Chemical Reaction Neural Network modelling of lignocellulosic biomass pyrolysis

Jiangkuan Xing[a,d,*], Ryoichi Kurose[a], Kun Luo[b,c], Jianren Fan[b,c]

[a] *Department of Mechanical Engineering and Science, Kyoto University, Kyoto, 615–8540, Japan*
[b] *State Key Laboratory of Clean Energy Utilization, Zhejiang University, Hangzhou, 310027, China*
[c] *Shanghai Institute for Advanced Study of Zhejiang University, Shanghai 200120, China*
[d] *JSPS International Research Fellow, Kyoto University, Japan*

Abstract

Lignocellulosic biomass pyrolysis is a complicated reaction process that involves complex components and reaction pathways. Due to measurement limitations, the intermediate components are difficult to be detected, therefore their detailed kinetics are still not well established. To address this issue, the Chemical Reaction Neural Network (CRNN) (Ji et al., 2022) was further applied to derive the lignocellulosic biomass pyrolysis kinetics. The derived pyrolysis kinetics, involving eight species and eleven reactions, could accurately reproduce the pyrolysis process for both the seen and unseen samples with R^2>0.95. The comparisons with the CRECK multi-step and Bio-CPD models also demonstrated the superiority of the CRNN model. This study provides a tool for establishing conversion kinetics of realistic solid fuels using chemistry-informed machine learning approaches.

Keywords: Lignocellulosic biomass; TGA; Machine learning; Chemical Reaction Neural Network; Pyrolysis kinetics

*Corresponding author:
Email address: xing.jiangkuan.64w@st.kyoto-u.ac.jp (Jiangkuan Xing)

1. Introduction

The utilization of biomass has gained traction in the recent years owing to the increasing demand for renewable and environmentally friendly energy sources. Generally, biomass can be utilized via thermo-chemical and bio-chemical routes to produce heat/electricity or biofuels and other chemical raw materials. Among the commonly-used thermo-chemical technologies, such as gasification and combustion (Wang et al., 2017), pyrolysis is always the primary process (Luo et al., 2017; Xing et al., 2018). Therefore, accurate modeling of this process is of great significance for the development and optimization of those utilization technologies.

Biomass pyrolysis is a complex reaction process, in which the chemical bonds of biomass components break and interact with each other during heating, thereby generating some intermediate species, and eventually terminating with the release of pyrolysis products, including tar and light gases. The key challenge in biomass pyrolysis modeling is to determine the complicated nonlinear reaction kinetics between the complex components, including the major and intermediate constituents. Numerous studies have been conducted to establish accurate kinetic models for biomass pyrolysis. The biomass pyrolysis model has progressed from simple global models, to multiple reactions models, and further to chemical structural/components based models. In the simple global models, the pyrolysis process is usually represented by a global reaction with different reaction mechanisms used (Mian et al., 2019; Singh et al., 2020; Li et al., 2020). In the multiple reactions models, such as the the distributed activation energy (Cai and Liu, 2008) and pseudo components (Sharma et al., 2019) models, the biomass pyrolysis is represented by multiple parallel reactions. In the chemical structural/components based models, such as the Bio-chemical percolation pyrolysis (Bio-CPD) (Sheng and Azevedo, 2002), Bio-functional group-depolymerization vaporization crosslinking (Chen et al., 1998), Bio-FLASHCHAIN (Niksa, 2020) and CRECK multi-step kinetic (Ranzi et al., 2017) models, the biomass structure and components are reasonably simplified, and the pyrolysis process is represented by multiple parallel reactions considering the biomass

structure and components. However, in the determination of the kinetic parameters of these models, the compensation effect between the pre-exponential factor and the activation energy (Xu et al., 2018), the reaction mechanism chosen in the global kinetic models (Mian et al., 2019), and the neglected interactions between major components of biomass (Wang et al., 2019) would introduce uncertainties into the modelling system (Gel et al., 2014). On the other hand, since the intermediate components species are difficult to be detected, the detailed components and their reaction kinetics cannot be determined. Therefore, the accurate biomass pyrolysis kinetics have not been well established and need further exploration (Xing et al., 2019d).

Machine learning approaches have been demonstrated to be an efficient tool for handling nonlinear issues well, and have been employed in biomass thermo-chemical fields, such as fuel properties modelling (Zhu et al., 2019; Xing et al., 2019b,c; Katongtung et al., 2022), pyrolysis modelling (Xing et al., 2019d; Sunphorka et al., 2017; Wei et al., 2022; Tang et al., 2021), and gasification modelling (Elmaz et al., 2020; Mutlu and Yucel, 2018). While it is noted that previous works on machine learning modelling of biomass pyrolysis could be divided into two types of framework: 1) kinetic-based framework (Xing et al., 2019d; Sunphorka et al., 2017; Xing et al., 2019a) and 2) weight-loss-based framework (Buyukada, 2019; Abbas et al., 2003; Xing et al., 2020). In the former one, the global kinetic models are often used with prescribed reaction mechanism, and machine learning methods are used to represent the nonlinear correlations between kinetic parameters, and biomass types/heating conditions (Xing et al., 2019d; Sunphorka et al., 2017). The choice of reaction mechanism and compensation effect would introduce uncertainty on the determination of kinetic parameters, further influence the machine learning modelling. In the later one, machine learning methods are used to represent the nonlinear correlations between particle mass loss and fuel types/heating conditions (Buyukada, 2019; Abbas et al., 2003; Wei et al., 2022). However, the particle heating history and intermediate species can not be considered in this framework, and the prediction capability for various fuel types needs to be further extended. In summary, the detailed components and their reaction kinetics still can

not be determined using the above machine learning frameworks. Very recently, Ji and Deng (Ji and Deng, 2021) developed the Chemical Reaction Neural Network (CRNN) to autonomously discover reaction pathways from the time-resolved species concentration data by making the chemistry informed by the neural network, which provided an efficient tool to handle the challenging problem of discovering unknown reaction kinetics from limited macroscopic measurements. Ji et al. (Ji et al., 2022) further applied this method to explore the kinetics of pyrolysis and oxidation of cellulose. However, the realistic lignocellulosic biomass is mainly composed of cellulose, hemicellulose, lignin and extractive contents, which is much more complex than the pure cellulose. So far, there has been no report on exploring the detailed pyrolysis kinetics of lignocellulosic biomass using such chemistry-informed machine learning method and directly comparing the developed machine learning models with previous well-established kinetic models.

In our previous study (Xing et al., 2022), this chemistry-informed machine learning method was further extended for lignocellulosic biomass pyrolysis, and the model performance was compared with those of the previous well-established kinetic models. Specifically, a database is first constructed by collecting thermogravimetric analysis (TGA) profiles for a wide range of lignocellulosic biomass types from available published literature. Pyrolysis kinetics are then derived from the database using the trained model. The performance of the derived kinetics is further evaluated by comparing with the unseen experimental data as well as the predictions of the well-established models, including the CRECK multi-step model and the Bio-CPD model. However, in the previous version (Xing et al., 2022), the above two very recent papers (Ji and Deng, 2021; Ji et al., 2022) were insufficiently cited and discussed, which is open to misunderstanding on their contributions. Therefore, we have decided to retract the previous version and republish this revised version in which those insufficient citations and discussions have been properly addressed, and our originality and new findings have also been explicitly stated.

2. Materials and approaches

2.1. Database construction

A pyrolysis database is constructed from available experimental data of lignocellulosic biomass pyrolysis under nitrogen atmosphere in published literature (Zhang et al., 2020; Li et al., 2004; Liu et al., 2011; Millán et al., 2017; Ding et al., 2019; Sobek and Werle, 2020; Rego et al., 2020; Gu et al., 2014) using the Digitizer module of Origin-Pro (Origin, Pro). Among the collected TGA measurements, the biomass samples include walnut shell, cotton stalk, corncob, legume straw, oil palm shells, bamboo guadua, beech wood, waste wood pellets, Ellert poplar, poplar wood sawdust, and synthetic biomass samples with the three major components in different proportions (see supplementary materials). The minimum temperature is ambient temperature, and the maximum temperature is 1373 K. The temperature ranges of TGA measurements are same for those collected from the same literature, but different for those collected from different literature (see supplementary materials). The collected data are pre-processed by transforming the major chemical compositions (cellulose, hemicellulose, and lignin) of biomass samples on a *daf* basis. Correspondingly, the TGA curves are then processed to exclude the ash and moisture components. Two notes are given here: 1) all collected samples have limited extractive content (less than 10%) so that the extractive content is not considered independently. Instead, it is attributed to the major chemical components with a weighted-average method; 2) each major component measured in each experiment is thought as the same pure-substance in the following modeling procedures. In total, 50 TGA curves for 28 types of biomass are collected, and each TGA curve contains 500 temporal measurements. The distribution of the three major components and heating rate are shown in Fig. 1. The ranges of cellulose, hemicellulose, lignin and heating rate are 0-100%, 0-100%, 0-100%, and 2-80 K/min, respectively (see supplementary materials). All 50 TGA curves are then randomly sorted into 40 "seen" and 10 "unseen" data. Curves from the seen data-set are randomly selected for model training and testing with proportions of 75% and 25%, respectively. The unseen data

will be used for model validation and comparison.

