Supporting Information for:

Self-Optimizing Bayesian for Continuous Flow Synthesis Process

Runzhe Liu,[a] Zihao Wang,[b] Wenbo Yang,[a]* Jinzhe Cao,[a]* and Shengyang Tao[a]*

[a] School of Chemistry, State Key Laboratory of Fine Chemicals, Frontier Science Center for Smart Materials, Dalian Key Laboratory of Intelligent Chemistry, Dalian University of Technology, Dalian 116024, China.

Email: wbyang@dlut.edu.cn (W.Y.)

Email: ginercao@mail.dlut.edu.cn (J.C.)

Email: taosy@dlut.edu.cn (S.T.)

[b] Faculty of Information, Beijing University of Technology, Beijing 100124, China.

Index

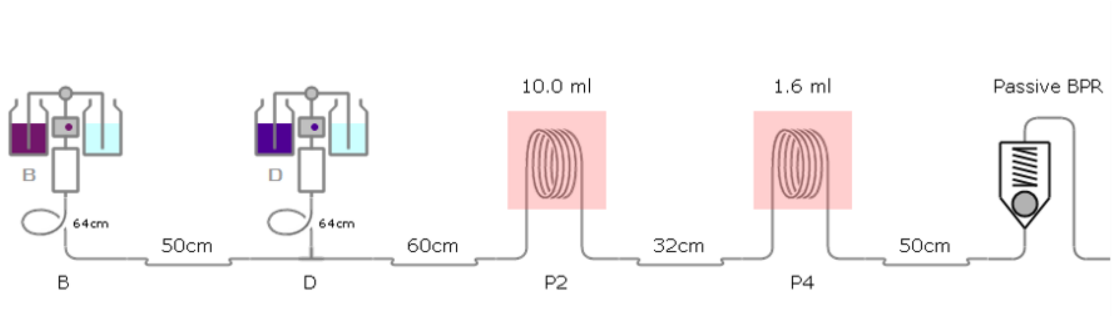

Figure S1. Schematic diagram of piping connections for continuous flow reactions with the Vapoutec R-Series Flow Synthesizer

Table S1. 14 Sets of Prior Data for Optimization Data Table

Exp #	Time / min	Tem / °C	Anisamide / eq.	Base / eq.	Ligand / mol%	Actual Yield / %	Predicted Yield / %
1	43	132	1.8	1.8	7.5	42.8	non
2	30	90	1.0	1.0	6.0	28.8	non
3	60	120	1.4	1.5	8.0	52.2	non
4	20	145	1.0	1.3	10.0	32.8	non
5	10	145	1.0	1.3	12.0	37.6	non
6	13	120	1.1	1.0	10.3	19.2	non
7	14	139	1.0	1.2	10.4	32.6	non
8	15	142	1.0	1.3	14.3	37.2	non
9	18	144	1.1	1.3	14.7	43.0	non
10	19	145	1.1	1.3	14.9	41.0	non
11	40	145	1.1	1.3	15.0	48.6	non
12	15	145	1.3	1.3	17.5	46.2	non
13	33	145	1.4	1.3	15.0	51.8	non
14	40	145	2.0	1.3	20.0	56.8	non
15	80	148	10.	2.9	19.9	71.6	111.6
16	80	148	1.9	1.0	19.9	62.8	80.2
17	80	148	1.9	2.9	6.0	45.4	88.2
18	80	148	1.0	2.9	19.9	65.5	72.2
19	80	148	1.0	2.9	19.9	63.6	69.0
20	80	41	1.3	2.7	19.9	24.6	70.6
21	80	148	1.0	1.0	19.3	57.8	73.6
22	80	148	1.9	1.0	19.9	63.3	68.0
23	80	148	1.0	1.0	19.7	57.8	65.5
24	78	147	1.9	2.6	19.9	68.1	66.5
25	80	148	1.9	2.9	6.3	50.0	76.6
26	80	148	1.9	2.9	19.9	79.1	67.4
27	80	148	1.9	2.9	19.9	74.9	70.3
28	80	148	1.9	2.9	19.9	79.3	70.4
29	80	148	1.9	2.9	19.9	79.1	70.2

