Supplementary information for: General Component Additivity, Reaction Engineering, and Machine Learning Models for Hydrothermal Liquefaction

Peter M. Guirguis, Philip E. Savage*
Chemical Engineering Department, Pennsylvania State University,
121D CBEB Building, 375 Science Dr, University Park, PA 16802, USA
* psavage@psu.edu

Table S1†: Parameters for the component additivity model in Eqn. 11.

Parameter label	Parameters values for $ln(SI)$					
	L	k_1	k_2	s_1	s_2	M
C_{St}	-56.5	0.00216	1.74	-0.881	-0.876	43.6
C_{Ce}	36.3	0.738	0.147	-1.40	0.738	11.6
C_{He}	168	1.79	0.905	0.101	1.45	18.0
C_{Pe}	-6.16	0.000147	551	-0.797	0.410	41.9
C_{Lp}	94.6	0.143	0.949	4.63	4.63	65.8
C_{Lg}	50.9	3.96	26.4	0.544	3.08	3.26
C_{Sa}	-7.75	10.8	78.9	-3.92	2.50	-5.90e-05
C_{AA}	-121	18.0	0.911	6.75	6.75	8.03
C_{FA}	106	0.216	18.4	-7.16	2.35	38.3
C_{Ph}	143	0.158	0.810	-3.14	2.54	21.6
C_{C_t,P_t}	284	0.323	1.97	0.893	3.16	-54.0
C_{C_t,F_t}	483	0.175	5.20	-6.52	-1.09	-62.6
C_{P_t,F_t}	-381	2.63e-05	67.0	-5.06	5.95	-0.360
C_{C_t,P_t,F_t}	-1810	0.000453	0.267	-0.377	-0.373	330
C_{C_t,P_t,L_t}	-825	0.554	0.221	-9.53	-0.285	-4.36
	Parameters values for loading mass					
label	k_3	k_4	s_3	s_4		
C_{St}	5.50	0.203	-0.882	0.00475		
C_{Ce}	2.22	0.0921	-0.0543	8.85		
C_{He}	0.578	0.138	-2.54	8.70		
C_{Pe}	2.41	1.38	10.1	96.1		
C_{Lp}	0.146	0.192	-1.08	39.5		
C_{Lg}	0.165	0.00201	2.59	12.7		
C_{Sa}	3.79e-05	0.0788	-1.19	-0.891		
C_{AA}	0.0625	0.00648	0.673	0.844		
C_{FA}	4.04	0.0102	2.50	2.52		
C_{Ph}	0.874	0.000256	2.50	8.93		
C_{C_t,P_t}	0.0454	14.5	1.74	18.9		
C_{C_t,F_t}	1.91	0.253	2.41	2.41		
C_{P_t,F_t}	3.21	0.0300	8.45	8.63		
C_{C_t,P_t,F_t}	0.218	32.0	2.99	10.8		
C_{C_t,P_t,L_t}	0.0421	3.90	5.21	5.84		

Table S2[†]: Statistical analysis for model prediction of the full prediction dataset (112 biocrude yields).

Model label	Number of parameters	$Med[\epsilon]$ (wt%)	$\overline{\lvert\epsilon\rvert}$ (wt%)	$Med[\lvert\epsilon\rvert]$ (wt%)	MAPE %	% < 5 %	% < 10 %
Component additivity model	150	1.61	8.33	6.33	67.9	42.9	67.0
Reaction engineering model	56	0.950	8.04	5.53	77.2	47.3	67.0
Decision Tree Algorithm	924	1.08	6.64	4.60	43.5	52.7	80.4
Valdez and Savage [1]	22	-4.14	14.7	10.6	132	26.8	46.3
Hietala *et al.* [2]	14	7.25	14.3	11.8	75.3	27.7	43.8
Obeid *et al.* [3]	34	11.0	16.4	14.7	107	21.4	33.0
Qian *et al.* [4]	10	1.30	14.5	12.0	94.3	24.1	41.1
Obeid *et al.* [5][a]	34	16.7	21.3	16.8	106	9.82	27.7
Obeid *et al.* [5][b]	34	26.2	27.8	26.2	89.5	11.6	15.2
Saral *et al.* [6]	12	-3.20	16.6	13.2	111	22.3	35.7
Déniel *et al.* [7]	8	7.87	14.1	12.7	61.4	23.2	43.8
Yang *et al.* [8]	7	4.49	12.1	9.04	92.4	27.7	55.4
Lu *et al.* [9]	11	2.76	10.6	8.41	89.0	34.8	57.1
Yang *et al.* [10]	10	5.99	13.9	11.2	94.0	24.1	39.3
Mahadevan Subramanya and Savage [11]	30	5.12	11.9	11.2	94.2	24.1	52.7