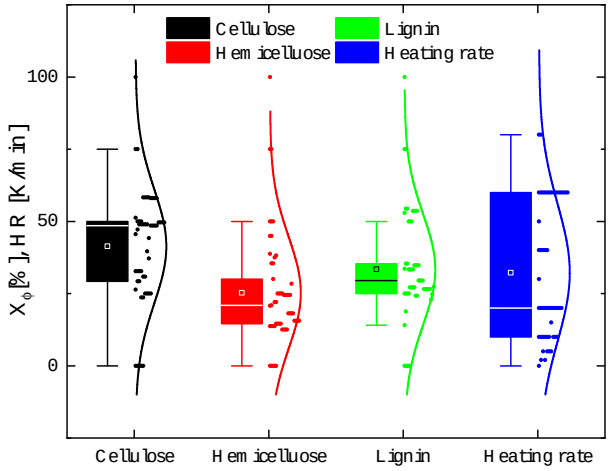

Figure 1: Data distribution of the major chemical compositions of biomass and heating rate in the constructed database.

2.2. CRNN for lignocellulosic biomass pyrolysis

In the present study, the lignocellulosic biomass pyrolysis kinetics are derived by combining the chemistry of lignocellulosic biomass pyrolysis with the artificial neural networks, which is consistent with the studies of Ji et al. (Ji and Deng, 2021; Ji et al., 2022). Different from the study of Ji et al. (Ji et al., 2022), the lignocellulosic biomass pyrolysis is modelled by considering the three major components (cellulose, hemicellulose, and lignin) and introducing some intermediate species. Taking one global pyrolysis reaction with six unknown intermediate species shown in Eq. (1) as an example,

$$v'_C Cel + v'_H Hem + v'_L Lig + v'_1 S_1 + v'_2 S_2 + v'_3 S_3 \longrightarrow +v''_4 S_4 + v''_5 S_5 + v''_6 S_6 + v''_v Vol. \quad (1)$$

Then, the reaction rate for Eq. (1) could be written and rearranged as,

$$r = [Cel]^{v'_C}[Hem]^{v'_H}[Lig]^{v'_L}[S_1]^{v'_1}[S_2]^{v'_2}[S_3]^{v'_3}[S_4]^0[S_5]^0[S_6]^0 A T_p^b exp(-E_a/RT_p), \quad (2)$$

$$r = exp\{v'_C ln[Cel] + v'_H ln[Hem] + v'_L ln[Lig] + v'_1 ln[S_1] + v'_2 ln[S_2] + v'_3 ln[S_3] + 0ln[S_4] + 0ln[S_5] + 0ln[S_6] + ln(A) + bln(T_p) - E_a/RT_p\}, \quad (3)$$

where Cel, Hem, and Lig are the cellulose, hemicellulose, and lignin fractions, respectively. S_1 to S_6 denote the unknown intermediate species. Vol signifies the released volatiles, including liquid and gaseous products. v_i' and v_i'' correspond to the stoichiometric coefficients of the reactants and products. A, E_a, T_p, b and R are the pre-exponential factor, activation energy, particle temperature, non-exponential temperature dependence factor, and universal constant, respectively. With the similar assumption that the stoichiometric coefficients and the reaction order are the same (Ji et al., 2022), the reaction rate of each species can then be written as (here only one reactant, one intermediate species, and one product are shown for brevity),

$$\frac{d[Cel]}{dt} = -v_C' r, \frac{d[S_1]}{dt} = -v_1' r, \frac{d[Vol]}{dt} = v_v'' r. \tag{4}$$

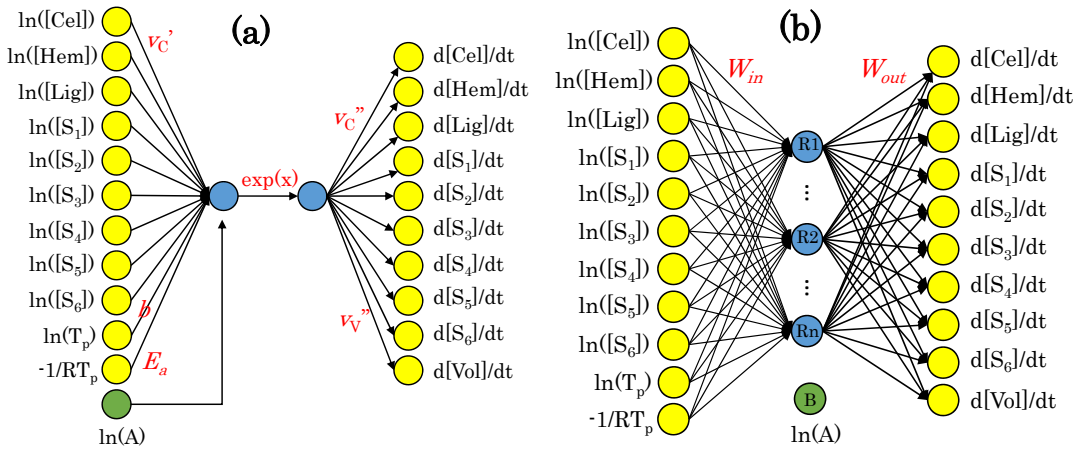

Figure 2: Neural network structure for lignocellulosic biomass pyrolysis considering the three major components and six unknown intermediate species as an example: (a) One global reaction with one neuron in the hidden layer; (b) Multiple reactions with multiple neurons in the hidden layer.

The output of the artificial neural network could be written as $\mathbf{y} = f(\mathbf{wx} + \mathbf{b})$, where $\mathbf{x}$ and $\mathbf{y}$ are the input and output vectors, $\mathbf{w}$ and $\mathbf{b}$ correspond to the weight and bias vectors, and f is the activation function which is an exponential function in the present study. It is noted here that the exponential function is used here to mathematically make the outputs of the artificial neural network to be the reaction

rate of pyrolysis reactions, which can not be achieved by using other types of activation functions. With $ln(Y_i)$, $Ln(T_p)$, and $-1/RT_p$ as the inputs and the exponential function as the activation function, the reaction rate in Eqs. (1) and (3) can then be represented by using a hidden neuron in the artificial neural network as shown in Fig. 2(a). Since the lignocellulosic biomass is mainly composed of three major components (Cel, Hem, and Lig), its pyrolysis should involve multi-step reactions between the three major components and some derived intermediate species, which could be further represented by using multiple neurons in the hidden layer, as shown in Fig. 2(b). It is noted that the structure of the artificial neural network is different from that of the previous work (Ji et al., 2022) with more inputs and outputs considering the complex components of lignocellulosic biomass (cellulose, hemicellulose, and lignin). In the artificial neural network, one neuron in the hidden layer denotes one reaction, whose reaction pathway and kinetics could be detailed with the weight vector and bias value. The interaction of those reactions are considered by calculating the instantaneous reaction rate of each species considering all involved species. The released volatiles are assumed to be only presented in the products and the mass stoichiometry balancing is employed between the reactants and products (Ji et al., 2022). The initial three major components of lignocellulosic biomass (cellulose, hemicellulose and lignin) are only presented in the reactants. In addition, with the assumption that the stoichiometric coefficients and reaction order are the same, the input and output weights could be shared to accelerate the training, increase the model regularization, and prevent over-fitting (Ji et al., 2022). These treatments make the chemistry informed in the artificial neural network, and eventually the CRNN model is achieved for lignocellulosic biomass pyrolysis.

