracy of the equilibrium based model and reduces the data requirement by preventing neural network to learn from atomic and heat balances (Brown et al., 2006). The combination of equilibrium and artificial neural networks models were further investigated and improved by the same authors (Brown et al., 2007). An attempt was made to develop an artificial neural networks model for predict to gasification characteristics of the municipal solid waste (Xiao et al., 2009). Two different artificial neural networks based models were introduced to predict gas production rate and heating value of the product gas in a steady state fluidized bed coal gasifier (Chavan et al., 2012). Recently, two artificial neural networks models were presented (Puig-Arnavat et al., 2013); one for a circulating fluidized bed gasifier and another for a bubbling fluidized bed gasifier for estimating the product gas composition (CO, CO_2, H_2 and CH_4) and gas yield. The results show good agreement with the experimental data.

Despite prediction capability of artificial intelligence based techniques, only artificial neural networks have been used in the modelling of FB gasifiers. Very few applications of genetic programming have been reported in recent literature focused on predicting syngas production and the lower heating value of syngas. An extensive literature review shows that so far only a few studies have been reported where the GP strategy has been employed for the modelling of fluidized bed gasifier. To the best of the author's knowledge this is the first study using the multi-gene genetic programming technique to predict the lower heating value and yield of syngas produced from municipal solid waste.

Recently, the multilayer perceptron neural network model and genetic programming have been used to predict $CO+H_2$ generation rate, syngas production rate, carbon conversion and heating value of the syngas in a pilot-plant scale FB coal gasifier (Patil-Shinde et al., 2014). The

output prediction accuracies of the models were indicated by correlation coefficients. The correlation coefficients were lying between 0.92 and 0.996. The authors have claimed that the prediction accuracy of genetic programming model has an advantage over the multilayer perceptron neural network.

The remaining part of the paper is organised as follows. Section 2 gives an overview of the genetic programming based modelling. Section 3 illustrates the simulation results and a comparison with the single gene genetic programming variant. The paper ends in section 4 with the conclusions followed by the references.

2. Method of genetic programming modelling

Genetic programming is an evolutionary approach which automatically evolves computer programs to solve the problem without specifying the structure of the solution in advance (Koza, 1992; Poli et al., 2008). Genetic programming is a branch of evolutionary algorithms and can be used for development of nonlinear mathematical models based on input-output training datasets. Genetic programming is based on the Darwinian principle of natural selection and survival of the fittest. The main advantage of the genetic programming formalism is that it automatically evolves an empirical mathematical model from the input- output datasets. Hence, the genetic programming modelling process does not require the detailed information of process phenomena. The genetic programming technique has been used for symbolic nonlinear regression problems to develop mathematical expressions that provide a good fit between a given set of independent variables and the associated dependent variables (Pan et al., 2013).

Over the years, several models such as thermodymic equilibrium model, kinetics rate model, pseudo equilibrium model, Aspen Plus based model and artificial intelligence based models have been developed for simulating a gasification system. Among those models, thermodynamic models are the simplest and mostly used for sensitivity analysis of the process parameters. The equilibrium models are independent of gasifier design but their prediction accuracy is not good in all cases. The kinetic rate models are more accurate and computationally intensive. However, their applicability is limited to specific plants (Puig-Arnavat et al., 2010). Most of the chemical process simulation tools like Aspen Plus which build up a model from the first principle's approach have a higher computational burden. A typical Aspen Plus model which includes the different stages of gasification like drying, devolatilization, gasification and combustion would

take a simulation time of the order of a couple of minutes. However, the problem is compounded in a situation which requires multiple function calls to the model (e.g. incorporating an optimisation algorithm with the Aspen model in the loop). In such cases if say there are a thousand function calls for one optimisation run, then the total computational time would be in the order of days. The multi-gene genetic programming method circumvents this problem since it is an explicit mathematical expression which can be calculated within milli-seconds on a personal computer.

Another advantage of the multi-gene genetic programming method over the other models is that it is a data-driven methodology which relies on experimental data to build models. This eliminates errors due to various assumptions (e.g. considering the reactors as point masses and neglecting their spatial effects, assuming perfect insulation and neglecting heat losses while modelling etc.) which are often employed while building up a differential equation based model.

The genetic programming starts with a high-level statement of the problem and attempts to invent a computer program to solve the problem. The evolved solution variables are represented in the form of genes or trees. At the beginning of the algorithm, the genes or trees are initiated randomly. To accomplish the best possible fitness function, the genes or trees undergo reproduction, crossover and mutation processes. Crossover involves the mutual interchange of genetic material between the parents to form new offspring. Mutation, refer to random change within the gene i.e. a randomly chosen element is replaced by another element. The nodes of the genetic programming tree are called operator nodes and operand nodes. The operator nodes represent mathematical operators such as addition, subtraction, division, multiplication, etc. while operand nodes define the input variables (x_i). The symbolic regression of multi-gene genetic programming is a weighted linear combination of several gene outputs. The multi-gene genetic programming methodology has been used for predicting the toxicity of chemical compounds (Searson et al., 2010). It was concluded that the multi-gene genetic programming model offers an alternative approach to currently accepted empirical modelling and data analysis techniques. The uniqueness of the multi-gene genetic programming based model is that it automatically evolves a mathematical expression in a symbolic form which can be analysed further to find which variables impact the final prediction and in what fashion. Figure 1 is a tree

representation of a multi-gene genetic programming with output (y) and input variables (x_1, x_2, x_3). The mutation and crossover operations are also shown diagrammatically in Figure 1.

For the genetic programming simulation, there are nine process parameters that have been used as model inputs i.e. carbon (x_1,wt%), hydrogen (x_2, wt%), nitrogen (x_3, wt%), sulphur (x_4, wt%), oxygen (x_5, wt%), moisture content (x_6, wt%), ash (x_7, wt%), equivalence ratio (x_8, ER) and the temperature of the gasifier (x_9, T^0C). ER is defined as the ratio of actual air to fuel ratio versus stoichiometric air to fuel ratio for complete combustion. The input parameters are represented as an input vector $\hat{x}_i = [x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9]$ and the output variables are lower heating value (kJ/Nm^3) (y_1) and gas yield (Nm^3/kg) (y_2).