a is the reaction network fitted to sewage sludge.
b is the reaction network fitted to pine wood.

Table S3[†]: Statistical analysis for model prediction for HTL of biomass with no lignin (88 biocrude yields).

Model label	Number of parameters	$Med[\epsilon]$ (wt%)	$\overline{\lvert\epsilon\rvert}$ (wt%)	$Med[\lvert\epsilon\rvert]$ (wt%)	MAPE %	% < 5 %	% < 10 %
Component additivity model	150	1.07	8.51	6.65	75.2	42.0	67.0
Reaction engineering model	56	0.864	7.85	5.29	83.6	48.9	67.0
Decision Tree Algorithm	924	1.50	7.29	5.38	44.7	50.0	77.3
Valdez and Savage [1]	22	-1.31	14.1	9.04	120	28.4	52.3
Valdez *et al.* [12]	20	1.91	15.9	9.28	99.4	31.8	51.1
Hietala *et al.* [2]	14	6.87	14.5	11.4	78.4	28.4	44.3
Vo *et al.* [13]	20	-3.95	15.7	11.6	104	25.0	44.3
Vo *et al.* [14]	34	-7.32	18.6	16.8	107	14.8	33.0
Sheehan and Savage [15][a]	20	-3.58	14.0	9.66	133	31.8	51.1
Sheehan and Savage [15][b]	32	-0.143	13.5	8.88	128	31.8	52.3
Palomino *et al.* [16]	6	-0.767	12.8	9.23	128	31.8	52.3
Qian *et al.* [4]	10	1.30	15.1	12.4	102	22.7	39.8
Hietala and Savage [17]	28	0.235	12.7	8.93	58.4	27.3	58.0
Obeid *et al.* [5][c]	30	14.4	21.4	17.6	127	14.8	20.5
Saral *et al.* [6]	12	-1.67	16.0	12.3	93.3	26.1	39.8
Biller and Ross [18]	3	9.90	13.3	11.4	86.3	23.9	40.9
Teri *et al.* [19][d]	6	0.0	11.2	7.49	130	35.2	63.6
Teri *et al.* [19][e]	6	1.61	11.3	7.73	125	31.8	63.6
Teri *et al.* [19][f]	12	-2.18	11.3	7.49	134	36.4	60.2
Teri *et al.* [19][g]	12	-1.66	11.3	7.40	128	34.1	59.1
Leow *et al.* [20]	3	-7.23	13.0	9.18	168	29.5	52.3
Li *et al.* [21]	3	-8.83	13.9	10.7	184	19.3	43.2
Shakya *et al.* [22]	6	-9.17	13.8	11.1	163	25.0	44.3
Hietala *et al.* [23]	10	-10.4	21.2	12.1	313	20.5	40.9
Sheng *et al.* [24]	6	-6.16	14.9	10.3	151	19.3	48.9
Wagner *et al.* [25]	12	4.98	12.2	9.52	83.3	31.8	54.5
Aierzhati *et al.* [26]	9	14.4	20.8	17.1	87.2	14.8	30.7

a is the reaction network without interactions.
b is the reaction network with interactions.
c is the reaction network fitted to microalgae.
d is the component additivity model fitted to HTL of soy protein, cornstarch, and castor oil data.
e is the component additivity model fitted to HTL of cellulose, albumin, and sunflower oil data.
f is the component additivity model in d with interactions.
g is the component additivity model in e with interactions.