Table S2. 5 Sets of Prior Data for Optimization Data Table

Exp #	Time / min	Tem / °C	Anisamide / eq.	Base / eq.	Ligand / mol%	Actual yield / %	Predicted Yield / %
1	10	145	1.0	1.3	12.0	37.6	non
2	13	120	1.1	1.0	10.3	19.2	non
3	14	139	1.0	1.2	10.4	32.6	non
4	15	142	1.0	1.3	14.3	37.2	non
5	15	145	1.3	1.3	17.5	46.2	non
6	2	30	1.9	1.0	19.9	21.2	366.4
7	2	148	1.0	2.9	19.9	42.2	148.0
8	80	148	1.9	2.9	19.9	74.2	127.1
9	80	148	1.9	2.9	19.9	80.6	74.1
10	80	148	1.9	1.0	19.9	62.8	80.9
11	2	148	1.9	2.9	19.9	43.0	77.1
12	80	148	1.0	2.9	19.9	63.4	76.4
13	80	148	1.9	1.0	19.9	63.3	78.9
14	80	148	1.0	2.9	19.9	65.5	77.1
15	80	148	1.9	2.9	19.9	79.7	73.7
16	80	148	1.9	2.9	19.9	78.1	75.4
17	80	148	1.9	2.9	6.0	45.4	77.7
18	80	148	1.9	2.9	19.9	79.1	72.9
19	80	148	1.9	2.9	19.9	74.9	73.6
20	80	148	1.9	2.9	19.9	79.3	74.2
21	80	148	1.9	2.9	19.9	79.1	74.1

Table S3. Table of Optimized Model Parameters for 14 Sets of Prior Data

Exp#	Best Kernel	MSE	Best Byperparameters
15	DotProduct(sigma_0=1)	3.186	OrderedDict([('alpha',0.006636085077612337 5),('n_restarts_optimizer', 16)])
16	Matern(length_scale=1, nu=1.5)	5.288	OrderedDict([('alpha',0.0066360850776123375),('n_restarts_optimizer', 16)])
17	RationalQuadratic(alpha=1,length_scale=1)	17.803	OrderedDict([('alpha', 0.0066360850776123375), ('n_restarts_optimizer', 16)])
18	RationalQuadratic(alpha=1,length_scale=1)	15.904	OrderedDict([('alpha', 1.0340016434251914e-06), ('n_restarts_optimizer', 17)])
19	RationalQuadratic(alpha=1,length_scale=1)	13.615	OrderedDict([('alpha', 0.0066360850776123375), ('n_restarts_optimizer', 16)])
20	RBF(length_scale=1)	18.923	OrderedDict([('alpha', 0.0066360850776123375), ('n_restarts_optimizer', 16)])
21	Matern(length_scale=1, nu=1.5)	1.067	OrderedDict([('alpha', 0.0066360850776123375), ('n_restarts_optimizer', 16)])
22	Matern(length_scale=1, nu=1.5)	4.538	OrderedDict([('alpha', 0.0066360850776123375), ('n_restarts_optimizer', 16)])
23	Matern(length_scale=1, nu=1.5)	22.253	OrderedDict([('alpha', 0.0066360850776123375), ('n_restarts_optimizer', 16)])
24	RationalQuadratic(alpha=1,length_scale=1)	0.386	OrderedDict([('alpha', 0.0066360850776123375), ('n_restarts_optimizer', 16)])
25	RationalQuadratic(alpha=1,length_scale=1)	3.364	OrderedDict([('alpha', 0.0066360850776123375), ('n_restarts_optimizer', 16)])
26	RationalQuadratic(alpha=1,length_scale=1)	9.107	OrderedDict([('alpha', 0.0066360850776123375), ('n_restarts_optimizer', 16)])
27	RationalQuadratic(alpha=1,length_scale=1)	9.474	OrderedDict([('alpha', 0.0066360850776123375), ('n_restarts_optimizer', 16)])
28	DotProduct(sigma_0=1)	1.430	OrderedDict([('alpha', 6.0163078295899284e-05), ('n_restarts_optimizer', 19)])
29	Matern(length_scale=1, nu=1.5)	6.772	OrderedDict([('alpha', 0.0066360850776123375), ('n_restarts_optimizer', 16)])