Two different models are presented to predict the lower heating value and gas yield from the biomass composition and process parameters. The experimental data was obtained from literature for wood, paper, kitchen garbage, polyethylene plastic and textile (Xiao et al., 2009) and for Hongkong municipal solid waste (Choy et al., 2004). To check the accuracies and robustness of the model, the experimental dataset is divided for training and testing purposes. From the available data 70% (47 data points) is randomly selected to use for training purposes and the remaining 30% (20 data points) is used for model validation. The performance of the model is compared with experimental data reported by the authors. The input and output datasets are normalized. The mean value of the nine input variable $(\hat{x}_i)$ and output variables (y_1, y_2) are represented by the vector μ_x and μ_y, respectively and is given in equations (1) and (2).

$$\mu_x = [43.815 \quad 5.11 \quad 0.685 \quad 0.17 \quad 36.53 \quad 4.21 \quad 9.55 \quad 0.4 \quad 581] \tag{1}$$

$$\mu_y = [3153 \quad 2.86] \tag{2}$$

Similarly, their corresponding standard deviations are given by σ_x and σ_y in equations (3) and (4).

$$\sigma_x = [0.1202 \quad 0.6929 \quad 0.5868 \quad 0.1838 \quad 6.2649 \quad 5.9538 \quad 10.6773 \quad 0.2828 \quad 154.1493] \tag{3}$$

$$\sigma_y = [835.8002 \quad 2.616295] \tag{4}$$

2.1 Parameter setting of genetic programming algorithm

For the genetic programming simulations, the population size and the maximum number of generations are set as 100 and 1000 respectively. For selecting the parent genes from the pool of available solutions, a tournament selection strategy is adopted. The tournament size is set to 3. The maximum depth of each tree in the multi-gene representation is set to 5 to allow some control over the complexity of the developed expressions. The set of instructions or functions used for symbolic regression are $\{+,-,\times,\div,\sin,\cos,(.)^2,\exp,\log\}$. The crossover, mutation and direct reproduction probabilities are taken as 0.85, 0.1 and 0.05 respectively. The multi-gene genetic programming simulations are compared with a single-gene genetic programming algorithm. The details of both genetic programming variants are given in Table 1. For single-gene genetic programming the number of trees (*T*) is set to one.

3. Results and Discussion

3.1 Multi-gene genetic programming based model for lower heating value calculation

Figure 2 shows the final population of the genetic programming run for lower heating value, presenting the trade-off between the precision of the fit and the complexity of the evolved multi-gene genetic programming solutions. The optimal evolved models fall on the curve of a non-dominated solution called the Pareto front. The blue dots represent the set of dominated solutions and those in green are the set of non-dominated solutions on the Pareto front. From the Pareto front, user can decide whether the incremental gain in performance is worth with associated model complexity. Three solutions A, B, and C on the Pareto front are selected based on their corresponding accuracies and model complexities. The solutions are indicated by arrows in the top right Figure 2.

Figure 2 also depicts the convergence characteristics of the genetic programming algorithm. It is evident that the mean fitness of the curve becomes smoother after 500 generations and that the change in objective function is not significant near the end of the genetic programming run. It indicates that running the genetic programming for more generations does not result in a more favourable outcome. However, as the best fitness is reported at 971 in this particular case, it suggests that the genetic programming algorithm should have to run for at least 1000 generations.

Figure 3 show twelve subplots (a1-c2) which indicates the correlation coefficient (R^2) and root mean squared error (RMSE) on the training and testing datasets. The subplots are

representing the number of data points used in this study for predicting the output variables $(y_1 - y_2)$. Solution A in Figure 3 (a1) shows the best fit for the lower heating value calculations amongst the other solutions on the Pareto front, but it also involves the highest model complexity as compared to the other solutions B and C. The evolved regression equation for Solution A is given by (5).

$$y_1 = 0.4047\sqrt{\left(x_4\cos\left(x_4\right)\right)}\left(x_9 - x_5\right) - 3.937\cos\left(\cos\left(x_5\cos\left(x_4\right)\right)\right)$$
$$+0.2127\sqrt{\mathrm{e}^{x_9}}\left(x_4 - x_8 x_4^{\,2}\right) + 0.289 x_4 e^{x_8}\cos\left(x_8^{\,1/4}\right) + 0.1625\left(x_2 - x_8\right)^2\left(\cos\left(x_6\right)\right)^2 x_5^{\,2}\left(x_2 + x_6\right) + 2.717 \tag{5}$$

The regression equation of Solution B (slightly less complex) as evolved by the genetic programming algorithm is given by (6).

$$y_1 = 0.1565\sqrt{\mathrm{e}^{x_9}}\left(x_5 - x_8 x_4^{\,2}\right) - 4.149\cos\left(\cos\left(x_5\cos\left(x_4\right)\right)\right) - 0.2079\log\left(\sin\left(x_6\right)\right)$$
$$+0.2297\sqrt{\mathrm{e}^{x_9}}\left(x_4 - \cos\left(x_4\right)x_4^{\,2}\right) + 0.1259\left(x_2 - x_8\right)^2 x_4 x_5^{\,2}\left(x_2 + x_6\right) + 2.459 \tag{6}$$

On the other hand, the models of solution B and solution C are less complex but at the same time their prediction capability is poorer as compared to solution A. Nevertheless, the prediction capability of model B is more than 95% and 57% on the training and testing datasets respectively. However, the prediction accuracy of model C is over 80% on the training dataset and just over 33% on the validation set suggesting that solution C has been over simplified. This clearly indicates that the decrease in model complexity not only deteriorates the model prediction capability on the training data but also the model does not fit well to unseen data.