We added lignin as a component for selected reaction engineering models in the literature and then reparameterized each model. For the model from Hietala and Savage [17], we added a pathway from lignin to oligolignols, Olg, to monolignols or lignols, Lgl, and these lignols can undergo decarboxylation producing biocrude and CO_2. This pathway is represented in Fig. S1†. Finally, because of the extra molecular level components in the model from Hietala and Savage [17], monomers and feedstocks that are more oligomer-like are assigned as initial condition for the corresponding model components.

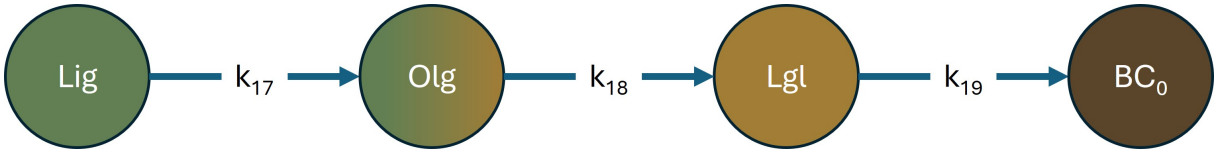

Figure S1†: The pathway added to the model from Hietala and Savage [17] where Lig: Lignin, Olg: Oligolignols, Lgl: lignols, and BC_0: biocrude with zero nitrogen content.

Table S4†: Statistics for correlating yields of solids, biocrude, aqueous, and gas wt.% from HTL using reaction engineering models and the fitting dataset (2949 yields).

Model based off of	Number of parameters	$Med[\epsilon]$ (wt%)	$\overline{\lvert\epsilon\rvert}$ (wt%)	$Med[\lvert\epsilon\rvert]$ (wt%)	MAPE %	AIC	% < 5 %	% < 10 %
Valdez *et al.* [12]	24	1.50	14.6	9.50	275	17915.6	29.9	51.7
Sheehan and Savage [15]	48	1.28	13.5	8.56	134	17533.6	32.1	53.9
Obeid *et al.* [3]	34	-0.165	13.9	10.3	176	17328.2	27.0	48.6
Hietala and Savage [17]	38	4.05	15.6	11.1	126	18255.3	26.9	46.8
This work	56	1.71	13.0	8.60	133	17360.6	33.2	55.5

Table S5†: Statistics for predicting yields of solids, biocrude, aqueous, and gas wt.% from HTL using reaction engineering models and the prediction dataset (325 yields).

Model based off of	Number of parameters	$Med[\epsilon]$ (wt%)	$\overline{\lvert\epsilon\rvert}$ (wt%)	$Med[\lvert\epsilon\rvert]$ (wt%)	MAPE %	% < 5 %	% < 10 %
Valdez *et al.* [12]	24	2.97	15.1	10.9	154	27.4	48.3
Sheehan and Savage [15]	48	2.93	13.7	8.71	128	35.1	53.2
Obeid *et al.* [3]	34	1.82	15.6	11.2	185	23.1	46.2
Hietala and Savage [17]	38	4.50	16.7	11.9	143	22.8	44.9
This work	56	2.92	13.2	9.52	135	32.9	51.4

Table S6†: Statistics for correlating biocrude wt.% only with reaction engineering models and the fitting dataset (1032 biocrude yields).