The table shows the MSE of the best kernel as well as the hyperparameters, e.g., the 15th set of experiments was calculated using the data from the previous 14 sets of experiments, where the best kernel was DotProduct (sigma_0=1) with an MSE of 3.186 and the best hyperparameters were OrderedDict([('alpha', 0.0066360850776123375), ('n_restarts_optimizer', 16)])

Table S4. Table of Optimized Model Parameters for 5 Sets of Prior Data

Exp#	Best Kernel	MSE	Best Byperparameters
6	RationalQuadratic(alpha=1, length_scale=1)	97.922	OrderedDict([('alpha',0.0066360850776123375),('n_restarts_optimizer', 16)])
7	Matern(length_scale=1, nu=1.5)	9.713	OrderedDict([('alpha', 0.0066360850776123375), ('n_restarts_optimizer', 16)])
8	RationalQuadratic(alpha=1, length_scale=1)	1.263	OrderedDict([('alpha', 1.0340016434251914e-06), ('n_restarts_optimizer', 17)])
9	Matern(length_scale=1, nu=1.5)	8.440	OrderedDict([('alpha', 0.0066360850776123375), ('n_restarts_optimizer', 16)])
10	RBF(length_scale=1)	34.442	OrderedDict([('alpha', 0.0066360850776123375), ('n_restarts_optimizer', 16)])
11	RationalQuadratic(alpha=1, length_scale=1)	40.960	OrderedDict([('alpha', 1.0340016434251914e-06), ('n_restarts_optimizer', 17)])
12	RationalQuadratic(alpha=1, length_scale=1)	21.496	OrderedDict([('alpha', 1.0340016434251914e-06), ('n_restarts_optimizer', 17)])
13	Matern(length_scale=1, nu=1.5)	4.282	OrderedDict([('alpha', 0.0066360850776123375), ('n_restarts_optimizer', 16)])
14	RationalQuadratic(alpha=1, length_scale=1)	30.208	OrderedDict([('alpha', 4.3693399475103194e-05), ('n_restarts_optimizer', 16)])
15	RationalQuadratic(alpha=1, length_scale=1)	2.446	OrderedDict([('alpha', 0.0066360850776123375), ('n_restarts_optimizer', 16)])
16	RationalQuadratic(alpha=1, length_scale=1)	2.442	OrderedDict([('alpha', 0.0066360850776123375), ('n_restarts_optimizer', 16)])
17	DotProduct(sigma_0=1)	83.223	OrderedDict([('alpha', 0.0066360850776123375), ('n_restarts_optimizer', 16)])
18	DotProduct(sigma_0=1)	35.948	OrderedDict([('alpha', 0.0066360850776123375), ('n_restarts_optimizer', 16)])
19	DotProduct(sigma_0=1)	35.756	OrderedDict([('alpha', 0.0001491462267977867), ('n_restarts_optimizer', 19)])
20	RationalQuadratic(alpha=1, length_scale=1)	35.631	OrderedDict([('alpha', 1.0340016434251914e-06), ('n_restarts_optimizer', 17)])
21	RationalQuadratic(alpha=1, length_scale=1)	1.625	OrderedDict([('alpha', 4.3693399475103194e-05), ('n_restarts_optimizer', 16)])

The table shows the MSE of the best kernel as well as the hyperparameters, e.g., the 6th set of experiments was calculated using the data from the previous 5 sets of experiments, where the best kernel was RationalQuadratic(alpha=1, length_scale=1) with an MSE of 97.922 and the best hyperparameters were OrderedDict([('alpha',0.0066360850776123375),('n_restarts_optimizer', 16)])