The regression equation for Solution C (the simplest expression) as evolved by the genetic programming algorithm is given by (7).

$$y_1 = 0.1862\left(x_2 - x_8\right)^2\left(\cos\left(x_6\right)\right)^2 x_5^{\,2}\left(x_2 + x_6\right) - 6.472\cos\left(\cos\left(x_5\cos\left(x_4\right)\right)\right) + 4.723 \tag{7}$$

The prediction accuracies and generalised performance of each model is tested between the experimental data and model predicted values. The model that performed the best on training and testing data is selected on the basis of R^2 and RMSE. Figure 4 (a1-c1) show the multi-gene genetic programming model predicted versus actual data and their R^2 values on the training and testing data sets for the three solutions A, B and C respectively. It can be observed that Solution A has less than 3% error in prediction of training dataset. Also, the evolved model shows good agreement with the testing data. The correlation coefficient and root mean squared error of the

multi-gene genetic programming based model for lower heating value prediction are listed in Table 2.

As a summary, solution A has the best fitness over solution B and C but is more complex as well. The prediction accuracy of solution A and B for the performance variable (y_1) shows good agreement with their experimental counterparts on the training data set. It demonstrates that multi-gene genetic programming can be exploited to simulate the complex thermochemical processes such as FB gasifiers.

3.2 Multi-gene genetic programming based model for Syngas yield production

In this model, the same strategy is used as explained in the previous case. Based on the complexity and fitness three different solutions are selected as shown in Figure 2 (bottom). As explained in section 4.1, it can be seen from the Figure 2 (bottom) that the mean fitness of the convergence curve is not changing significantly after 600 generations. On the other hand the best fitness of the genetic programming algorithm is reported at 991 generations.

Figure 3 (a2-c2) show the R^2 and RMSE on training and testing dataset for solution A, B and C respectively. It can be observed that solution A shows an excellent predictive ability in the training data; however, the prediction capability is poorer over the unseen data (validation). It indicates that model A is suffering from the over-fitting problem of the training data set. On the other hand, solution B and C have very good prediction accuracies on both the training and validation data sets. The regression equation of Solution A as evolved by the genetic programming algorithm for the Syngas yield prediction is given by (8).

$$\begin{aligned} y_2 &= 0.8001x_1 + 0.1986x_3 + 0.001104x_4 + 0.799x_8 + 0.2354\sin(x_4 - x_5) + 0.2354\cos\left(\frac{x_1 + x_8}{(x_4)^2}\right) + \\ &0.2732x_8\left(\log(x_7) + \frac{x_2}{x_4 - x_5}\right) - 0.05691(0.2266x_8 - 0.2266x_9)(x_8 + 4.382) - \\ &\frac{0.05691\left(\log(x_9) - (x_4)^2\right)}{2x_3 + 4.375} - \frac{0.001104(x_3 - x_4)}{x_8} - \frac{0.1986(x_4 - x_7)(x_4 - x_8)}{x_8 + 4.3} - \\ &\frac{0.001104(x_8 + 0.1765)(x_9 + 0.1765)}{(x_3 - x_4)^2} - 0.01526 \end{aligned} \tag{8}$$

The regression equation of Solution B as evolved by the genetic programming algorithm for gas yield prediction is given by(9).

$$y_2 = 0.7994x_1 + 0.1956x_3 + 0.7994x_8 + 0.2352\sin(x_4 - x_5) + 0.2352\cos\left(\frac{x_1 + x_8}{(x_4)^2}\right)$$
$$+0.266x_8\left(\log(x_7) + \frac{x_2}{x_4 - x_5}\right) - 0.06401(0.2266x_8 - 0.2266x_9)(x_8 + 4.382) - \tag{9}$$
$$0.008318x_8\left(\log(x_7) - \frac{x_2}{x_8}\right) - \frac{0.06401\left(\log(x_9) - (x_4)^2\right)}{2x_3 + 4.375} - \frac{0.1956(x_4 - x_7)(x_4 - x_8)}{x_8 + 4.3} - 0.01241$$

The regression equation of Solution C as evolved by the genetic programming algorithm for gas yield prediction is given by(10).

$$y_2 = 0.7784x_1 + 0.1748x_3 + 0.7784x_8 + 0.229\sin(x_4 - x_5) + 0.229\cos\left(\frac{x_1 + x_8}{(x_4)^2}\right)$$
$$+0.277x_8\left(\log(x_7) + \frac{x_2}{x_4 - x_5}\right) + \frac{0.02553x_2(x_4 - x_8)}{x_8 + 4.3} - \frac{0.2003(x_4 - x_7)(x_4 - x_8)}{x_8 + 4.3} - 0.004875 \tag{10}$$

Concisely, the multi-gene genetic programming paradigm evolves multiple models which provide more number of choices to the designer. A single model can be selected based on the application requirements. It is evident that solution A (Figure 3 a2) suffers from the over-fitting problem. Hence, the selection of the model must be carried out on the basis of the application. In cases where the model is used for predicting the data which is already present in the training data set, using model A is advisable, as it gives a better prediction on the training datasets. On the other hand, if the model is used to predict the syngas yield for untrained datasets then solution B or C may be used, since they show a better prediction capability on unseen (test) or untrained datasets. Figure 4 (a2-c2) shows the prediction accuracy of the multi-gene genetic programming model for syngas yield production. It is noticed that the multi-gene genetic programming based model for syngas yield production shows slightly better accuracy to that possessed by the multi-gene genetic programming model for lower heating value prediction.

3.3 Multi-gene genetic programming algorithm and comparison with single-gene genetic programming Model

This section focuses on comparing the results obtained using multi-gene genetic programming algorithm with single-gene genetic programming. It is well known that the back propagation artificial neural networks techniques often get stuck in local minima without having learned the entire dataset (Gori and Tesi, 1992). However, the genetic programming technique uses an evolutionary approach in which the model exchanges their information through mutation and crossover. This characteristic helps genetic programming to converge at global minima and get out of local minima.

