Electronic Supplementary Material (ESI) for Environmental Science: Nano.
This journal is © The Royal Society of Chemistry 2023

Supporting Information

Potential threat of microplastic to humans: toxicity prediction modeling by small data analysis

Daheui Choi[a,#], Chul Kim[b,#], Taihyun Kim[a], Kyungtae Park[a], Jongho Im[b,c,*], and Jinkee Hong[a,*]

Table S1. The results of performing confirmative latent Dirichlet allocation (CLDA) on 731 discovered scientific articles.

No.	Topic 1: Microplastic		Topic 2: Human Toxicity	
	Word	**Association Degree**	**Word**	**Association Degree**
1	Marine	284	Environment	172
2	Environment	269	Water	121
3	Water	210	Particle	101
4	Particle	110	Concentration	100
5	Sea	100	Effect	80
6	Pollution	100	Marine	80
7	Sediment	100	Exposure	75
8	Debris	95	Surface	60
9	Soil	35	Soil	31
10	River	29	Increase	27

The data are the top 10 words highly related to the topic. The larger the value of the association degree, the more relevant it is to the topic.

Table S2. Experimental condition for 3 different microparticles

Exp. No	Particle Concentration (μg/mL)			Number of microparticles (*n*)		
	PS	**PVC**	**ABS**	**PS**	**PVC**	**ABS**
1	33.33	33.33	33.33	184.14 ± 24.48	39.90 ± 5.04	21.38 ± 9.81
2	50	50	0	276.22 ± 36.72	59.86 ± 7.56	0
3	50	0	50	276.22 ± 36.72	0	32.07 ± 14.71
4	0	50	50	0	59.86 ± 7.56	32.07 ± 14.71
5	100	0	0	553.43 ± 36.72	0	0
6	0	100	0	0	119.71 ± 15.13	0
7	0	0	100	0	0	64.13 ± 29.43

PS, polystyrene; PVC, poly(vinyl chloride); ABS, acrylonitrile butadiene styrene.

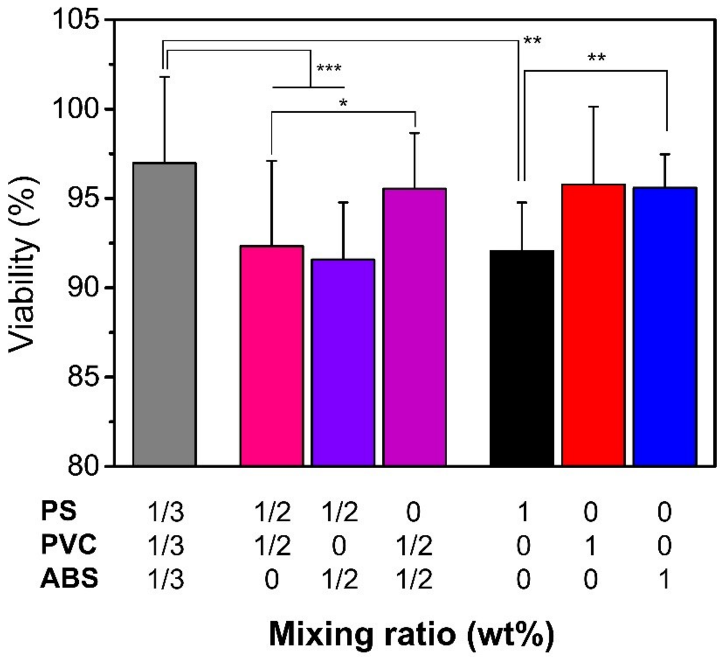

Figure S1. Normalized viability (%) of each microplastic combination (n= 37~50). The significance between two groups was denoted as *, **, *** for the $p<0.05$, $p<0.01$, $p<0.001$, respectively.

Table S3. The data obtained under the simplex-centroid design.