Computerized benchmarking of algorithms

Based on the Arrhenius equation, we have constructed a chemical reaction model for computational benchmarking of algorithms. In this section, we compare our designed SOBayesian algorithm with the Dragonfly algorithm. The testing was conducted on a model that optimizes the yield as a single objective, with reaction temperature and reaction time as independent variables. The model was constructed according to the following formula, and the yield was calculated using first-order reaction kinetics.

$$k = Ae^{\frac{-Ea}{RT}}$$

$$Yield = (1 - e^{-kt}) \times 100$$

Where k is the rate constant, A is the frequency factor, Ea is the activation energy, R is the ideal gas constant, T is the absolute temperature, and t is the reaction time. The full parameter space search range is set to (time: 10–120min, temperature: 25–75°C), and the quarter parameter space search range is (time: 10–65min, temperature: 25–50°C).

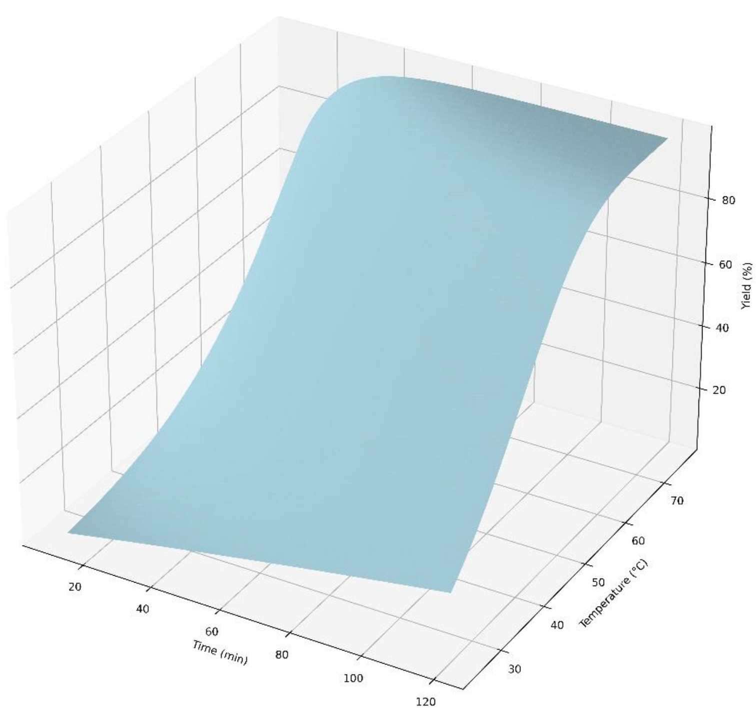

Figure S2. Optimizing the full parameter space of a model

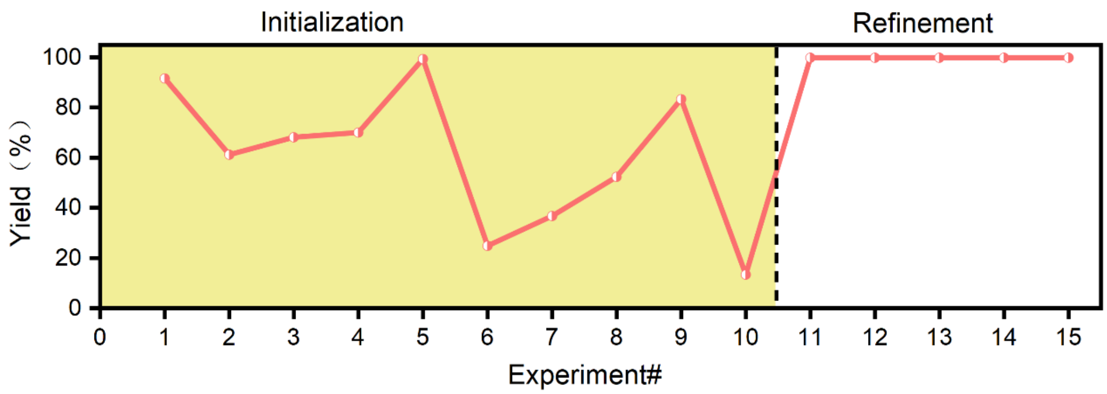

Figure S3. Full-space random sampling using SOBayesian

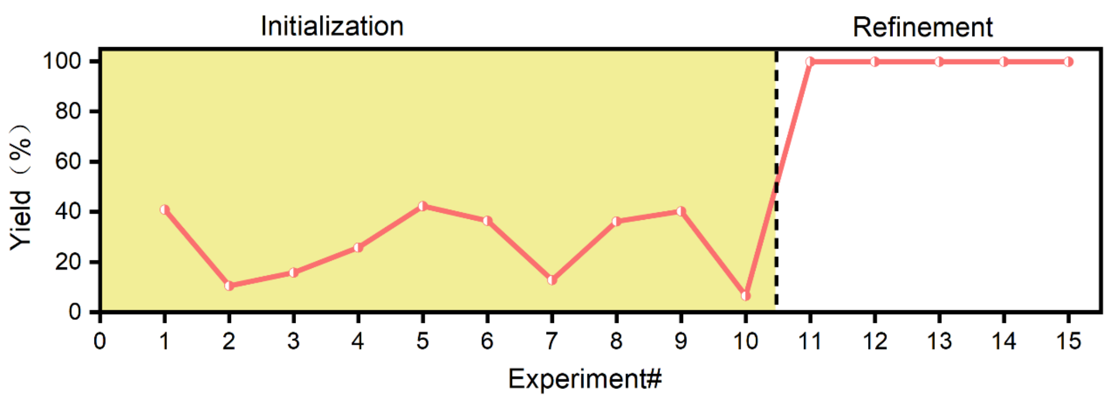

Figure S4. Quarter-space random sampling using SOBayesian

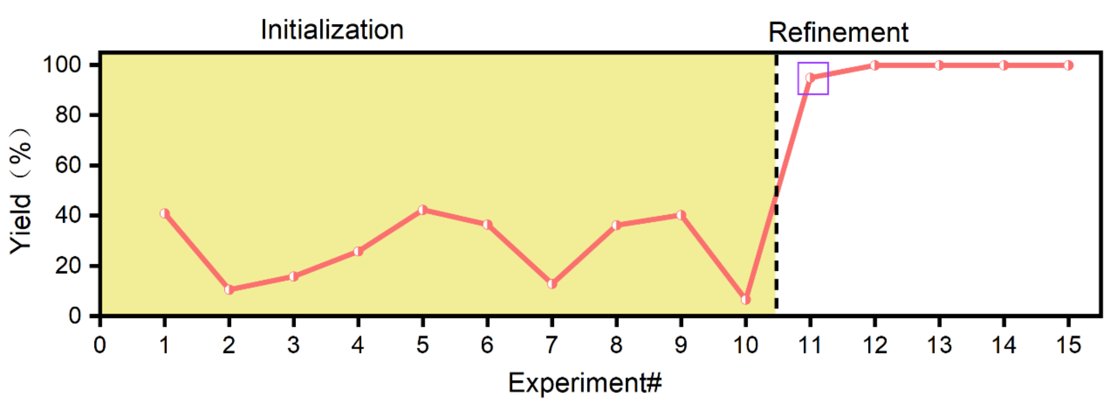

Figure S5. Quarter-space random sampling using SOBayesian（Includes error data）

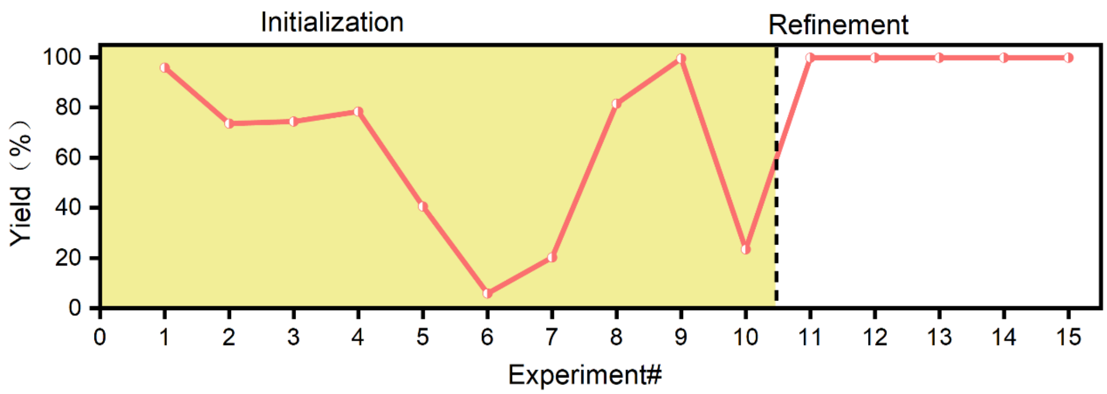

Figure S6. Full-space Latin hypercube sampling using SOBayesian