3.3.1 Comparison of multi-gene genetic programming and single-gene genetic programming model for lower heating value

Figure 5 shows the Pareto plot of single-gene genetic programming depicting the trade-off between fitness *vs.* complexity. The convergence characteristic of single-gene genetic programming algorithm is shown in Figure 5 (top) for predicting the lower heating value of the syngas. The actual *vs* predicted value of lower heating value and Syngas yield production from the best single-gene genetic programming based solution on training and testing data are reported in Figure 6 (top). In the case of lower heating value prediction, single-gene genetic programming algorithm is fitting over 83% and 34% on training and testing datasets. The best single-gene genetic programming based regression equation (as indicated by the red circle in Figure 5 (top)) for lower heating value prediction is (11).

$$y_1 = 0.1925x_9 - 0.1925x_6 + 0.1925\left(\frac{x_4}{x_3}\right)^2 + 0.1925e^{\frac{x_2 x_8}{x_5 - x_6}} - 0.1925x_7 x_8 - 0.6675 \qquad (11)$$

The RMSE and R^2 of the multi-gene genetic programming and the single-gene genetic programming solutions are reported in Table 2. The RMSE and R^2 values reported in Table 2 are useful while selecting the model. For predicting the lower heating value, the mean fitness and model complexity of the single-gene genetic programming algorithm reported in Figure 5 (top) are poor compared to the multi-gene genetic programming algorithm in Figure 2 (top) for the best case. It is evident that the percentage fit of the single-gene genetic programming based solution is not better than multi-gene genetic programming based solutions A and B. However, it is worth mentioning that the best single-gene genetic programming algorithm shows better prediction capability over solution C. This shows that decreasing the model complexity lessens the model prediction capability.

Apart from the statistics of the two genetic programming variants, the mean and standard deviation of the best solution for 30 independent runs are reported in Table 3. The mean and standard deviation of the best solutions clearly show that the multi-gene genetic programming algorithm consistently converges to a better fit than the single-gene genetic programming.

3.3.2 Comparison of multi-gene genetic programming and single-gene genetic programming model for Syngas yield production

As explained in the section 4.3.1, a similar trend is observed in the case of yield prediction as well. The Pareto front of the evolved single-gene genetic programming solution showing the fitness and accuracy of the model is given in Figure 5 (bottom). Figure 6 (bottom) shows that the single-gene genetic programming model is in good agreement with training and validation data. The percentage fit of single-gene genetic programming prediction for training and testing data is close to 98% and 95% respectively. However, the comparison clearly shows that the multi-gene genetic programming evolved solutions are more accurate and can be applied more generally over the single-gene genetic programming solutions. Similarly, the mean fitness of the convergence curve single-gene genetic programming algorithm in Figure 5 (bottom) is not better than the mean fitness reported in Figure 2 (bottom). The single-gene genetic programming based regression equation for gas yield prediction is (12),

$$y_2 = 0.1638x_2 + 0.1638x_8 + 0.1638\sin(x_2 + x_8) + 0.1638\log(x_4) + 0.1638e^{x_1}(x_1 + x_4 + x_8) - 0.06335 \tag{12}$$

The comparative analysis of three different multi gene genetic programming solutions with single-gene genetic programming is reported in Table 2. It can be seen from the Table 2 that multi-gene genetic programming based algorithms show better fitness over single-gene genetic programming based formalism on both training and testing datasets. It is reported that the R^2 value of single-gene genetic programming model on the testing data is over 94% compared to Solution A. This indicates that the Solution A is over-fitted in the training phase. However, multi-gene genetic programming based Solutions B and C show improved fitness on the both training and validation dataset compared to single-gene genetic programming algorithm. Hence, R^2 and RMSE of the training and validation (testing) dataset will be helpful while choosing the best model for the prediction.

The mean and standard deviation of the best solution of single-gene genetic programming and multi-gene genetic programming algorithm for 30 independent runs are reported in Table 3. The mean and standard deviation of the best solutions clearly show that the multi-gene genetic programming algorithm consistently converges to a better fit than the single-gene genetic programming.

4. Conclusion

In this study, a multi-gene genetic programming based mathematical model is developed to predict the gas yield and lower heating value of the syngas produced from the wastes during FB gasification using their physio-chemical characteristics and a few process parameters. The multi-gene genetic programming models shows better performance ($R^2 > 97\%$ for lower heating value and 99.8% for Syngas yield production) over the single-gene genetic programming model ($R^2 \approx 83\%$ for lower heating value and 97.9% for Syngas yield production). The accuracies of the predicted values using the multi-gene genetic programming approach are in good agreement with experimental data. The results of this work are encouraging and will be used to model other similar gasification processes.

Acknowledgement

The research leading to these results has received funding from the People Programme (Marie Curie Actions) of the European Union's Seventh Framework Programme FP7/2007-2013/ under REA grant agreement n° [289887].

References

Arena, U., 2012. Process and technological aspects of municipal solid waste gasification. A review. Waste Management 32, 625–639.

Brown, D., Fuchino, T., Maréchal, F., 2006. Solid fuel decomposition modelling for the design of biomass gasification systems. Computer Aided Chemical Engineering 21, 1661–1666.

Brown, D.W., Fuchino, T., MA Maréchal, F., 2007. Stoichiometric equilibrium modelling of biomass gasification: Validation of artificial neural network temperature difference parameter regressions. Journal of Chemical Engineering of Japan 40, 244–254.

Chavan, P., Sharma, T., Mall, B., Rajurkar, B., Tambe, S., Sharma, B., Kulkarni, B., 2012. Development of data-driven models for fluidized-bed coal gasification process. Fuel 93, 44–51.

Cheng, H., Hu, Y., 2010. Municipal solid waste (MSW) as a renewable source of energy: Current and future practices in China. Bioresource Technology 101, 3816–3824.

Choy, K.K., Porter, J.F., Hui, C.-W., McKay, G., 2004. Process design and feasibility study for small scale MSW gasification. Chemical Engineering Journal 105, 31–41.