The training of the model is performed to derive the reaction kinetics that satisfy $\dot{Y}$=$CRNN(Y)$. Ideally, the actual values of reaction rates should be known to calculate the loss function and back-propagate the error to the weight and bias vectors. However, as mentioned before, usually only the mass loss profiles could be directly obtained from TGA measurements, and the detailed intermediate components are hard to be measured. Therefore, the model is augmented with neural ordinary differential equation

(ODE) (Chen et al., 2018) to predict the residual mass and back propagate the error into the model parameters (Ji and Deng, 2021; Ji et al., 2022). In the current work, the stochastic gradient descent optimization approach is used to learn the weight and bias vectors of the model. Specifically, the commonly-used Adam optimizer proposed by Kingma and Ba (Kingma and Ba, 2014) is employed. The training process is then performed to low the mean sum error (MSE) as expressed in Eqs. (5) and (6), which is different from the mean absolute error in the study of Ji et al. (Ji and Deng, 2021; Ji et al., 2022).

$$Y^{CRNN}(t) = ODE(CRNN(Y), Y_0; T_0, \beta), \tag{5}$$

$$Loss = MSE(m_{res}^{CRNN}(t), m_{res}^{TGA}(t)) + MSE(m_{vol}^{CRNN}(t), m_{vol}^{TGA}(t)), \tag{6}$$

where $ODE()$ represents an ODE solution that could be numerically obtained with the reaction kinetics and initial conditions. The reaction kinetics are obtained from the weight and bias vectors of the CRNN model. Y_0, T_0 and β are the initial chemical composition, initial particle temperature, and heating rate, which are used to determine the initial status and the temperature evolution for the ODE solution. MSE is defined as $MSE = \frac{\sum_{i=1}^{N}(y_{actual}-y_{pred})^2}{N}$, where N is the total sample points, and y_{actual}/y_{pred} are the actual/predicted values, respectively. $m_{res}^{TGA}(t)$ and $m_{vol}^{TGA}(t)$ are the instantaneous mass residuals and released volatiles, respectively, which are measured by TGA. $m_{res}^{CRNN}(t)$ and $m_{vol}^{CRNN}(t)$ are the instantaneous mass residuals and released volatiles predicted by the CRNN model, respectively.

The numbers of species and reactions are hyper-parameters of the model, which correspond to the number of neurons in the input and output layers and the number of neurons in the hidden layer. The two numbers are determined with a trial-and-error grid-research method (*see Section 3.1.1*). In the model training, the mass fraction of each species is set within the range of [ϵ,1] to satisfy the mass rule and facilitate the logarithmic operation in Eq. (3), in which ϵ is a small value (10^{-8}). The model is trained for 2500 epochs, with each epoch consisting 30 times of weight and bias updates by iterating over 30 training samples. The updated model is then tested on

the remaining 10 test samples to obtain the test loss in each epoch. The initial learning rate is 5×10^{-3} and decays to 10^{-5} with a ratio of 0.2 every 500 epochs. Finally, a hard threshold pruning method is employed to further encourage sparsity in the learned weights (Ji et al., 2022). This pruning is operated by clipping the reaction orders and stoichiometric coefficients below a critical threshold, and the critical threshold is also determined by a trial-and-error grid-search method (*see Section 3.1.1*).

2.3. The CRECK multi-step and Bio-CPD models

The performance of the derived pyrolysis kinetics will be compared with those of the CRECK multi-step and Bio-CPD models for the unseen samples, with the experimental results as benchmarks. Therefore, the two models are briefly introduced here for completeness.

The CRECK multi-step model was developed by Ranzi et al. (Ranzi et al., 2008) and then modified by Blondeau and Jeanmart (Blondeau and Jeanmart, 2012). A total of 36 species and 19 reactions are involved in this model (see supplementary materials). Five biomass chemicals, *Cel*, *Hem*, *Lig-C*, *Lig-H*, and *Lig-O*, are set as the initial components, and the *Lig*-components could be determined based on the conservation of the C/H/O elemental composition from the study of Faravelli et al. (Faravelli et al., 2010) or a 2D interpolation based on elemental results from the reference data corresponding to a library whose *Lig* components are available (Ranzi et al., 2017). With the known heating rate and the initial particle temperature, the pyrolysis process could be calculated using the Arrhenius equations, and the pyrolysis products, including tar, light gases and solid residue, could be obtained by summing certain species. Specifically, 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, LIGOH, LIGCC, G{COg}, G{H_2}, G{CO_2}, G{COH_2} and Char.

The Bio-CPD model was developed by Sheng and Azevedo (Sheng and Azevedo,

2002) and further extended by Fletcher et al. (Fletcher et al., 2012). In this model, the general reaction framework of the coal CPD model was used, and three major components (cellulose, hemicellulose and lignin) were considered with different structural parameters and reaction kinetics (see supplementary materials). The pyrolysis process of each component was calculated and then merged based on the weight fraction of each component to obtain the final prediction.

3. Results and discussions

3.1. Hyper-parameters optimization and threshold pruning

The hyper-parameters that need to be optimized are the numbers of species and reactions (numbers of input neurons and hidden neurons), which are determined with a trial-and-error grid-research method. The number of species ranges from 5 to 12, which should equal to 4+n, where 4 represents the number of known species (Cel, Hem, Lig, and Vol), and n represents the number of unknown intermediate species. Thus, the number of intermediate species ranges from 1 to 8. The number of reactions ranges from the number of species to 12 to allow interactions between different components. Figure 3(a) shows the MSE values of the predictions of the trained models with different numbers of species and reactions. Overall, the increasing of the numbers of species and reactions helps to improve the model performance. For a certain species number, the MSE values firstly decrease and then stabilize around 0.004 with the increasing of reaction number (Luo et al., 2018). However, when reaction number is smaller than 9, the model performance would not be improved so much with the increasing of species number, indicating a less-fitting of the trained models. When reaction number is greater than 9, the model performance would be first improved and then stabilized with the increasing of species number. Therefore, to make a balance between the model complexity and accuracy, the model size of eight species and eleven reactions is selected in the present work (as marked with purple box in Fig. 3(a)). The selected model is further treated by a hard threshold pruning method to produce a sparse model and

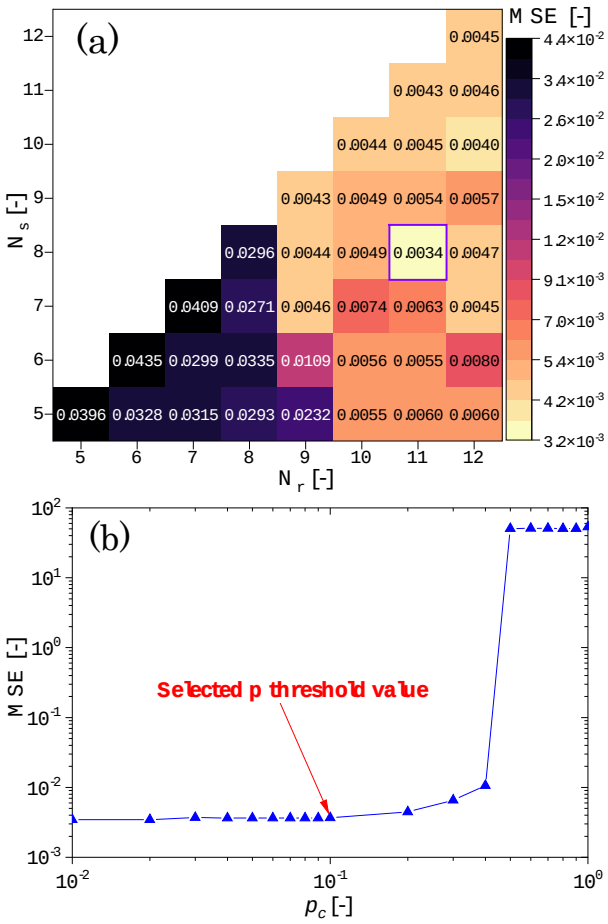

Figure 3: Hyper-parameters optimization and threshold pruning results: (a) MESs of the trained models with different numbers of species and reactions. (b) MSEs of the trained models with different pruning thresholds.

a better interpretation of the reaction kinetics. Figure 3(b) shows the MSE values of sparse models after different pruning thresholds from 0.01 to 1. The model performance is stabilized when the pruning threshold is smaller than 0.1, but obviously decreased when the pruning threshold exceeds this value. Therefore, this critical value is selected for the model pruning. The performance of the selected model after pruning is assessed on the seen data-sets by comparing with the experimental data. Figure 4 shows the comparisons of the TGA profiles of some seen (train and test) samples predicted by the trained model and measured in the experiments. The sample number and evaluation indicators including R^2 and MSE are also given in the plots, and the labels "train"