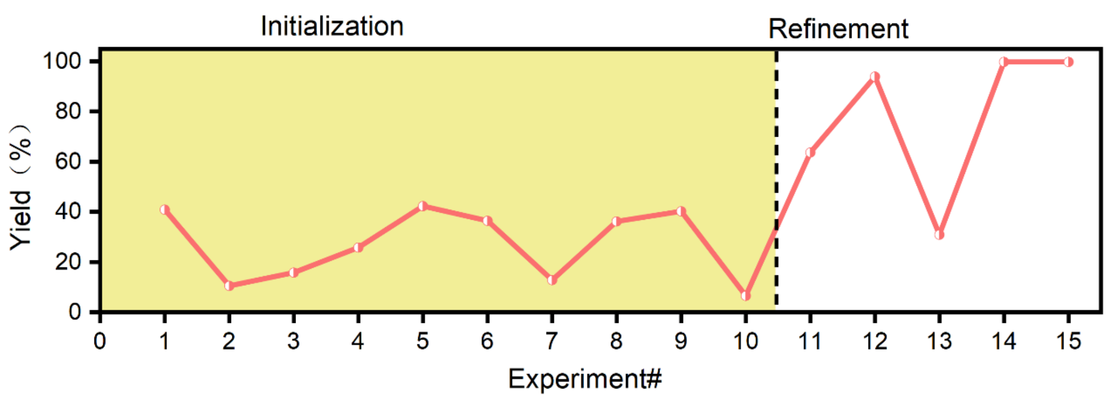

Figure S7. Quarter-space random sampling using Dragonfly

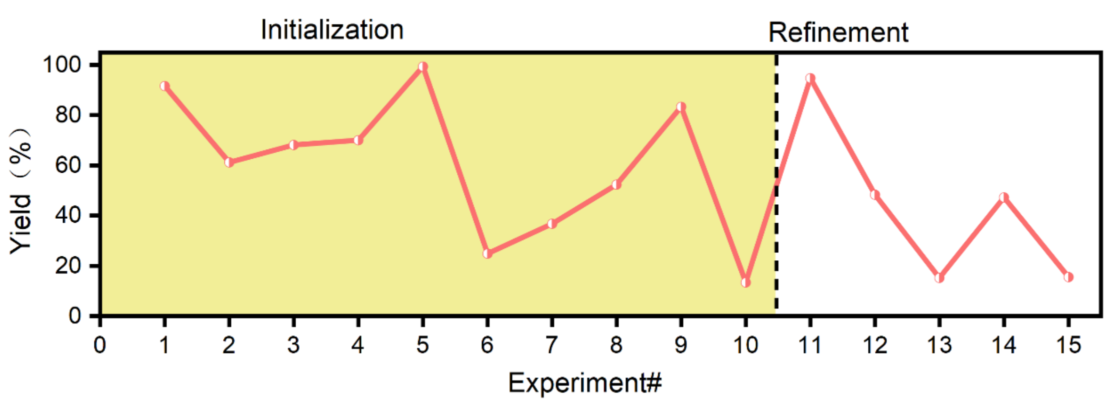

Figure S8. Full-space random sampling using Dragonfly

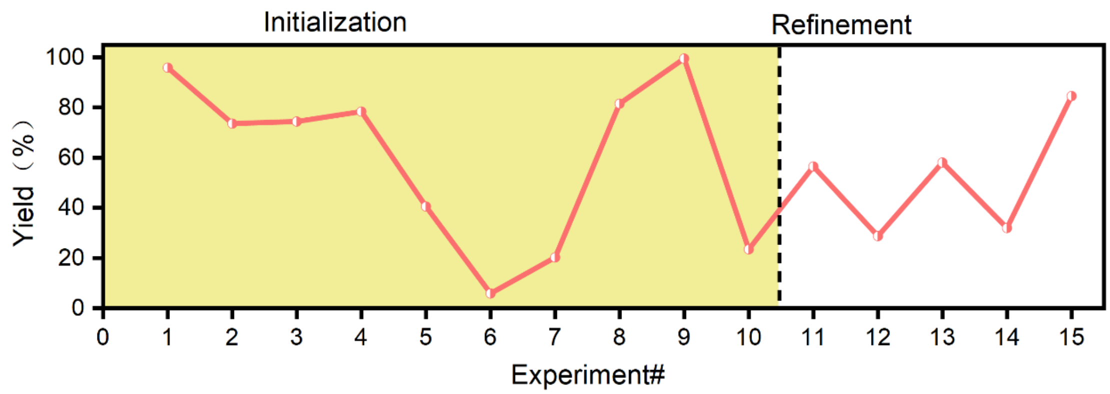

Figure S9. Full-space Latin hypercube sampling using Dragonfly

Table S5. Full-space random sampling using SOBayesian optimized data

Exp#	Time (min)	Temperature (℃)	Yield (%)	predicted value
1	112.22222	54.7979798	91.68316	
2	44.444444	54.29292929	61.30069	
3	20	68.43434343	68.27146	
4	85.555556	48.73737374	70.18882	
5	105.55556	66.41414141	99.49449	
6	18.888889	49.74747475	25.04631	
7	74.444444	38.13131313	36.8923	
8	94.444444	41.16161616	52.4791	
9	75.555556	55.80808081	83.49493	
10	27.777778	36.11111111	13.56466	
11	120	75	100	128.9
12	120	75	100	113.65
13	120	75	100	113.45
14	120	75	100	109.5
15	120	75	100	107.39

Table S6. Quarter-space random sampling using SOBayesian optimized data

Exp#	Time (min)	Temperature (℃)	Yield (%)	predicted value
1	36.666667	48.98989899	41.06085	
2	52.777778	25.50505051	10.67239	
3	61.111111	28.78787879	15.94564	
4	30	44.19191919	25.8952	
5	45.555556	46.71717172	42.46581	
6	48.333333	43.43434343	36.56735	
7	58.333333	26.76767677	13.00093	
8	34.444444	47.72727273	36.32587	
9	52.777778	43.93939394	40.36865	
10	16.666667	33.33333333	6.719041	
11	120	75	100	119.33
12	120	75	100	101.91
13	120	75	100	101.12
14	120	75	100	100.57
15	120	75	100	100.38

Table S7. Quarter-space random sampling using SOBayesian optimized data(Includes error data)

