A novel sensitivity study for biomass directional devolatilization by random forest models

Jiangkuan Xing^a, Kun Luo^a, Haiou Wang^a, Tai Jin^b, Jianren Fan^{a,*}

^aState Key Laboratory of Clean Energy Utilization, Zhejiang University, Hangzhou, 310027, China ^bSchool of Aeronautics and Astronautics, Zhejiang University, Hangzhou 310027, China

Supporting Information

1. Biomass devolatilization detailed mechanism

In the present study, we construct the devolatilization products database with a multi-step detailed mechanism (as shown in Table S1) proposed by Ranzi et al [1]. In this mechanism, five biomass chemical compositions are set as the initial reactants, and there are 19 reactions in total involving 36 species. Mass fractions of $C_2H_4O_2$, $C_2H_2O_2$, C_3H_6O , $C_3H_4O_2$, $C_6H_6O_3$, $C_6H_{10}O_5$, $C_5H_8O_4$, $C_9H_{10}O_2$, C_6H_6O , and $C_{11}H_{12}O_4$ are summed to obtain the tar fraction. The solid fraction is calculated through summing the mass fractions of CELL, CELLA, HCE, HCE1, HCE2, LIG, LIG-C, LIG-H, LIG-O, LIG-OH, LIG-CC, G{CO}, G{H_2}, G{CO_2}, G{COH_2} and Char. A and E are the pre-exponential factor and activation energy of the reaction.

2. Details of biomass components determination

There are two possible ways can be used. The first one is to use a linear rule based on the conservation of the elemental C/H/O composition from the study of Faravelli et al. [2], in which three reference components (LIG-C, LIG-H and LIG-O) have their C/H/O compositions, therefore mass fraction of each LIG component can be obtained through solving the following equations,

$$X_{LIG-C}f^C_{LIG-C,ref} + X_{LIG-H}f^C_{LIG-H,ref} + X_{LIG-O}f^C_{LIG-O,ref} = X_C$$
(1)

$$X_{LIG-C}f^H_{LIG-C,ref} + X_{LIG-H}f^H_{LIG-H,ref} + X_{LIG-O}f^H_{LIG-O,ref} = X_H$$

$$\tag{2}$$

$$X_{LIG-C}f^{O}_{LIG-C,ref} + X_{LIG-H}f^{O}_{LIG-H,ref} + X_{LIG-O}f^{O}_{LIG-O,ref} = X_{O}$$
(3)

*Corresponding author: Email address: fanjr@zju.edu.cn (Jianren Fan)