Dong, C., Jin, B., Li, D., 2003. Predicting the heating value of MSW with a feed forward neural network. Waste Management 23, 103–106.

Gómez-Barea, A., Leckner, B., 2010. Modeling of biomass gasification in fluidized bed. Progress in Energy and Combustion Science 36, 444–509.

Gori, M., Tesi, A., 1992. On the problem of local minima in backpropagation. IEEE Transactions on Pattern Analysis and Machine Intelligence 14, 76–86.

Guerrero, L.A., Maas, G., Hogland, W., 2013. Solid waste management challenges for cities in developing countries. Waste Management 33, 220–232.

Guo, B., Li, D., Cheng, C., Lü, Z., Shen, Y., 2001. Simulation of biomass gasification with a hybrid neural network model. Bioresource Technology 76, 77–83.

Guo, B., Shen, Y., Li, D., Zhao, F., 1997. Modelling coal gasification with a hybrid neural network. Fuel 76, 1159–1164.

Higman, C., Van der Burgt, M., 2011. Gasification. Gulf professional publishing.

Kalogirou, S.A., 2001. Artificial neural networks in renewable energy systems applications: a review. Renewable and Sustainable Energy Reviews 5, 373–401.

Koza, J.R., 1992. Genetic programming: on the programming of computers by means of natural selection. MIT press.

Malkow, T., 2004. Novel and innovative pyrolysis and gasification technologies for energy efficient and environmentally sound MSW disposal. Waste Management 24, 53–79.

Murphy, J.D., McKeogh, E., 2004. Technical, economic and environmental analysis of energy production from municipal solid waste. RenewableEenergy 29, 1043–1057.

Narayana, T., 2009. Municipal solid waste management in India: From waste disposal to recovery of resources? Waste Management 29, 1163–1166.

Pan, I., Pandey, D.S., Das, S., 2013. Global solar irradiation prediction using a multi-gene genetic programming approach. Journal of Renewable and Sustainable Energy 5, 063129.

Patil-Shinde, V., Kulkarni, T., Kulkarni, R., Chavan, P.D., Sharma, T., Sharma, B.K., Tambe, S.S., Kulkarni, B.D., 2014. Artificial Intelligence based Modelling of High Ash Coal Gasification in a Pilot-plant Scale Fluidized Bed Gasifier. Industrial & Engineering Chemistry Research.

Pires, A., Martinho, G., Chang, N.-B., 2011. Solid waste management in European countries: A review of systems analysis techniques. Journal of Environmental Management 92, 1033–1050.

Poli, R., Langdon, W., McPhee, N., 2008. A field guide to genetic programming. (Lulu Enterprises UK Ltd.).

Puig-Arnavat, M., Bruno, J.C., Coronas, A., 2010. Review and analysis of biomass gasification models. Renewable and Sustainable Energy Reviews 14, 2841–2851.

Puig-Arnavat, M., Hernández, J.A., Bruno, J.C., Coronas, A., 2013. Artificial neural network models for biomass gasification in fluidized bed gasifiers. Biomass and Bioenergy 49, 279–289.

Searson, D.P., Leahy, D.E., Willis, M.J., 2010. GPTIPS: an open source genetic programming toolbox for multigene symbolic regression, in: Proceedings of the International Multiconference of Engineers and Computer Scientists.

Sridevi, K., Sivaraman, E., Mullai, P., 2014. Back propagation neural network modelling of biodegradation and fermentative biohydrogen production using distillery wastewater in a hybrid upflow anaerobic sludge blanket reactor. Bioresource technology 165, 233–240.

Xiao, G., Jin, B., Zhong, Z., Chi, Y., Ni, M., Cen, K., Xiao, R., Huang, Y., Huang, H., 2007. Experimental study on MSW gasification and melting technology. Journal of Environmental Sciences 19, 1398–1403.

Xiao, G., Ni, M., Chi, Y., Jin, B., Xiao, R., Zhong, Z., Huang, Y., 2009. Gasification characteristics of MSW and an ANN prediction model. Waste Management 29, 240–244.