Design Point	Component fraction PS	PVC	ABS	Observed Response Values ($N = 314$)	Average Response Value $\bar{y}$
1	1/3	1/3	1/3	104.06, 92.85, 95.19, 92.38, 79.30, 93.08, 85.07, 88.86, 90.13, 96.46, 112.86, 89.81, 89.2, 86.17, 90.41, 115.16, 113.8, 85.65, 94.28, 91.56, 104.13, 97.83, 98.88, 93.28, 88.37, 127.60, 114.12, 90.75, 104.97, 106.13, 72.88, 81.72, 73.36, 76.72, 93.84, 107.5, 103.48, 97.90, 93.49, 93.36, 99.37, 104.30, 106.06, 96.66, 103.83, 118.53, 101.05, 101.79, 108.37, 95.97 ($n_1 = 50$)	96.9
2	1/2	1/2	0	101.49, 87.71, 77.44, 93.55, 81.17, 95.19, 87.60, 90.13, 90.55, 104.05, 105.58, 81.31, 89.20, 80.10, 82.52, 106.08, 101.09, 92.01, 98.82, 95.64, 105.53, 93.28, 96.08, 92.23, 89.77, 105.39, 86.23, 85.44, 75.92, 91.67, 82.69, 76.43, 72.88, 79.99, 77.78, 106.59, 94.40, 101.27, 90.37, 99.20, 96.43, 97.6, 83.16, 87.86, 87.62,	92.1

				110.91, 102.99, 100.00, 104.33 ($n_2 = 49$)	
3	1/2	0	1/2	102.66, 88.18, 87.25, 83.27, 79.54, 93.93, 93.93, 91.82, 106.16, 93.5, 112.26, 86.17, 94.05, 86.77, 83.74, 109.26, 99.27, 98.82, 82.48, 87.47, 104.83, 94.33, 95.03, 90.47, 83.47, 105.33, 82.39, 79.10, 74.28, 92.28, 84.71, 76.43, 79.32, 90.38, 85.09, 103.09, 92.19, 92.19, 93.49, 93.49, 96.08, 91.73, 84.21, 73.53, 87.86, 114.34, 103.29, 87.45, 91.48, 90.74 ($n_3 = 50$)	91.5
4	0	1/2	1/2	103.36, 99.86, 94.25, 89.35, 93.32, 97.72, 94.35, 98.57, 106.58, 100.67, 118.33, 93.45, 92.84, 92.84, 94.05, 116.07, 102.91, 97.91, 91.10, 97.00, 115.69, 97.83, 91.87, 97.13, 98.88, 85.93, 72.26, 64.57, 69.76, 86.72, 117.51, 96.25, 86.34, 89.42, 94.52, 105.29, 97.25, 92.58, 106.2, 103.48, 97.13, 99.25, 85.86, 88.09, 96.31, 114.79, 99.70,, 68.92, 95.37, 88.35 ($n_4 = 50$)	95.6
5	1	0	0	96.88, 90.55, 81.69, 90.55, 90.55,	92.2

				86.17, 84.34, 83.74, 80.10, 95.87, 109.26, 96.55, 92.01, 88.83, 89.74, 96.78, 92.93, 101.33, 103.08, 95.77, 96.15, 107.69, 110.77, 88.56, 76.75, 87.39, 96.08, 90.76, 74.00, 83.04, 80.34, 79.75, 112.40, 97.61, 89.99, 92.53, 99.4 ($n_5 = 37$)	
6	0	1	0	100.67, 81.27, 88.86, 89.29, 98.57, 90.41, 72.21, 81.92, 74.03, 92.84, 114.71, 109.72, 101.54, 97.46, 101.54, 117.79, 87.67, 104.83, 100.28, 103.08, 78.84, 79.51, 85.67, 88.27, 106.20, 101.53, 87.26, 103.61, 93.10, 110.41, 93.85, 98.66, 106.65, 104.89, 106.42, 98.36, 93.28, 93.73, 101.79 ($n_6 = 39$)	95.9
7	0	0	1	89.71, 84.22, 87.6, 88.44, 99.83, 93.45, 93.45, 101.33, 82.52, 94.66, 112.89, 92.92, 102.00, 98.37, 95.19, 111.49, 99.93, 103.78, 107.99, 96.08, 88.75, 90.19, 90.86, 97.40, 90.24, 101.79, 98.03, 93.10, 99.00, 95.61, 99.37, 95.02, 95.49, 93.26, 94.17, 89.84, 91.93, 94.62, 92.08	95.6

$(n_7 = 39)$

PS, polystyrene; PVC, poly(vinyl chloride); ABS, acrylonitrile butadiene styrene.