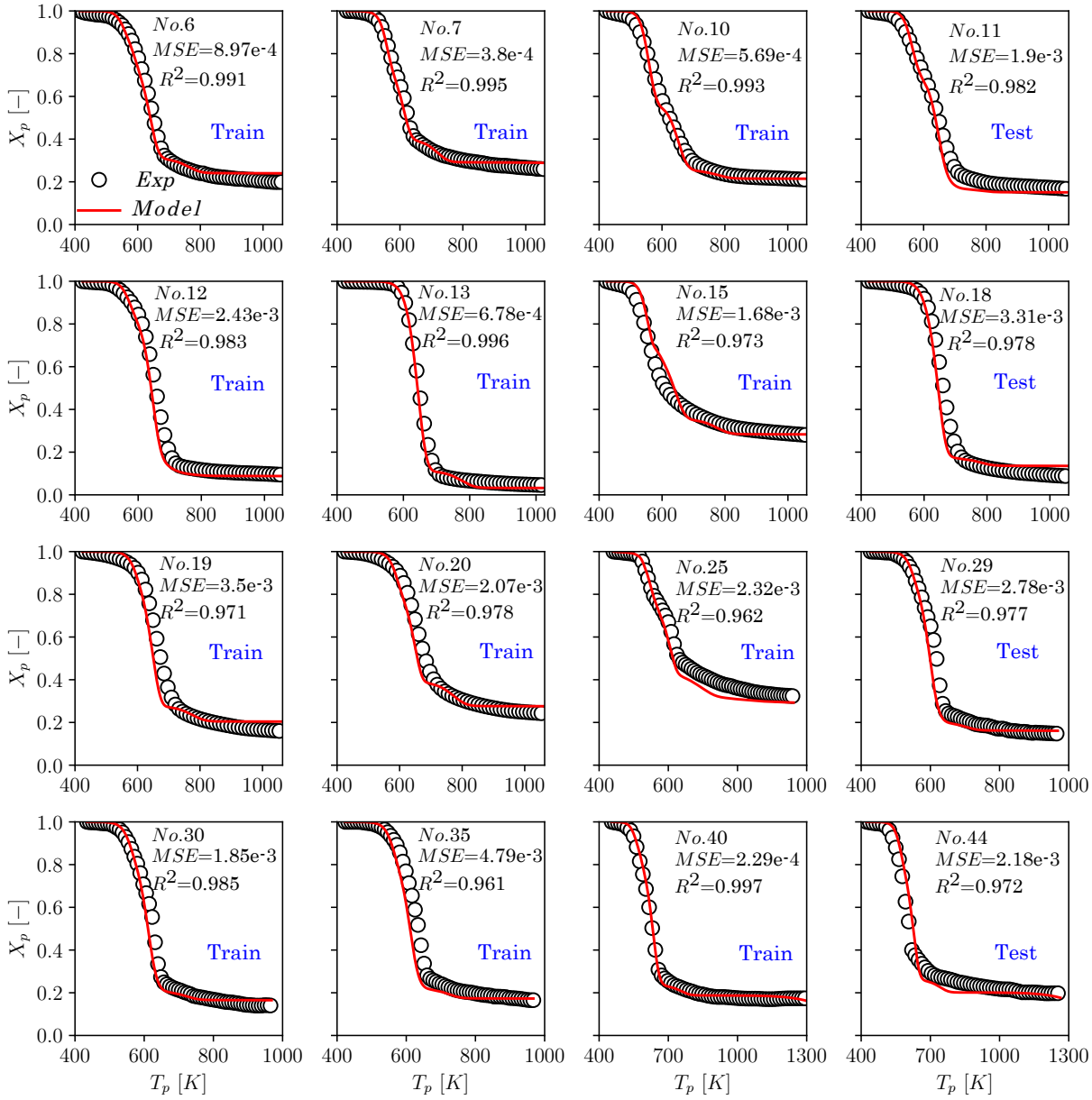

Figure 4: Comparisons of the TGA profiles predicted using the optimal trained model and measured from experiments for the seen samples. The evaluation indicators are also given in the plots, and the labels "train" and "test" mean that the samples are used for training and testing, respectively.

and "test" mean that the samples are used for training and testing, respectively. Note that only some samples are shown here for saving space, and the comparisons for the remaining seen samples are also conducted (see supplementary materials). In the machine learning study, the determination coefficient (R^2) is often used as an indicator

to measure the difference between the predicted and actual values. The perfect model should have a R^2 value of 1.0 and a MSE value of 0. The closer the value of R^2 is to 1, the higher the accuracy of the model. It is found that the model could accurately reproduce both the pyrolysis process and the final volatiles yields for various samples with R^2s larger than 0.96, and the mean/rms of R^2 and MSE for all seen samples are 0.978/0.01 and $3.4\times10^{-3}/1.1\times10^{-3}$, respectively. This is higher than those of the previous kinetic-based (Sunphorka et al., 2017) and mass-loss-based (Wei et al., 2022) machine learning models, validates the model on the seen samples and lays a solid foundation for the following interpretation of the pyrolysis kinetics.

3.2. Derived pyrolysis kinetics

Table 1: Reaction pathway and kinetic parameters of the pyrolysis kinetics derived by the model.

No	Reaction pathway	ln(A) [1/s]	E [kcal/mol]	b [-]
1	1.1Cel $\rightarrow$0.12S_4+0.98Vol	24.87	151.23	0.02
2	1.1S_2+1.1$S_4$$\rightarrow$1.48$S_1$+1.1Vol	28.28	164.76	0.07
3	1.1$S_2$$\rightarrow$0.54$S_3$+1.7$S_4$+0.94Vol	16.33	139.31	0.08
4	0.96S_1+0.58S_2+1.1$S_3$$\rightarrow$2.76$S_4$+1.1Vol	27.78	178.73	0.55
5	0.76S_1+0.98$S_4$$\rightarrow$0.2$S_2$+0.45$S_3$+1.1Vol	30.94	164.85	0.42
6	0.86$S_4$$\rightarrow$0.14$S_1$+0.38$S_2$+0.13$S_3$+0.21Vol	7.69	125.59	0.05
7	1.1$S_4$$\rightarrow$0.29$S_1$+0.44$S_2$+0.22$S_3$+0.22Vol	11.05	124.28	0.16
8	1.1Hem$\rightarrow$0.44S_1+0.66S_4	24.02	50.64	0.14
9	1.1Lig$\rightarrow$0.33S_1+0.14S_2+0.42S_3+0.28Vol	23.63	131.17	0.01
10	1.1$S_4$$\rightarrow$1.13$S_1$+0.36$S_2$+0.55$S_3$+0.47Vol	14.52	147.81	0.47
11	1.1Lig+1.06$S_1$$\rightarrow$ 0.91S_2+1.25S_4	4.00	14.89	0.01

In the trained model, the number of neurons in the hidden layer reflects the number of reactions. The weight vectors from the input layer to the hidden layer are the reaction orders, the activation energies, and temperature dependence factors of the reactions. The weight vectors from the hidden layer to the output layer are the stoichiometric coefficients of the reactions, and positive/negative values indicate that the

corresponding species are produced/consumed, whereas zero values mean that the corresponding species do not participate in the reaction. The bias vectors are the natural logarithms of the pre-exponential factors of the reactions. Pyrolysis kinetics, involving eight species and eleven reactions, could be interpreted from the weight and bias vectors (see supplementary materials) of the optimal model as listed in Table 1. The most important features of the optimal model are as follows: (i) the inputs are $ln(Y_i)$, $Ln(T_p)$, and $-1/RT_p$, which are obviously different from those of previous kinetic-based and mass-loss based models; (ii) the activation function is the exponential function, which is used to make the outputs of artificial neural networks be the reaction rates in the mathematical form and obviously different from those of previous machine learning studies; (iii) it features eight species and eleven reactions, and the reaction pathway and kinetics could be directly interpreted from the learned weight and bias vectors. In the derived pyrolysis kinetics, the three major components and volatiles are only existed in the reactants and products, respectively. The cyan rows (R1, R8, R9 and R11) represent the primary reactions involving the three initial major components (Cel, Hem and Lig). In these reactions, the cellulose, hemicellulose and lignin are consumed to form inter-mediate species (S_1 to S_4) and volatiles. The intermediate species are interacted with each other to form the released volatiles and solid residuals through the remaining reactions (white rows in Table 1). The extractive content is excluded in the modelling considering the following reasons: (i) among all the collected measurements, the extractive content is very small compared with the three major components (as claimed in Section 2.1); (ii) the dominant temperature range of the pyrolysis of this component would overlap with the temperature range of water evaporation, and it is hard to distinguish them and convert the TGA measurements to a *daf* basis; (iii) in the CRECK multi-step and Bio-CPD models, only the three major components are considered without the extractive content included. Therefore, this process is excluded for the convenience of direct comparisons between the derived kinetics, the CRECK multi-step model and Bio-CPD model. This will be comprehensively considered together with water evaporation by adding the two species into inputs, collecting more

samples to further rich the database, and further increasing the model size in the future study. Note that since the inputs are $ln(Y_i)$, $Ln(T_p)$, and $-1/RT_p$, which should be simultaneously updated in the calculation of pyrolysis process, the partial importance analysis of the reaction kinetics is hard to be conducted and has little physical meaning. However, a sensitivity analysis method similar as that of the gas-phase chemistry (Saltelli et al., 2005) could be used to identify the dominant reaction routes when the biomass compositions and heating conditions are known.

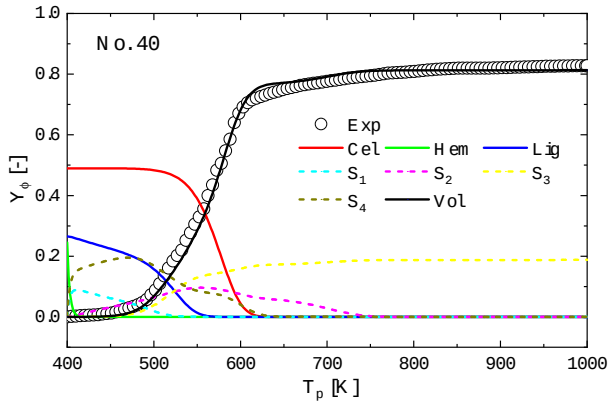

Figure 5: Temporal evolution of the mass fractions of the components in the derived kinetics for sample No. 40. The experimental data are also plotted here as a reference.