List of figures

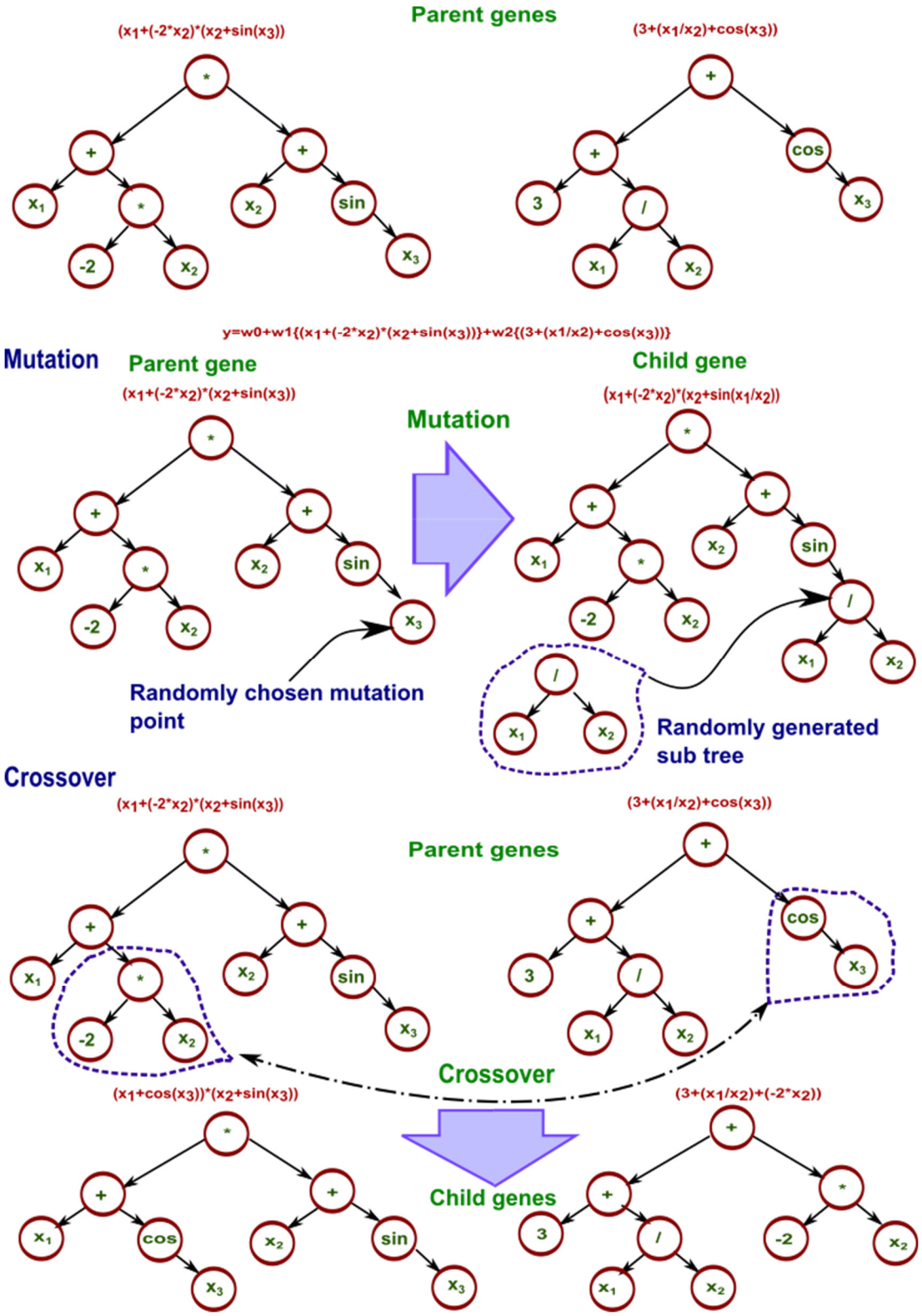

Figure 1. Schematic diagram of the tree structure of a multi-gene genetic programming model

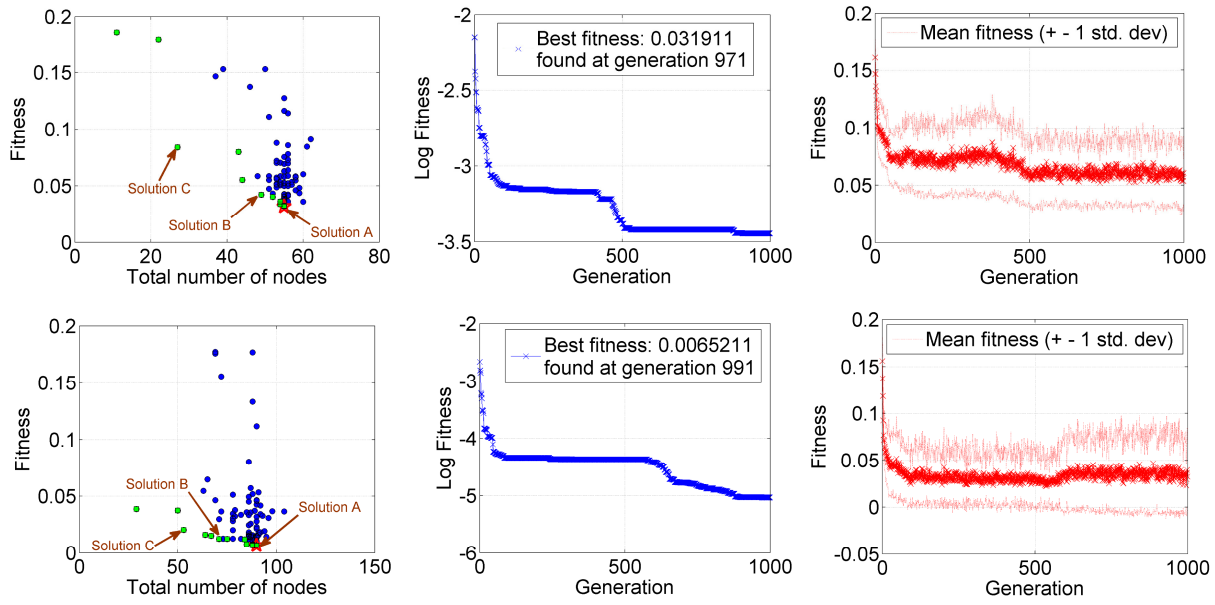

Figure 2. The Pareto front and convergence of the multi-gene genetic programming solutions for lower heating value calculation (top row) and Syngas yield production (bottom row).