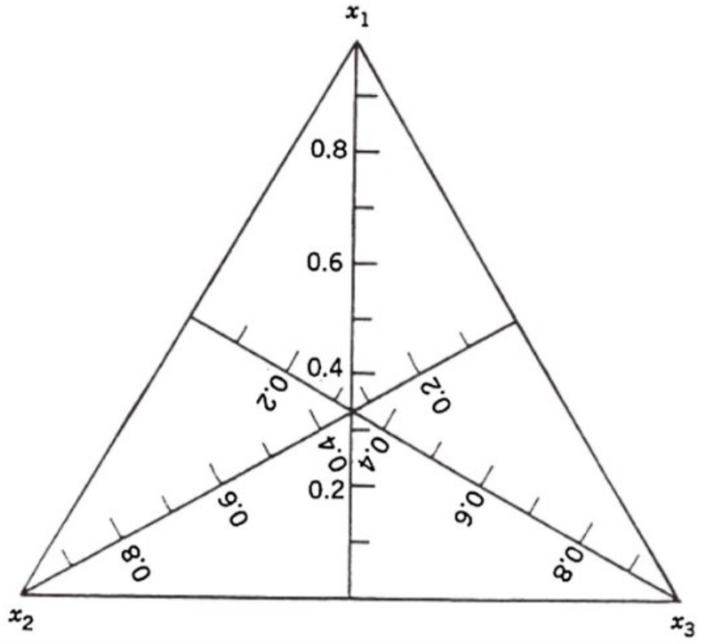

Figure S2. The simplex coordinate system.

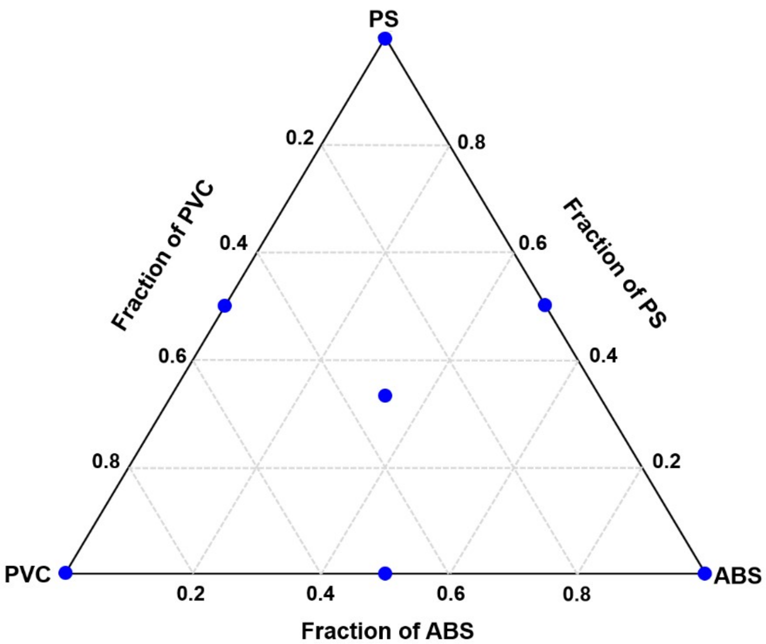

Figure S3. The simplex-centroid design. The blue dots mean the seven experimental points (fraction of three microplastics) at which the experiment will be performed.

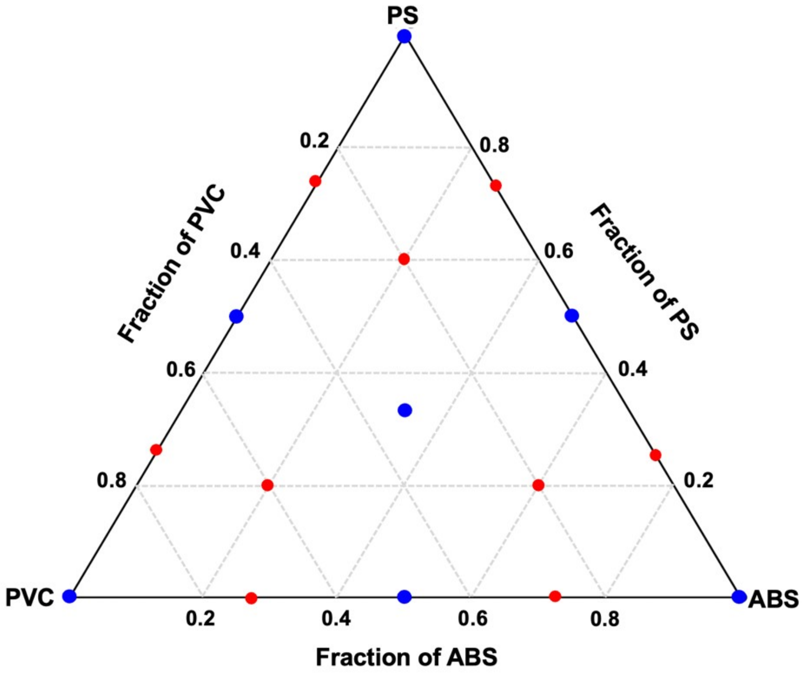

Figure S4. Nine experimental points (the red dots) for test data.