Taking the temporal evolution of the pyrolysis process of sample No. 40 shown in Fig. 5 as an example, it is evident that the initial species are the cellulose, hemicellulose and lignin, and the mass fractions of volatiles and the intermediate species are initially zero. After particles are heated, the mass fractions of cellulose, hemicellulose and lignin start to decrease, and the mass fractions of intermediate species and volatiles start to increase, which is caused by the occurring of primary reactions (R1, R8, R9 and R11). Then the intermediate species are interacted with each other and consumed to form the volatiles and solid residuals through the rest reactions. Their mass fractions would be temporally updated through the primary and interacted reactions during pyrolysis process. The final remaining components are S_3 and Vol, which corresponds to the solid residuals (char) and released volatiles, respectively. However, the types of the remaining intermediate species are still unknown. This is because the model

derives pyrolysis kinetics from the TGA measurements, which only measure the mass loss without any products information available. To determine the types of the rest intermediate species, atomic balance and pyrolysis products measurements (including gases and tar) are expected, which could be further explored in the future study.

3.3. Comparisons with the CRECK multi-step and Bio-CPD models

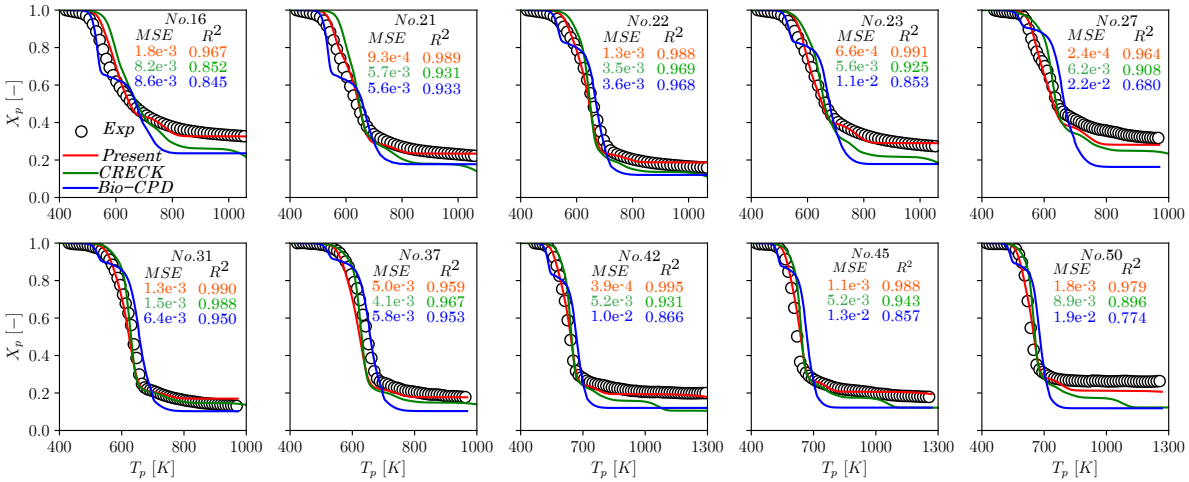

Figure 6: Comparisons of the TGA profiles measured in the experiments and predicted by the present derived kinetics, the CRECK multi-step model and the Bio-CPD model for the unseen samples. The sample numbers and evaluation indicators are also given in the plots.

To assess the performance of the pyrolysis kinetics derived by the optimal model on the unseen samples, the derived kinetics listed in Table 1 are further employed to predict the pyrolysis processes of the unseen samples. The performance is also compared with the well-established biomass pyrolysis models: the CRECK multi-step and Bio-CPD models. Figure 6 shows the comparisons of the TGA profiles measured in the experiments and predicted by the derived pyrolysis kinetics, the CRECK multi-step model, and the Bio-CPD model for the unseen samples. The sample numbers and evaluation indicators including R^2 and MSE are also given in each sub-figure. It shows that the Bio-CPD model could not accurately predict the pyrolysis process and overestimates the final volatiles yields. This could be attributed to the fact that the

Bio-CPD predictions are obtained by merging the pyrolysis process of each component with a weighted-sum method, and there are none interactions between the three major components considered. Therefore, the predicted TGA profiles feature three individual reacting stages. The CRECK multi-step model demonstrates a better performance than the Bio-CPD model but still overestimates the final volatiles yields for samples Nos. 16, 21 23, 27, 42 and 50, and the temperature at which the pyrolysis reactions start for samples Nos. 16, 21, 22, 23 and 27. In contrast, the pyrolysis kinetics derived by the present model (see the red lines) could (i) accurately predict the starting of pyrolysis reactions and the finale volatiles yield; (ii) well reproduce the pyrolysis process with the determination coefficients larger than 0.95 for all unseen samples, which is obviously higher than those of the previous two models. This indicates the superior performance of the derived pyrolysis kinetics to the those of the two well-established models and demonstrates the advantages of the CRNN method.

4. Conclusions

Chemical Reaction Neural Network was applied to explore the lignocellulosic biomass pyrolysis kinetics by directly combining the pyrolysis chemistry with the artificial neural networks. The derived pyrolysis kinetics could accurately reproduce the pyrolysis process for both the seen and unseen samples, demonstrating superiority against the kinetics of CRECK multi-step and Bio-CPD models in predicting the final volatiles yield and the pyrolysis process. This study provides a tool for establishing conversion kinetics of realistic solid fuels using chemistry informed machine learning approaches.

E-supplementary data of this work can be found in online version of the paper.

Acknowledgments

The authors thank for the supports from Japan Society for the Promotion of Science (Grant No. 21P20351) and MEXT as "Program for Promoting Researches on the Supercomputer Fugaku" (Digital Twins of Real World's Clean Energy Systems with

Integrated Utilization of Super-simulation and AI) (Grant No. JPMXP1020200303). JX and KL are grateful for the support from the National Natural and Science Foundation of China (Grant No: 52106211 and 51925603). JX especially thanks to the helpful discussions on the code development with Dr. Weiqi Ji and valuable comments on the manuscript from Prof. Sili Deng at Massachusetts Institute of Technology.

References

Abbas, T., Awais, M.M., Lockwood, F.C., 2003. An artificial intelligence treatment of devolatilization for pulverized coal and biomass in co-fired flames. Combustion and Flame 132, 305–318.

Blondeau, J., Jeanmart, H., 2012. Biomass pyrolysis at high temperatures: Prediction of gaseous species yields from an anisotropic particle. Biomass and bioenergy 41, 107–121.

Buyukada, M., 2019. Investigation of thermal conversion characteristics and performance evaluation of co-combustion of pine sawdust and lignite coal using TGA, artificial neural network modeling and likelihood method. Bioresource Technology 287, 121461.

Cai, J., Liu, R., 2008. New distributed activation energy model: Numerical solution and application to pyrolysis kinetics of some types of biomass. Bioresource Technology 99, 2795–2799.

Chen, R.T., Rubanova, Y., Bettencourt, J., Duvenaud, D., 2018. Neural ordinary differential equations. arXiv preprint arXiv:1806.07366 .

Chen, Y., Charpenay, S., Jensen, A., Wójtowicz, M.A., Serio, M.A., 1998. Modeling of biomass pyrolysis kinetics. Symposium (International) on Combustion 27, 1327–1334.

Ding, Y., Huang, B., Wu, C., He, Q., Lu, K., 2019. Kinetic model and parameters study of lignocellulosic biomass oxidative pyrolysis. Energy 181, 11–17.

Elmaz, F., Yucel, O., Mutlu, A.Y., 2020. Predictive modeling of biomass gasification with machine learning-based regression methods. Energy 191, 116541.

Faravelli, T., Frassoldati, A., Migliavacca, G., Ranzi, E., 2010. Detailed kinetic modeling of the thermal degradation of lignins. Biomass and bioenergy 34, 290–301.

Fletcher, T.H., Pond, H.R., Webster, J., Wooters, J., Baxter, L.L., 2012. Prediction of tar and light gas during pyrolysis of black liquor and biomass. Energy & Fuels 26, 3381–3387.

Gel, A., Chaudhari, K., Turton, R., Nicoletti, P., 2014. Application of uncertainty quantification methods for coal devolatilization kinetics in gasifier modeling. Powder Technology 265, 66–75.

Gu, X., Liu, C., Jiang, X., Ma, X., Li, L., Cheng, K., Li, Z., 2014. Thermal behavior and kinetics of the pyrolysis of the raw/steam exploded poplar wood sawdust. Journal of Analytical and Applied Pyrolysis 106, 177–186.