No	Reaction pathway	A (1/s)	E (kJ/mol)
1	$CELL \rightarrow CELLA$	2.80×10^{19}	242.4
2	$CELLA \rightarrow 0.95C_{2}H_{4}O_{2} + 0.25C_{2}H_{2}O_{2} + 0.2CH_{3}CHO + 0.25C_{6}H_{6}O_{3} + 0.2C_{3}H_{6}O + 0.16CO_{2} + 0.23CO + 0.9H_{2}O + 0.1CH_{4} + 0.61Char$	1.30×10^{10}	150.5
3	CELLA $\rightarrow C_6 H_{10} O_5$	3.28×10^{14}	196.5
4	$CELL \rightarrow 5H_2O + 6Char$	8.00×10^7	133.9
5	$\text{HCE} \rightarrow 0.4 \text{HCE1} + 0.6 \text{HCE2}$	2.10×10^{16}	186.7
6	$\begin{aligned} \mathrm{HCE1} &\rightarrow 0.75\mathrm{G}\{\mathrm{H}_2\} + 0.8\mathrm{CO}_2 + 1.4\mathrm{CO} + 0.5\mathrm{CH}_2\mathrm{O} + 0.25\mathrm{CH}_3\mathrm{OH} \\ &+ 0.125\mathrm{C}_2\mathrm{H}_6\mathrm{O} + 0.125\mathrm{H}_2\mathrm{O} + 0.625\mathrm{CH}_4 + 0.25\mathrm{C}_2\mathrm{H}_4 + 0.675\mathrm{Char} \end{aligned}$	2.60×10^{11}	145.7
7	$\rm HCE1 \rightarrow C_5 H_8 O_4$	$8.75 imes 10^{15}$	202.4
8	$HCE2 \rightarrow 0.2CO_2 + 0.5CH_4 + 0.25C_2H_4 + 0.8G\{CO_2\} + 0.8G\{COH_2\}$ + 0.7CH_O + 0.25CH_OH + 0.125C_2H_2O + 0.125H_2O + Char	1.00×10^{10}	138.1
9	$+ 0.7CH_{2}O + 0.25CH_{3}OH + 0.125C_{2}H_{6}O + 0.125H_{2}O + Char$ LIG-C $\rightarrow 0.35$ LIGCC + 0.1C ₉ H ₁ 0O ₂ + 0.08C ₆ H ₆ O + 0.41C ₂ H4 + H ₂ O + 0.495CH ₄ + 0.32CO + G{COH ₂ } + 5.735Char	4.00×10^{15}	202.9
10	$\rm LIG-H \rightarrow \rm LIGOH + \rm C_3H_6O$	$2.00 \times \ 10^{13}$	156.9
11	$\rm LIG-O \rightarrow \rm LIGOH + \rm CO2$	1.00×10^9	106.7
12	$LIGCC \rightarrow 0.3C_9H_{10}O_2 + 0.2C_6H_6O + 0.35C3H_4O_2 + 0.7H_2O + 0.65CH_4 + 0.6C_2H_4 + G\{COH_2\} + 0.8G\{CO\} + 6.4Char$	5.00×10^{6}	131.8
13	$LIGOH \rightarrow LIG + H_2O + CH_3OH + 0.45CH_4 + 0.2C_2H_4 + 1.4G\{CO\} + 0.6GCOH_2 + 0.1GH_2 + 4.15Char$	3.00×10^8	125.5
14	$LIG \rightarrow C_{11}H_{12}O_4$	1.50×10^{9}	143.8
15	$LIG \rightarrow H_2O + 0.5CO + 0.2CH_2O + 0.4CH_3OH + 0.2CH_3CHO + 0.2C_3H_6O + 0.6CH_4 + G\{CO\} + 0.5G\{COH_2\} + 5.5Char$	7.70×10^6	111.4
16	$G{CO_2} \rightarrow CO_2$	1.00×10^{6}	100.4
17	$G\{CO\} \rightarrow CO$	$5.00 imes 10^{12}$	209.2
18	$G\{COH_2\} \rightarrow CO + H_2$	1.50×10^{12}	297.1
19	$G{H2} \rightarrow H_2$	5.00×10^{11}	313.8

Table S1: Reaction pathway and kinetic parameters of the detailed biomass pyrolysis mechanism.



Figure S1: Schematic diagram of the 2D interpolation. The solid dots represent the reference biomass library, and the hollow dot denotes the new biomass whose LIG components need to be determined. L1, L2 and L3 are the nearest three library biomass.

where X_{LIG-C} , X_{LIG-H} and X_{LIG-O} are the mass fraction of LIG component we want to obtained. $f_{LIG-C,ref}^{i}$, $f_{LIG-H,ref}^{i}$ and $f_{LIG-O,ref}^{i}$ are the already-known mass fraction of element *i* in the reference three components (*i*=C, H or O). X_C , X_H and X_O are the C/H/O of the lignin component. In this method, the elemental conservation ensures the sum of all determined components should be 100% (in a base of lignin mass), and the sum of mass fractions of all biomass components, including the hemicellulose, cellulose and lignin would not exceed 100%.