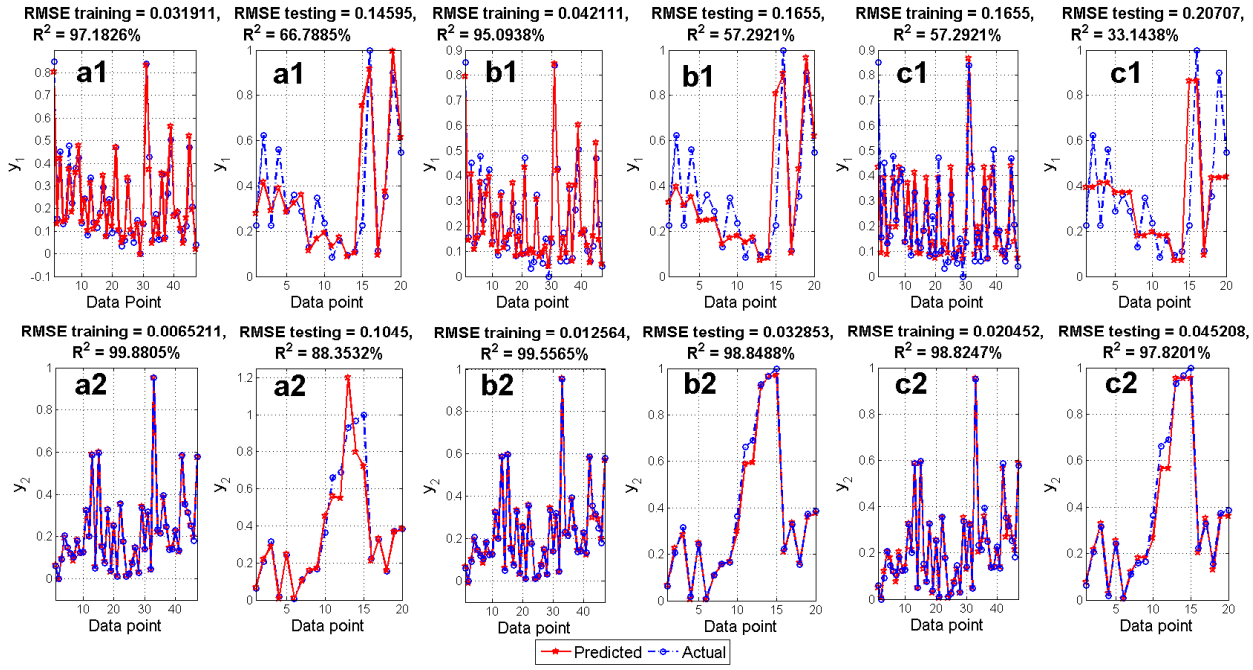

Figure 3. Prediction results of the multi-gene genetic programming solutions with R^2 and RMSE on the training and the testing datasets: lower heating value prediction Solution A (a1), Solution B (b1), Solution C (c1) and Syngas yield production Solution A (a2), Solution B (b2), Solution C (c2)

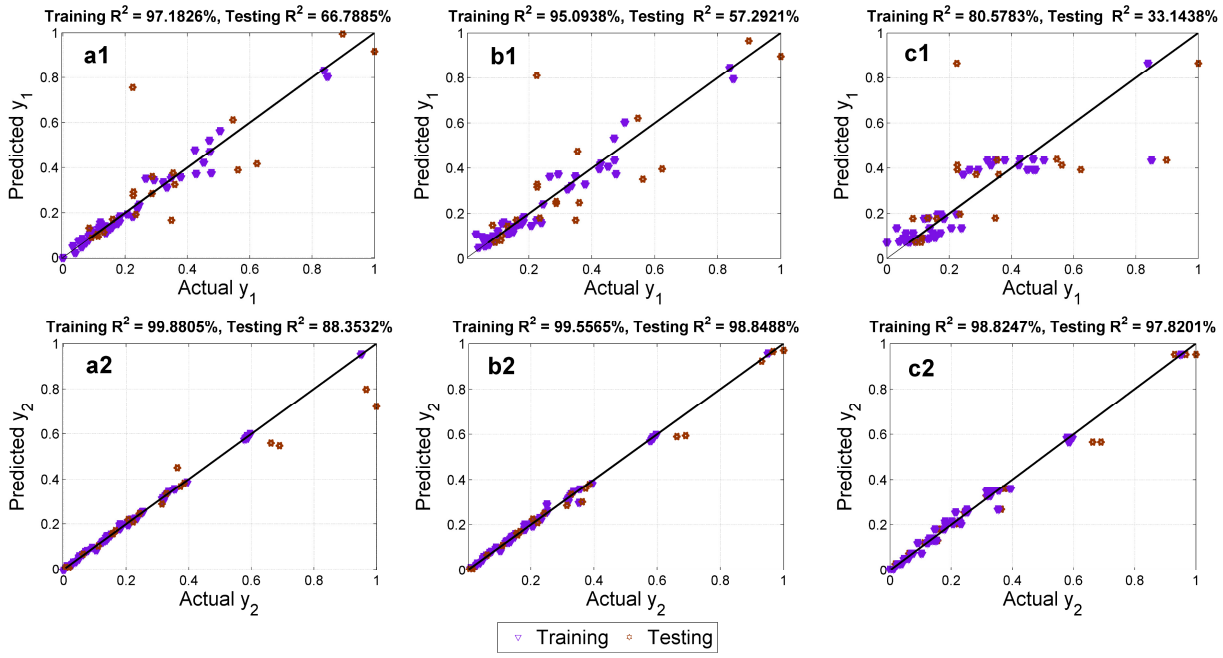

Figure 4. Prediction results of the multi-gene genetic programming solutions with R^2 on the training and the testing datasets: lower heating value prediction Solution A (a1), Solution B (b1), Solution C (c1) and Syngas yield production Solution A (a2), Solution B (b2), Solution C (c2)