Ji, W., Deng, S., 2021. Autonomous discovery of unknown reaction pathways from data by chemical reaction neural network. The Journal of Physical Chemistry A 125, 1082–1092.

Ji, W., Richter, F., Gollner, M.J., Deng, S., 2022. Autonomous kinetic modeling of biomass pyrolysis using chemical reaction neural networks. Combustion and Flame 240, 111992.

Katongtung, T., Onsree, T., Tippayawong, N., 2022. Machine learning prediction of biocrude yields and higher heating values from hydrothermal liquefaction of wet biomass and wastes. Bioresource Technology 344, 126278.

Kingma, D.P., Ba, J., 2014. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 .

Li, H., Zhou, N., Dai, L., Cheng, Y., Cobb, K., Chen, P., Ruan, R., 2020. Effect of lime mud on the reaction kinetics and thermodynamics of biomass pyrolysis. Bioresource Technology 310, 123475.

Li, S., Xu, S., Liu, S., Yang, C., Lu, Q., 2004. Fast pyrolysis of biomass in free-fall reactor for hydrogen-rich gas. Fuel Processing Technology 85, 1201–1211.

Liu, Q., Zhong, Z., Wang, S., Luo, Z., 2011. Interactions of biomass components during pyrolysis: A tg-ftir study. Journal of Analytical and Applied Pyrolysis 90, 213–218.

Luo, K., Xing, J., Bai, Y., Fan, J., 2017. Universal Devolatilization Process Model for Numerical Simulations of Coal Combustion. Energy Fuels 31, 6525–6540.

Luo, K., Xing, J., Bai, Y., Fan, J., 2018. Prediction of product distributions in coal devolatilization by an artificial neural network model. Combustion and Flame 193, 283–294.

Mian, I., Li, X., Jian, Y., Dacres, O.D., Zhong, M., Liu, J., Ma, F., Rahman, N., 2019. Kinetic study of biomass pellet pyrolysis by using distributed activation energy model and Coats Redfern methods and their comparison. Bioresource Technology 294, 122099.

Millán, L.M.R., Vargas, F.E.S., Nzihou, A., 2017. Kinetic analysis of tropical lignocellulosic agrowaste pyrolysis. BioEnergy Research 10, 832–845.

Mutlu, A.Y., Yucel, O., 2018. An artificial intelligence based approach to predicting syngas composition for downdraft biomass gasification. Energy 165, 895–901.

Niksa, S., 2020. bio-FLASHCHAIN® theory for rapid devolatilization of biomass 1. Lignin devolatilization. Fuel 263, 116649.

Origin(Pro), 2021. Version 2021. Originlab Corporation, Northampton, MA, USA. URL: https://www.originlab.com/origin.

Ranzi, E., Cuoci, A., Faravelli, T., Frassoldati, A., Migliavacca, G., Pierucci, S., Sommariva, S., 2008. Chemical kinetics of biomass pyrolysis. Energy & Fuels 22, 4292–4300.

Ranzi, E., Debiagi, P.E.A., Frassoldati, A., 2017. Mathematical modeling of fast biomass pyrolysis and bio-oil formation. note i: Kinetic mechanism of biomass pyrolysis. ACS Sustainable Chemistry & Engineering 5, 2867–2881.

Rego, F., Dias, A.P.S., Casquilho, M., Rosa, F.C., Rodrigues, A., 2020. Pyrolysis kinetics of short rotation coppice poplar biomass. Energy 207, 118191.

Saltelli, A., Ratto, M., Tarantola, S., Campolongo, F., 2005. Sensitivity analysis for chemical models. Chemical reviews 105, 2811–2828.

Sharma, P., Pandey, O.P., Diwan, P.K., 2019. Non-isothermal kinetics of pseudo-components of waste biomass. Fuel 253, 1149–1161.

Sheng, C., Azevedo, J., 2002. Modeling biomass devolatilization using the chemical percolation devolatilization model for the main components. Proceedings of the Combustion Institute 29, 407–414.

Singh, R.K., Pandey, D., Patil, T., Sawarkar, A.N., 2020. Pyrolysis of banana leaves biomass: Physico-chemical characterization, thermal decomposition behavior, kinetic and thermodynamic analyses. Bioresource Technology 310, 123464.

Sobek, S., Werle, S., 2020. Kinetic modelling of waste wood devolatilization during pyrolysis based on thermogravimetric data and solar pyrolysis reactor performance. Fuel 261, 116459.

Sunphorka, S., Chalermsinsuwan, B., Piumsomboon, P., 2017. Artificial neural network model for the prediction of kinetic parameters of biomass pyrolysis from its constituents. Fuel 193, 142–158.

Tang, Q., Chen, Y., Yang, H., Liu, M., Xiao, H., Wang, S., Chen, H., Naqvi, S.R., 2021. Machine learning prediction of pyrolytic gas yield and compositions with feature reduction methods: Effects of pyrolysis conditions and biomass characteristics. Bioresource Technology 339, 125581.

Wang, J., Shen, B., Kang, D., Yuan, P., Wu, C., 2019. Investigate the interactions between biomass components during pyrolysis using in-situ DRIFTS and TGA. Chemical Engineering Science 195, 767–776.

Wang, S., Dai, G., Yang, H., Luo, Z., 2017. Lignocellulosic biomass pyrolysis mechanism: A state-of-the-art review. Progress in Energy and Combustion Science 62, 33–86.

Wei, H., Luo, K., Xing, J., Fan, J., 2022. Predicting co-pyrolysis of coal and biomass using machine learning approaches. Fuel 310, 122248.

Xing, J., Bai, Y., Zhao, C., Gao, Z., Wang, H., 2018. Numerical Studies of Coal Devolatilization Characteristics with Gas Temperature Fluctuation. Energy Fuels 32, 8760–8767.

Xing, J., Kurose, R., Luo, K., Fan, J., 2022. Retracted: Chemistry-informed neural networks modelling of lignocellulosic biomass pyrolysis. Bioresource Technology 355, 127275. URL: https://www.sciencedirect.com/science/article/pii/S0960852422006046, doi:https://doi.org/10.1016/j.biortech.2022.127275.

Xing, J., Luo, K., Pitsch, H., Wang, H., Bai, Y., Zhao, C., Fan, J., 2019a. Predicting kinetic parameters for coal devolatilization by means of artificial neural networks. Proceedings of the Combustion Institute 37, 2943–2950.

Xing, J., Luo, K., Wang, H., Fan, J., 2019b. Estimating biomass major chemical constituents from ultimate analysis using a random forest model. Bioresource Technology , 121541.

Xing, J., Luo, K., Wang, H., Gao, Z., Fan, J., 2019c. A comprehensive study on estimating higher heating value of biomass from proximate and ultimate analysis with machine learning approaches. Energy , 116077.

Xing, J., Luo, K., Wang, H., Jin, T., Fan, J., 2020. Novel Sensitivity Study for Biomass Directional Devolatilization by Random Forest Models. Energy Fuels 34, 8414–8423.

Xing, J., Wang, H., Luo, K., Wang, S., Bai, Y., Fan, J., 2019d. Predictive single-step kinetic model of biomass devolatilization for CFD applications: A comparison study of empirical correlations (EC), artificial neural networks (ANN) and random forest (RF). Renewable Energy 136, 104–114.

Xu, D., Chai, M., Dong, Z., Rahman, M.M., Yu, X., Cai, J., 2018. Kinetic compensation effect in logistic distributed activation energy model for lignocellulosic biomass pyrolysis. Bioresource Technology 265, 139–145.

Zhang, L., Yang, Z., Li, S., Wang, X., Lin, R., 2020. Comparative study on the two-step pyrolysis of different lignocellulosic biomass: Effects of components. Journal of Analytical and Applied Pyrolysis 152, 104966.

Zhu, X., Li, Y., Wang, X., 2019. Machine learning prediction of biochar yield and carbon contents in biochar based on biomass characteristics and pyrolysis conditions. Bioresource Technology 288, 121527.

List of Tables

List of Figures

Chemical Reaction Neural Network modelling of lignocellulosic biomass pyrolysis

Jiangkuan Xing[a,d,*], Ryoichi Kurose[a], Kun Luo[b,c], Jianren Fan[b,c]

[a] *Department of Mechanical Engineering and Science, Kyoto University, Kyoto, 615–8540, Japan*
[b] *State Key Laboratory of Clean Energy Utilization, Zhejiang University, Hangzhou, 310027, China*
[c] *Shanghai Institute for Advanced Study of Zhejiang University, Shanghai 200120, China*
[d] *JSPS International Research Fellow, Kyoto University, Japan*

Supplementary materials

Table S1: Chemical compositions and heating conditions of the samples in the constructed database.