Exp#	Time (min)	Temperature (℃)	Yield (%)	predicted value
1	36.666667	48.98989899	41.06085	
2	52.777778	25.50505051	10.67239	
3	61.111111	28.78787879	15.94564	
4	30	44.19191919	25.8952	
5	45.555556	46.71717172	42.46581	
6	48.333333	43.43434343	36.56735	
7	58.333333	26.76767677	13.00093	
8	34.444444	47.72727273	36.32587	
9	52.777778	43.93939394	40.36865	
10	16.666667	33.33333333	6.719041	
11	120	75	95	119.33
12	120	75	100	97.35
13	120	75	100	101
14	120	75	100	98
15	120	75	100	98.5

***Red color is human input error data, the real value is 100**

Table S8. Full-space Latin hypercube sampling using SOBayesian optimized data

Exp#	Time (min)	Temperature (℃)	Yield (%)	predicted value
1	79.104833	63.39847226	96.01944	
2	47.012046	58.25865231	73.71803	
3	74.074571	52.31249795	74.54596	
4	111.7994	48.37339799	78.53282	
5	62.188869	41.97700279	40.69459	
6	16.085575	32.40808484	6.024781	
7	88.361538	27.6598472	20.3815	
8	24.903954	71.00778837	81.663	
9	101.26575	68.1683552	99.66862	
10	41.034061	38.86084161	23.59186	
11	119	75	100	137.42
12	120	75	100	113.7
13	120	75	100	117.38
14	120	75	100	112.24
15	120	75	100	109.46

Table S9. Full-space random sampling using Dragonfly optimized data

Exp#	Time (min)	Temperature (℃)	Yield (%)	predicted value
1	54.79798	112.2222	91.68315779	
2	54.29293	44.44444	61.30068847	
3	68.43434	20	68.27145842	
4	48.73737	85.55556	70.18882037	
5	66.41414	105.5556	99.49448548	
6	49.74747	18.88889	25.04630696	
7	38.13131	74.44444	36.89229896	
8	41.16162	94.44444	52.4790997	
9	55.80808	75.55556	83.49492992	
10	36.11111	27.77778	13.56466235	
11	69.4208	48.4686	94.88	13.59078971
12	52.7814	34.64461	48.44	91.82506587
13	42.63544	18.82434	15.34	47.57556277
14	49.82325	41.84347	47.39	15.93938494
15	28.09313	63.66097	15.67	46.89615898

Table S10. Quarter-space random sampling using Dragonfly optimized data

Exp#	Time (min)	Temperature (℃)	Yield (%)	predicted value
1	48.9899	36.66667	41.06085371	
2	25.50505	52.77778	10.67239167	
3	28.78788	61.11111	15.94564229	
4	44.19192	30	25.89519853	
5	46.71717	45.55556	42.46580546	
6	43.43434	48.33333	36.56735459	
7	26.76768	58.33333	13.00092791	
8	47.72727	34.44444	36.32587499	
9	43.93939	52.77778	40.36864628	
10	33.33333	16.66667	6.719041148	
11	50.88222	61.15377	63.8	98.63242936
12	56.64995	111.1941	94.02	61.98361048
13	41.74114	45.0976	31.06	93.34251994
14	74.49987	78.89305	99.88	30.77780044
15	72.77515	86.55885	99.87	99.48475163

Table S11. Full-space Latin hypercube sampling using Dragonfly optimized data

Exp#	Time (min)	Temperature (℃)	Yield (%)	predicted value
1	63.39847	79.10483	96.01944277	
2	58.25865	47.01205	73.71803277	
3	52.3125	74.07457	74.54595961	
4	48.3734	111.7994	78.53282252	
5	41.977	62.18887	40.69458584	
6	32.40808	16.08557	6.024781104	
7	27.65985	88.36154	20.3815012	
8	71.00779	24.90395	81.66300029	
9	68.16836	101.2658	99.6686218	
10	38.86084	41.03406	23.5918594	
11	51.50317	48.04056	56.65	25.50481269
12	39.33475	50.02496	28.88	59.07126091
13	60.31874	26.51596	58.19	29.10699484
14	52.00571	21.48252	32.15	64.0154715
15	52.63696	99.25099	84.71	38.99264697