The second one is to use a 2D interpolation based on elemental data from the reference data corresponding to a library whose LIG components are available [3, 4] as shown in Fig. S1. The LIG components can be obtained by the following expressions,

$$r = \frac{S_{\triangle(L1,N,L3)}}{S_{\triangle(L1,L2,L3)}} \tag{4}$$

$$s = \frac{S_{\triangle(L1,L2,U)}}{S_{\triangle(L1,L2,L3)}} \tag{5}$$

$$\phi = (1 - r - s)\phi_{L1}^0 + r\phi_{L2}^0 + s\phi_{L3}^0 \tag{6}$$

where ϕ is the component fraction of the new biomass that needs to be determined. S_{Δ} is the triangular surface. ϕ^0 is the component fraction of the reference biomass. With this method, the mass sum of all components might exceed 100% if the unknown biomass exists in some extreme regions.

3. Comparisons with other approaches

We also test other two approaches, including the multi-variable regression (MVR, the degree ranges from 1 to 3) and artificial neural network (ANN). Their modelling results are then compared with the random forest results as shown in Fig. S2 (only some outputs are presented here for brevity) and the comparisons of the statistical indicators are listed in Table S2. It is noted that the hyper-parameter in the ANN model is the number of neurons in the hidden layer, which is also optimized using the trial-and-test method with the neurons number ranging from 1 to 300, and the determined optimal numbers are 97, 99, 93, 101, 98, 91, 99 and 107 for C_2H_4 , CH_4 , SR, CO, CO_2 , H_2 , H_2O and tar, respectively. As expected, the random forest models show advantages over the other two approaches with a higher R^2 , and therefore are selected for the following sensitivity analysis.

Models	Data set	C_2H_4	CH_4	СО	CO_2	H_2	H_2O	Tar	SR
DE	Train	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999
KF	Test	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999
	Train	0.980	0.982	0.979	0.994	0.990	0.973	0.993	0.991
ANN	Test	0.980	0.981	0.980	0.994	0.990	0.973	0.993	0.990
MVD 1	Train	0.718	0.725	0.728	0.836	0.702	0.725	0.891	0.742
M V K-1	Test	0.720	0.726	0.727	0.836	0.701	0.721	0.890	0.742
	Train	0.843	0.846	0.827	0.930	0.858	0.839	0.943	0.879
M V K-2	Test	0.842	0.844	0.826	0.929	0.859	0.839	0.944	0.877
MVD 9	Train	0.934	0.930	0.885	0.965	0.908	0.931	0.969	0.935
WIVK-3	Test	0.933	0.931	0.886	0.965	0.907	0.931	0.968	0.934

Table S2: Statistical indicators of the training and test performances of the MVR, ANN and RF models.

4. Relative importance of inputs on outputs with Gini importance method

The random forest model could also give us the relative importance ranking based on Gini importance method as shown in Fig. S3. The measured relative importance ranking is generally consistent with the Sobol's sensitivity analysis results but with different values due to their different implementations.



Figure S2: Comparisons between predictions of different models with the actual values for the entire database. From top to bottom, results of the RF, ANN, MVR-1, MVR-2 and MVR-3 are presented. From left to right, the results of CO, tar and SR are presented (only some components are shown for brevity). It is noted that in each sub-figure, the scatters are plotted with every 30 samples to decrease the figure size due to the huge number of samples in the entire database.



Figure S3: Measured relative importance of each input on each output using the random forest models.

References

- Ranzi, E., Cuoci, A., Faravelli, T., et al. Chemical kinetics of biomass pyrolysis. Energy fuels 2008, 22, 4292-4300.
- [2] Faravelli, T., Frassoldati, A., Migliavacca, G., et al. Detailed kinetic modeling of the thermal degradation of lignins. Biomass Bioenerg. 2010, 93, 60-71.
- [3] Ranzi, E.; Debiagi, P.; Frassoldati, A. Mathematical Modeling of Fast Biomass Pyrolysis and Bio-Oil Formation. Note I: Kinetic Mechanism of Biomass Pyrolysis. ACS Sustain Chem Eng 2017, 5, 2867-2881.
- [4] Jupudi, R.S.; Zamansky, V.; Fletcher, T.H. Prediction of Light Gas Composition in Coal Devolatilization. Energy fuels 2009, 23, 3063-3067.