Model based off of	Number of parameters	$Med[\epsilon]$ (wt%)	$\overline{\lvert\epsilon\rvert}$ (wt%)	$Med[\lvert\epsilon\rvert]$ (wt%)	MAPE %	AIC	% < 5 %	% < 10 %
Valdez *et al.* [12]	24	-1.22	9.44	7.07	119	5355.4	36.9	65.1
Sheehan and Savage [15]	48	-1.08	8.83	6.44	118	5247.5	41.2	66.7
Obeid *et al.* [3]	34	-0.815	12.9	10.5	162	5853.1	26.7	47.6
Hietala and Savage [17]	38	1.07	12.4	9.30	93.0	5910.3	30.4	52.8
This work	56	-0.0195	7.61	5.84	112	4939.7	44.5	72.0

Table S7†: Statistics for predicting biocrude wt.% only with reaction engineering models and the prediction dataset (112 biocrude yields).

Model based off of	Number of parameters	$Med[\epsilon]$ (wt%)	$\overline{\lvert\epsilon\rvert}$ (wt%)	$Med[\lvert\epsilon\rvert]$ (wt%)	MAPE %	% < 5 %	% < 10 %
Valdez *et al.* [12]	24	0.436	10.7	8.91	98.4	32.1	55.4
Sheehan and Savage [15]	48	-0.0952	9.56	7.05	86.8	42.0	61.6
Obeid *et al.* [3]	34	3.91	14.4	11.9	116	23.2	42.9
Hietala and Savage [17]	38	0.960	14.1	10.5	70.7	17.9	50.0
This work	56	0.950	8.04	5.53	77.2	47.3	67.0

Table S8†: Statistics for initial fitting of machine learning models to the training dataset (825 biocrude yields).

Model label	Number of parameters	$Med[\epsilon]$ (wt%)	$\overline{\lvert\epsilon\rvert}$ (wt%)	$Med[\lvert\epsilon\rvert]$ (wt%)	MAPE %	AIC	% < 5 %	% < 10 %
Ridge Regression	16	0.906	9.25	8.04	118	4080	31.6	60.7
Lasso Regression	16	0.698	9.26	7.89	118	4083	31.4	60.7
Elastic Net Regression	16	0.698	9.26	7.89	118	4083	31.4	60.7
Random Forest Algorithm	5930	-0.037	2.07	1.18	17.9	13,900	90.8	97.9
Gradient Boosting Regression	175	-0.240	3.93	3.05	35.1	3097	71.6	94.2
Support Vector Machine	816	-0.000403	10.6	8.80	144	5978	30.9	56.6
k-Nearest Neighbors (k-NN)	5	-0.6	5.99	4.34	73.4	3523	54.7	82.1
Bayesian Regression	18	0.727	9.27	7.95	118	4087	30.7	60.8
Gaussian Process Regression	7	3.00e-10	0.173	3.00e-10	1.36	-160	99.2	99.8
Huber Regression	16	0.243	9.14	7.69	128	4107	35.0	63.0
Decision Forest Regression	754	0.000	0.168	0.000	1.36	1393	99.0	99.8
RANSAC Regression	16	-0.107	11.2	7.58	122	4793	37.5	61.0
Generalized Additive Model	301	0.229	6.17	4.33	75.4	4190	55.8	83.5

Table S9†: Statistics for initial prediction for machine learning models from the testing dataset (207 biocrude yields).

Model label	Number of parameters	$Med[\epsilon]$ (wt%)	$\overline{\lvert\epsilon\rvert}$ (wt%)	$Med[\lvert\epsilon\rvert]$ (wt%)	MAPE %	% < 5 %	% < 10 %
Ridge Regression	16	2.05	9.13	7.81	95.5	30.0	69.6
Lasso Regression	16	1.94	9.12	7.78	95.9	31.9	67.1
Elastic Net Regression	16	1.94	9.12	7.78	95.9	31.9	67.1
Random Forest Algorithm	5930	-0.426	4.79	2.86	41.7	65.7	87.9
Gradient Boosting Regression	175	-0.236	5.97	4.44	61.1	54.1	85.5
Support Vector Machine	816	1.05	10.5	8.16	114	28.0	60.4
k-Nearest Neighbors (k-NN)	5	-0.301	6.44	4.62	66.1	52.7	79.2
Bayesian Regression	18	1.75	9.17	7.75	96.0	29.5	67.1
Gaussian Process Regression	7	22.3	24.7	22.3	104	21.3	31.9
Huber Regression	16	1.29	8.85	7.39	100	35.7	68.1
Decision Forest Regression	754	0.000	6.55	3.19	55.9	60.9	79.7
RANSAC Regression	16	0.252	12.8	8.11	115	30.9	58.0
Generalized Additive Model	301	1.23	6.81	4.95	82.2	50.2	78.7

Table S10†: Statistics for final fitting machine learning models to the fitting dataset (1032 biocrude yields).