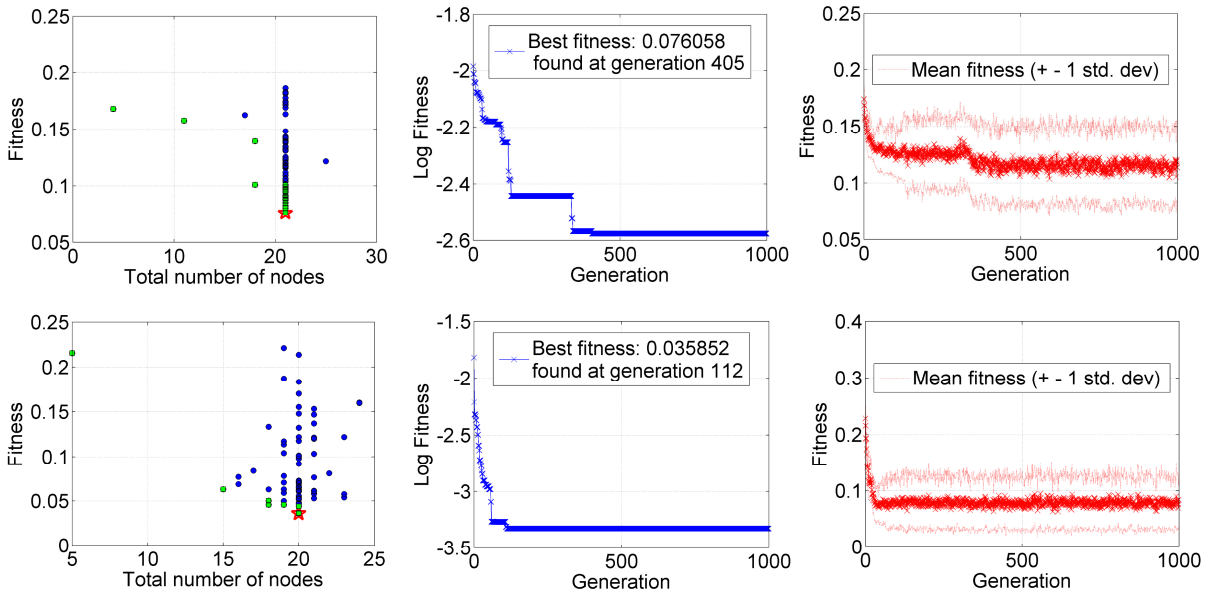

Figure 5. The Pareto front and convergence of the best single-gene genetic programming solution for lower heating value calculation (top row) and Syngas yield production (bottom row).

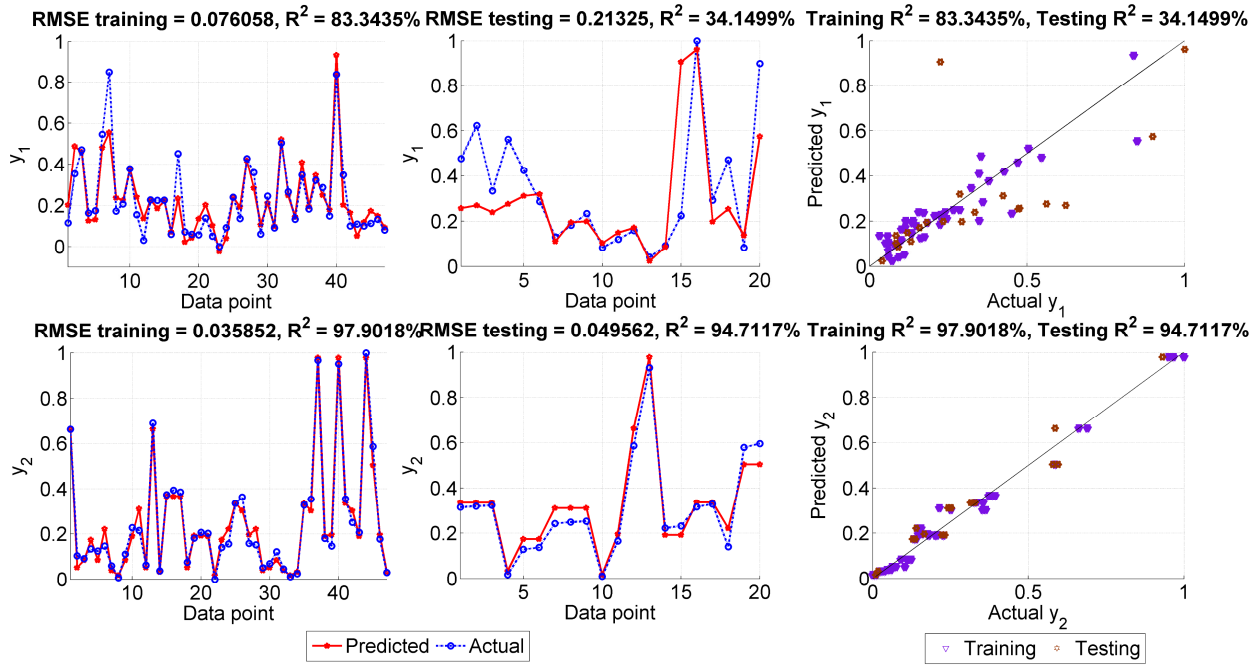

Figure 6. Prediction results of the best single-gene genetic programming solution with R^2 and RMSE on the training and the testing datasets: for the lower heating value (top) and Syngas yield production (bottom).

List of tables:

Table 1: Parameter settings for the multi-gene genetic programming and single-gene genetic programming variants.

GP Algorithm parameters	Parameter settings
Population size	100
Number of generation	1000
Selection method	Plain lexicographic tournament selection
Tournament size	3
Termination criteria	1000 generation or fitness value less than 0.00001 whichever is earlier
Maximum depth of tree	5
Maximum number of trees in an individual (for MGGP only)	15
Mathematical operations	$\{+,-,\times,\div,\sin,\cos,(.)^2,\exp,\log\}$

Table 2. Statistics of the best solutions of two genetic programming variants for lower heating value prediction and Syngas yield production.

Goodness of Fit	**MGGP**						**SGGP**	
	Solution A		**Solution B**		**Solution C**		**Best solution**	
	Training	**Testing**	**Training**	**Testing**	**Training**	**Testing**	**Training**	**Testing**
GP variants for LHV prediction								
RMSE	0.03191	0.14595	0.04211	0.1655	0.83785	0.20707	0.07605	0.21324
R^2 (%)	97.1826	66.7885	95.0938	57.2921	80.5783	33.1438	83.3435	34.1499
GP variant for Syngas yield production								
RMSE	0.00652	0.1045	0.01256	0.03285	0.02045	0.04520	0.03585	0.04956
R^2 (%)	99.8805	88.3532	99.5565	98.8488	98.8247	97.8201	97.9018	94.7117

Table 3. Statistics of the best solution for 30 independent runs of two genetic programming variants for lower heating value prediction and Syngas yield production.

Algorithm	**Mean (μ)**	**Standard deviation (σ)**	**Minimum**
GP variants for LHV prediction			
MGGP	0.050605	0.010224	0.031911
SGGP	0.116638	0.025553	0.076058
GP variants for Syngas yield production			
MGGP	0.013192	0.00331	0.006521
SGGP	0.041831	0.015197	0.020957