No[-]	Type	Cel[%]	Hem[%]	Lig[%]	Ext[%]	β[K/min]	T_p[K]	References
1	AWS	25.15	19.86	50.55	4.44	20	298-1023	(Zhang et al., 2020)
2	ACS	44.48	20.44	32.81	2.26	20	298-1023	(Zhang et al., 2020)
3	CC	45.26	37.20	13.50	4.04	20	298-1023	(Zhang et al., 2020)
4	ACC	49.70	29.10	18.27	2.92	20	298-1023	(Zhang et al., 2020)
5	LS	28.65	34.71	34.65	1.99	10	298-1073	(Li et al., 2004)
6	LS	28.65	34.71	34.65	1.99	40	298-1073	(Li et al., 2004)
7	AS	22.40	20.87	51.52	5.21	10	298-1073	(Li et al., 2004)
8	AS	22.40	20.87	51.52	5.21	40	298-1073	(Li et al., 2004)
9	HEC	0.00	100.0	0.00	0.00	60	303-1073	(Liu et al., 2011)
10	HC31	25.00	75.00	0.00	0.00	60	303-1073	(Liu et al., 2011)
11	HC11	50.00	50.00	100.0	0.00	60	303-1073	(Liu et al., 2011)
12	HC13	75.00	25.00	0.00	0.00	60	303-1073	(Liu et al., 2011)
13	CEL	100.0	0.00	0.00	0.00	60	303-1073	(Liu et al., 2011)
14	LIG	0.00	0.00	100.0	0.00	60	303-1073	(Liu et al., 2011)
15	HL31	0.00	75.00	25.00	0.00	60	303-1073	(Liu et al., 2011)
16	HL11	0.00	50.00	50.00	0.00	60	303-1073	(Liu et al., 2011)

Continued on next page

*Corresponding author:
Email address: xing.jiangkuan.64w@st.kyoto-u.ac.jp (Jiangkuan Xing)

Table S1 – continued from previous page

No[-]	Type	Cel[%]	Hem[%]	Lig[%]	Ext[%]	β[K/min]	T_p[K]	References
17	HL13	0.00	25.00	75.00	0.00	60	303-1073	(Liu et al., 2011)
18	CL31	75.00	0.00	25.00	0.00	60	303-1073	(Liu et al., 2011)
19	CL11	50.00	0.00	50.00	0.00	60	303-1073	(Liu et al., 2011)
20	CL13	25.00	0.00	75.00	0.00	60	303-1073	(Liu et al., 2011)
21	CHL121	25.00	50.00	25.00	0.00	60	303-1073	(Liu et al., 2011)
22	CHL211	50.00	25.00	25.00	0.00	60	303-1073	(Liu et al., 2011)
23	CHL112	25.00	25.00	50.00	0.00	60	303-1073	(Liu et al., 2011)
24	OPS	30.40	12.70	49.80	7.10	2	298-1073	(Millán et al., 2017)
25	OPS	30.40	12.70	49.80	7.10	5	298-1073	(Millán et al., 2017)
26	OPS	30.40	12.70	49.80	7.10	10	298-1073	(Millán et al., 2017)
27	OPS	30.40	12.70	49.80	7.10	20	298-1073	(Millán et al., 2017)
28	BG	53.90	13.50	25.10	7.50	2	298-1073	(Millán et al., 2017)
29	BG	53.90	13.50	25.10	7.50	5	298-1073	(Millán et al., 2017)
30	BG	53.90	13.50	25.10	7.50	10	298-1073	(Millán et al., 2017)
31	BG	53.90	13.50	25.10	7.50	20	298-1073	(Millán et al., 2017)
32	LB	28.30	41.20	22.00	8.30	5	400-800	(Ding et al., 2019)
33	LB	28.30	41.20	22.00	8.30	10	400-800	(Ding et al., 2019)
34	WW	54.80	11.80	27.90	5.50	5	298-973	(Sobek and Werle, 2020)
35	WW	54.80	11.80	27.90	5.50	10	298-973	(Sobek and Werle, 2020)
36	WW	54.80	11.80	27.90	5.50	15	298-973	(Sobek and Werle, 2020)
37	WW	54.80	11.80	27.90	5.50	20	298-973	(Sobek and Werle, 2020)
38	EG	48.00	24.00	26.00	2.00	10	303-1373	(Rego et al., 2020)
39	EG	48.00	24.00	26.00	2.00	20	303-1373	(Rego et al., 2020)
40	EG	48.00	24.00	26.00	2.00	30	303-1373	(Rego et al., 2020)
41	EG	48.00	24.00	26.00	2.00	40	303-1373	(Rego et al., 2020)
42	EG	48.00	24.00	26.00	2.00	50	303-1373	(Rego et al., 2020)
43	PWS	45.51	17.01	31.24	6.24	10	298-1273	(Gu et al., 2014)
44	PWS	45.51	17.01	31.24	6.24	20	298-1273	(Gu et al., 2014)
45	PWS	45.51	17.01	31.24	6.24	40	298-1273	(Gu et al., 2014)
46	PWS	45.51	17.01	31.24	6.24	80	298-1273	(Gu et al., 2014)

Continued on next page

Table S1 – continued from previous page

No[-]	Type	Cel[%]	Hem[%]	Lig[%]	Ext[%]	β[K/min]	T_p[K]	References
47	SPWS	45.35	14.28	31.83	8.54	10	298-1273	(Gu et al., 2014)
48	SPWS	45.35	14.28	31.83	8.54	20	298-1273	(Gu et al., 2014)
49	SPWS	45.35	14.28	31.83	8.54	40	298-1273	(Gu et al., 2014)
50	SPWS	45.35	14.28	31.83	8.54	80	298-1273	(Gu et al., 2014)

Table S2: Chemical compositions and heating conditions of the samples in the constructed database.

No	Reaction pathway	A[1/s]	E[kJ/mol]
1	CELL→ CELLA	2.80×10^{19}	242.4
2	CELLA→ $0.95C_2H_4O_2 + 0.25C_2H_2O_2 + 0.2CH_3CHO + 0.25C_6H_6O_3 + 0.2C_3H_6O + 0.16CO_2 + 0.23CO + 0.9H_2O + 0.1CH_4 + 0.61$Char	1.30×10^{10}	150.5
3	CELLA→$C_6H_{10}O_5$	3.28×10^{14}	196.5
4	CELL→ $5H_2O$ + 6Char	8.00×10^{7}	133.9
5	HCE→ 0.4HCE1 + 0.6HCE2	2.10×10^{16}	186.7
6	HCE1→ $0.75G\{H_2\} + 0.8CO_2 + 1.4CO + 0.5CH_2O + 0.25CH_3OH + 0.125C_2H_6O + 0.125H_2O + 0.625CH_4 + 0.25C_2H_4 + 0.675$Char	2.60×10^{11}	145.7
7	HCE1→ $C_5H_8O_4$	8.75×10^{15}	202.4
8	HCE2→ $0.2CO_2 + 0.5CH_4 + 0.25C_2H_4 + 0.8G\{CO_2\} + 0.8G\{COH_2\} + 0.7CH_2O + 0.25CH_3OH + 0.125C_2H_6O + 0.125H_2O$ + Char	1.00×10^{10}	138.1
9	LIG-C→ 0.35LIGCC + $0.1C_9H_10O_2 + 0.08C_6H_6O + 0.41C_2H4 + H_2O + 0.495CH_4 + 0.32CO + G\{COH_2\}$ + 5.735Char	4.00×10^{15}	202.9
10	LIG-H→ LIGOH + C_3H_6O	2.00×10^{13}	156.9
11	LIG-O→ LIGOH + CO2	1.00×10^{9}	106.7
12	LIGCC→ $0.3C_9H_{10}O_2 + 0.2C_6H_6O + 0.35C3H_4O_2 + 0.7H_2O + 0.65CH_4 + 0.6C_2H_4 + G\{COH_2\} + 0.8G\{CO\}$ + 6.4Char	5.00×10^{6}	131.8
13	LIGOH→ 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→ $C_{11}H_{12}O_4$	1.50×10^{9}	143.8
15	LIG→ $H_2O + 0.5CO + 0.2CH_2O + 0.4CH_3OH + 0.2CH_3CHO + 0.2C_3H_6O + 0.6CH_4 + G\{CO\} + 0.5G\{COH_2\}$ + 5.5Char	7.70×10^{6}	111.4

Continued on next page

Table S2 – continued from previous page

No	Reaction pathway	A[1/s]	E[kJ/mol]
16	G{CO_2}→ CO_2	1.00×10^6	100.4
17	G{CO}→ CO	5.00×10^{12}	209.2
18	G{COH_2}→ CO + H_2	1.50×10^{12}	297.1
19	G{H2}→ H_2	5.00×10^{11}	313.8

Table S3: Structural and kinetic parameters for different components in the Bio-CPD model (Fletcher et al., 2012).