Model label	Number of parameters	$Med[\epsilon]$ (wt%)	$\overline{\lvert\epsilon\rvert}$ (wt%)	$Med[\lvert\epsilon\rvert]$ (wt%)	MAPE %	AIC	% < 5 %	% < 10 %
Ridge Regression	16	0.785	9.16	7.81	115	5096	33.0	62.3
Lasso Regression	16	0.751	9.18	7.86	116	5097	33.0	61.9
Elastic Net Regression	16	0.751	9.18	7.86	116	5097	33.0	61.9
Random Forest Algorithm	7339	-0.089	1.90	1.22	18.0	16,926	92.5	99.0
Gradient Boosting Regression	174	0.0111	4.02	3.03	35.8	3835	70.4	93.2
Support Vector Machine	1029	-0.0222	10.3	8.44	139	7433	31.2	58.4
k-Nearest Neighbors (k-NN)	5	-0.385	5.44	3.76	66.5	4253	60.5	83.4
Bayesian Regression	18	0.804	9.19	7.93	115	5101	32.2	62.3
Gaussian Process Regression	7	0.000	0.207	0.000	1.50	5.01	99.1	99.8
Huber Regression	16	0.416	9.00	7.48	118	5133	35.5	65.2
Decision Forest Regression	924	0.000	0.230	0.000	1.56	2050	98.5	99.8
RANSAC Regression	16	-2.00	12.4	8.54	151	6022	30.8	56.0
Generalized Additive Model	301	0.214	6.21	4.32	76.5	5115	55.1	82.8

Table S11†: Statistics for prediction from machine learning models from the prediction dataset (112 biocrude yields).

Model label	Number of parameters	$Med[\epsilon]$ (wt%)	$\overline{\lvert\epsilon\rvert}$ (wt%)	$Med[\lvert\epsilon\rvert]$ (wt%)	MAPE %	$\% < 5$ %	$\% < 10$ %
Ridge Regression	16	2.09	9.65	8.35	90.9	31.3	61.6
Lasso Regression	16	2.07	9.71	8.28	91.2	31.3	61.6
Elastic Net Regression	16	2.07	9.71	8.28	91.2	31.3	61.6
Random Forest Algorithm	7339	0.729	5.63	4.04	36.5	55.4	83.9
Gradient Boosting Regression	174	0.338	6.53	4.04	39.1	58.9	78.6
Support Vector Machine	1029	3.88	11.2	8.21	106	23.2	59.8
k-Nearest Neighbors (k-NN)	5	1.21	7.43	6.42	65.2	43.8	75.9
Bayesian Regression	18	1.80	9.65	8.23	91.5	32.1	61.6
Gaussian Process Regression	7	29.7	31.0	29.7	96.0	8.93	16.1
Huber Regression	16	1.91	9.73	7.93	91.6	33.0	61.6
Decision Forest Regression	924	1.08	6.64	4.60	43.5	52.7	80.4
RANSAC Regression	16	1.41	12.3	8.62	114	24.1	55.4
Generalized Additive Model	301	1.88	6.86	4.92	48.1	50.0	80.4

References

[1] P. J. Valdez and P. E. Savage, *Algal Research*, 2013, **2**, 416–425.

[2] D. C. Hietala, J. L. Faeth and P. E. Savage, *Bioresource Technology*, 2016, **214**, 102–111.