Table S4. Summary of statistical inference for the quadratic model.

Parameter	**Estimate**	**Standard Error**	**p-Value**
β_1	91.77	1.65	$< 2.00 \times 10^{-16}$
β_2	95.54	1.60	$< 2.00 \times 10^{-16}$
β_3	95.19	1.60	$< 2.00 \times 10^{-16}$
β_{12}	-1.29	6.89	0.852
β_{13}	-3.41	6.86	0.619
β_{23}	5.41	6.82	0.428

F-statistic: 4572 on 6 and 308 degrees of freedom; p-value: $< 2.20 \times 10^{-16}$

Shapiro-Wilk normality test: W = 0.99, p-value = 0.21

Multiple $R^2 = 0.9889$, Adjusted $R^2 = 0.9887$,

Root Mean Square Error (RMSE) for test data = 0.45 (%)

Mean Absolute Error (MAE) for test data = 0.36 (%)

Table S5. Summary of statistical inference for the linear model.

Parameter	**Estimate**	**Standard Error**	**p-Value**
β_1	91.16	1.30	$< 2.00 \times 10^{-16}$
β_2	96.03	1.28	$< 2.00 \times 10^{-16}$
β_3	95.42	1.28	$< 2.00 \times 10^{-16}$

F-statistic: 9204 on 3 and 311 degrees of freedom; p-value: $< 2.20 \times 10^{-16}$

Shapiro-Wilk normality test: W = 0.99, p-value = 0.24

Multiple $R^2 = 0.9889$, Adjusted $R^2 = 0.9888$

Root Mean Square Error (RMSE) for test data = 0.72 (%)

Mean Absolute Error (MAE) for test data = 0.57 (%)

Table S6. The test data to evaluate the predictive performance of our linear model.

Experimental group	Component fraction			Observed Response Values ($N = 46$)	Average Response Value $\bar{y}$	Predicted Response Value $\hat{y}$
	PS	PVC	ABS			
1	0	0.25	0.75	102.53, 113.23, 111.54, 85.31, 74.50, 90.66 ($n_1 = 6$)	96.30	95.57
2	0	0.75	0.25	109.79, 94.20, 102.56, 89.74, 96.18, 84.86 ($n_2 = 6$)	96.22	95.88
3	0.2	0.2	0.6	99.96, 108.41, 104.35, 80.91, 78.11 ($n_3 = 5$)	94.35	94.69
4	0.2	0.6	0.2	98.19, 108.23, 81.61, 92.08 ($n_4 = 4$)	95.03	94.94
5	0.25	0	0.75	110.38, 109.92, 102.37, 76.37, 65.12 ($n_5 = 5$)	92.83	94.35
6	0.25	0.75	0	88.78, 97.05,	93.74	94.82

				96.95, 92.53, 93.36 ($n_6 = 5$)		
7	0.6	0.2	0.2	108.44, 76.74, 108.10, 104.27, 88.71, 67.82 ($n_7 = 6$)	92.35	92.99
8	0.75	0	0.25	106.55, 105.01, 85.31, 74.50, 90.66 ($n_8 = 5$)	92.41	92.23
9	0.75	0.25	0	97.33, 91.29, 96.36, 85.22 ($n_9 = 4$)	92.55	92.38

Root Mean Square Error (RMSE) for test data = 0.448(%)

Mean Absolute Error (MAE) for test data = 0.357(%)

Appendix 1.