Parameters	Cellulose	Hemicellulose	Lignin
p_0 [-]	1.0	1.0	0.71
c_0 [-]	0.0	0.0	0.0
$\sigma + 1$ [-]	3.0	3.0	3.5
m_w [-]	81	77.5	208
m_{del} [-]	22.7	21.5	39
A_b [s^{-1}]	2×10^{16}	1.2×10^{20}	7×10^{16}
E_b [kcal/mol]	55400	51500	55400
E_b^σ [kcal/mol]	4100	100	5009
A_c [s^{-1}]	100	1.35	1.7
E_c [kcal/mol]	0	0	0
A_g [s^{-1}]	3×10^{15}	3×10^{15}	2.3×10^{19}
E_g [kcal/mol]	61200	38200	69000
E_g^σ [kcal/mol]	8100	5000	39
A_{cr} [s^{-1}]	3×10^{15}	3×10^{15}	3×10^{15}
E_{cr} [kcal/mol]	65000	65000	65000

Table S4: Weight and bias vectors of the optimal model with eight species and eleven reactions.

		R1	R2	R3	R4	R5	R6	R7	R8	R9	R10	R11
W_in	Cell	1.1	0	0	0	0	0	0	0	0	0	0
	Hem	0	0	0	0	0	0	0	1.1	0	0	0
	Lig	0	0	0	0	0	0	0	0	1.1	0	1.1
	S1	0	0	0	0.96	0.76	0	0	0	0	0	1.06
	S2	0	1.1	1.1	0.58	0	0	0	0	0	0	0
	S3	0	0	0	1.1	0	0	0	0	0	0	0
	S4	0	1.1	0	0	0.98	0.86	1.1	0	0	1.1	0
	Vol	0	0	0	0	0	0	0	0	0	0	0
	E	151.2	164.7	139.3	178.7	164.8	125.6	124.3	50.6	131.2	147.8	14.9
	b	0.02	0.07	0.08	0.55	0.42	0.05	0.16	0.14	0.01	0.47	0.01
B	ln(A)	24.87	28.28	16.33	27.78	30.94	7.69	11.05	10.86	23.63	14.52	4
W_out'	Cell	-1.1	0	0	0	0	0	0	0	0	0	0
	Hem	0	0	0	0	0	0	0	-1.1	0	0	0
	Lig	0	0	0	0	0	0	0	0	-1.1	0	-1.1
	S1	0	1.48	0	-0.96	-0.76	0.14	0.29	0.44	0.33	1.13	-1.06
	S2	0	-1.1	-1.1	-0.58	0.2	0.38	0.44	0	0.14	0.36	0.91
	S3	0	0	0.54	-1.1	0.45	0.13	0.22	0	0.42	0.55	0
	S4	0.12	-1.1	1.7	2.76	-0.98	-0.86	-1.1	0.66	0	-1.1	1.25
	Vol	0.98	1.1	0.94	1.1	1.1	0.21	0.22	0	0.28	0.47	0

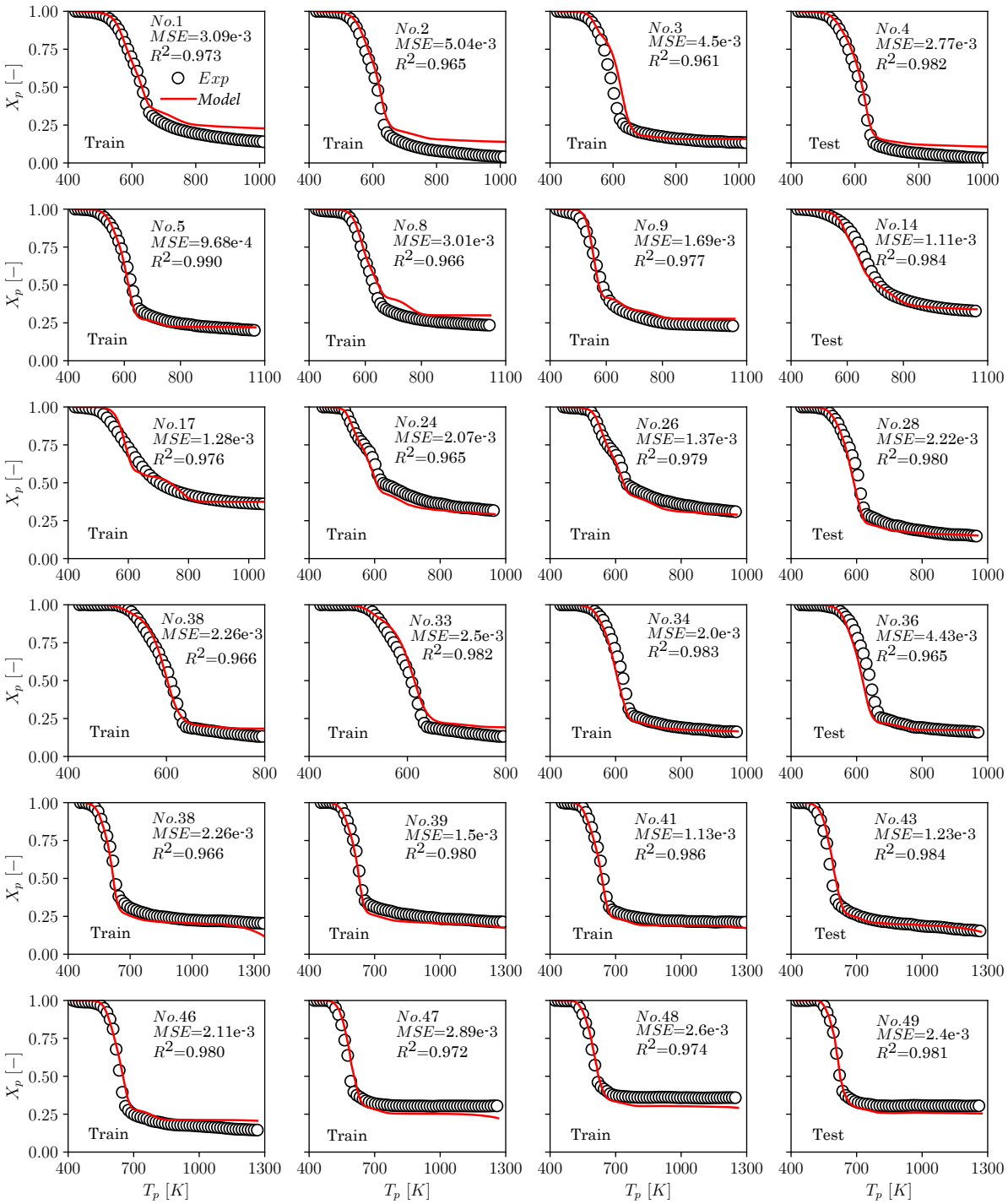

Figure S1: Comparisons of the TGA profiles predicted by the optimal model and measured from the experiments for the remaining "seen" samples. Sample number and R^2/MSE of the predictions are also given in each sub-figure. The labels "train" and "test" means that the samples are used for train and test, respectively.

References

Ding, Y., Huang, B., Wu, C., He, Q., Lu, K., 2019. Kinetic model and parameters study of lignocellulosic biomass oxidative pyrolysis. Energy 181, 11–17.

Fletcher, T.H., Pond, H.R., Webster, J., Wooters, J., Baxter, L.L., 2012. Prediction of tar and light gas during pyrolysis of black liquor and biomass. Energy & Fuels 26, 3381–3387.

Gu, X., Liu, C., Jiang, X., Ma, X., Li, L., Cheng, K., Li, Z., 2014. Thermal behavior and kinetics of the pyrolysis of the raw/steam exploded poplar wood sawdust. Journal of Analytical and Applied Pyrolysis 106, 177–186.

Li, S., Xu, S., Liu, S., Yang, C., Lu, Q., 2004. Fast pyrolysis of biomass in free-fall reactor for hydrogen-rich gas. Fuel Processing Technology 85, 1201–1211.

Liu, Q., Zhong, Z., Wang, S., Luo, Z., 2011. Interactions of biomass components during pyrolysis: A tg-ftir study. Journal of Analytical and Applied Pyrolysis 90, 213–218.

Millán, L.M.R., Vargas, F.E.S., Nzihou, A., 2017. Kinetic analysis of tropical lignocellulosic agrowaste pyrolysis. BioEnergy Research 10, 832–845.

Rego, F., Dias, A.P.S., Casquilho, M., Rosa, F.C., Rodrigues, A., 2020. Pyrolysis kinetics of short rotation coppice poplar biomass. Energy 207, 118191.

Sobek, S., Werle, S., 2020. Kinetic modelling of waste wood devolatilization during pyrolysis based on thermogravimetric data and solar pyrolysis reactor performance. Fuel 261, 116459.

Zhang, L., Yang, Z., Li, S., Wang, X., Lin, R., 2020. Comparative study on the two-step pyrolysis of different lignocellulosic biomass: Effects of components. Journal of Analytical and Applied Pyrolysis 152, 104966.