[3] R. Obeid, D. M. Lewis, N. Smith, T. Hall and P. van Eyk, *Chemical Engineering Journal*, 2020, **389**, 124397.

[4] L. Qian, S. Wang and P. E. Savage, *Applied Energy*, 2020, **260**, 114312.

[5] R. Obeid, N. Smith, D. M. Lewis, T. Hall and P. van Eyk, *Chemical Engineering Journal*, 2022, **428**, 131228.

[6] J. S. Saral, D. G. C. V. Reddy and P. Ranganathan, *Biomass Conversion and Biorefinery*, 2022, **1**, 1–9.

[7] M. Déniel, G. Haarlemmer, A. Roubaud, E. Weiss-Hortala and J. Fages, *Waste and Biomass Valorization*, 2017, **8**, 2087–2107.

[8] J. Yang, Q. S. He, H. Niu, K. Corscadden and T. Astatkie, *Applied Energy*, 2018, **228**, 1618–1628.

[9] J. Lu, Z. Liu, Y. Zhang and P. E. Savage, *ACS Sustainable Chemistry & Engineering*, 2018, **6**, 14501–14509.

[10] J. Yang, Q. S. He, K. Corscadden, H. Niu, J. Lin and T. Astatkie, *Applied Energy*, 2019, **233-234**, 906–915.

[11] S. Mahadevan Subramanya and P. E. Savage, *ACS Sustainable Chemistry & Engineering*, 2021, **9**, 13874–13882.

[12] P. J. Valdez, V. J. Tocco and P. E. Savage, *Bioresource Technology*, 2014, **163**, 123–127.

[13] T. K. Vo, O. K. Lee, E. Y. Lee, C. H. Kim, J.-W. Seo, J. Kim and S.-S. Kim, *Chemical Engineering Journal*, 2016, **306**, 763–771.

[14] T. K. Vo, S.-S. Kim, H. V. Ly, E. Y. Lee, C.-G. Lee and J. Kim, *Bioresource Technology*, 2017, **241**, 610–619.

[15] J. D. Sheehan and P. E. Savage, *Bioresource Technology*, 2017, **239**, 144–150.

[16] A. Palomino, L. C. Montenegro-Ruíz and R. D. Godoy-Silva, *Algal Research*, 2019, **44**, 101669.

[17] D. C. Hietala and P. E. Savage, *Chemical Engineering Journal*, 2021, **407**, 127007.

[18] P. Biller and A. B. Ross, *Bioresource technology*, 2011, **102**, 215–25.

[19] G. Teri, L. Luo and P. E. Savage, *Energy & Fuels*, 2014, **28**, 7501–7509.

[20] S. Leow, J. R. Witter, D. R. Vardon, B. K. Sharma, J. S. Guest and T. J. Strathmann, *Green Chemistry*, 2015, **17**, 3584–3599.

[21] Y. Li, S. Leow, A. C. Fedders, B. K. Sharma, J. S. Guest and T. J. Strathmann, *Green Chemistry*, 2017, **19**, 1163–1174.

[22] R. Shakya, S. Adhikari, R. Mahadevan, S. R. Shanmugam, H. Nam, E. B. Hassan and T. A. Dempster, *Bioresource Technology*, 2017, **243**, 1112–1120.

[23] D. C. Hietala, C. K. Koss, A. Narwani, A. R. Lashaway, C. M. Godwin, B. J. Cardinale and P. E. Savage, *Algal Research*, 2017, **26**, 203–214.

[24] L. Sheng, X. Wang and X. Yang, *Bioresource Technology*, 2018, **247**, 14–20.

[25] J. Wagner, R. Bransgrove, T. A. Beacham, M. J. Allen, K. Meixner, B. Drosg, V. P. Ting and C. J. Chuck, *Bioresource Technology*, 2016, **207**, 166–174.

[26] A. Aierzhati, M. J. Stablein, N. E. Wu, C.-T. Kuo, B. Si, X. Kang and Y. Zhang, *Bioresource Technology*, 2019, **284**, 139–147.