As shown in eq (1), the simplex-centroid design considering the second-order interaction we used can be generalized when there are three or more variables.

$$y = \sum_{i=1}^{m} \beta_i x_i + \sum\sum_{i<j}^{m} \beta_{ij} x_i x_j + \epsilon \quad (1)$$

Where m is the number of independent variables and $\sum\sum_{i<j}^{m} \beta_{ij} x_i x_j$ means that adding all possible combinations where $i > j$. In addition, the constraints for each independent variable are generalized as follows.

$$x_i \geq 0,\, i = 1,2,\ldots,m \quad (2)$$

$$\sum_{i=1}^{m} x_i = 1 \quad (3)$$

Due to the above constraints, the experimental space is not defined in the m-dimensional orthogonal coordinate system but in the (m-1)-dimensional simplex coordinate system. a simplex is a generalization of the concept of a triangle or tetrahedron to any dimension in geometry. For example, as shown in the following figure, the 1D simplex means a line, the 2D simplex (experimental space of our model) means a triangle, and the 3D simplex means a tetrahedron.

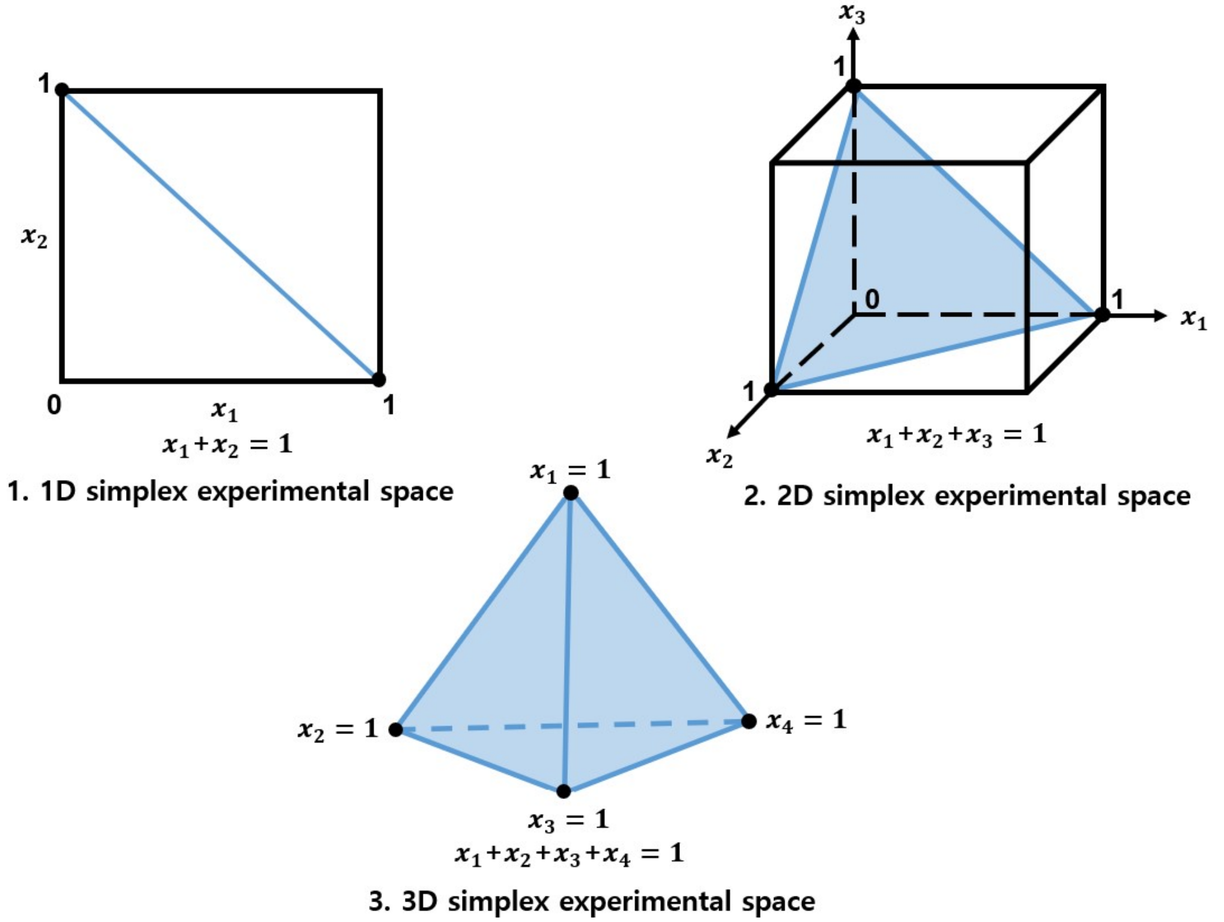

Figure A1. Simplex experiment space with 2, 3, and 4 independent variables.