Supplementary Materials

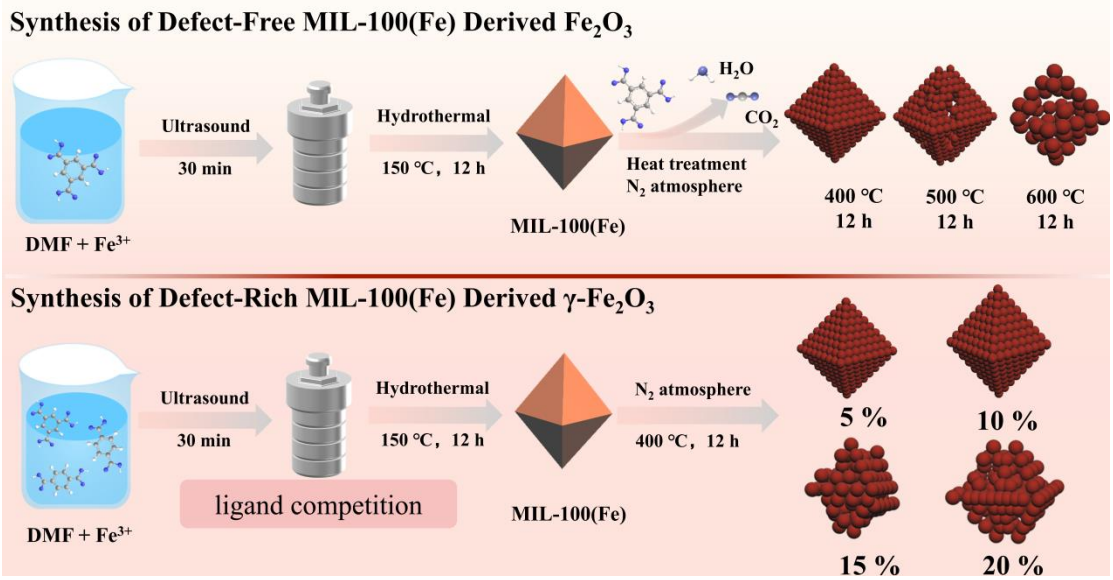
Defect engineering and machine learning enhance *n*-butanol sensing of γ-Fe_2O_3

Liang-Bo Bo[#], Zhi-Lei Li[#], Xiao-Hong Zheng[*], Su-Yue Ren, Zi-Qi Gu, Jing-Wen Pan, Yu-Feng Liu[*]

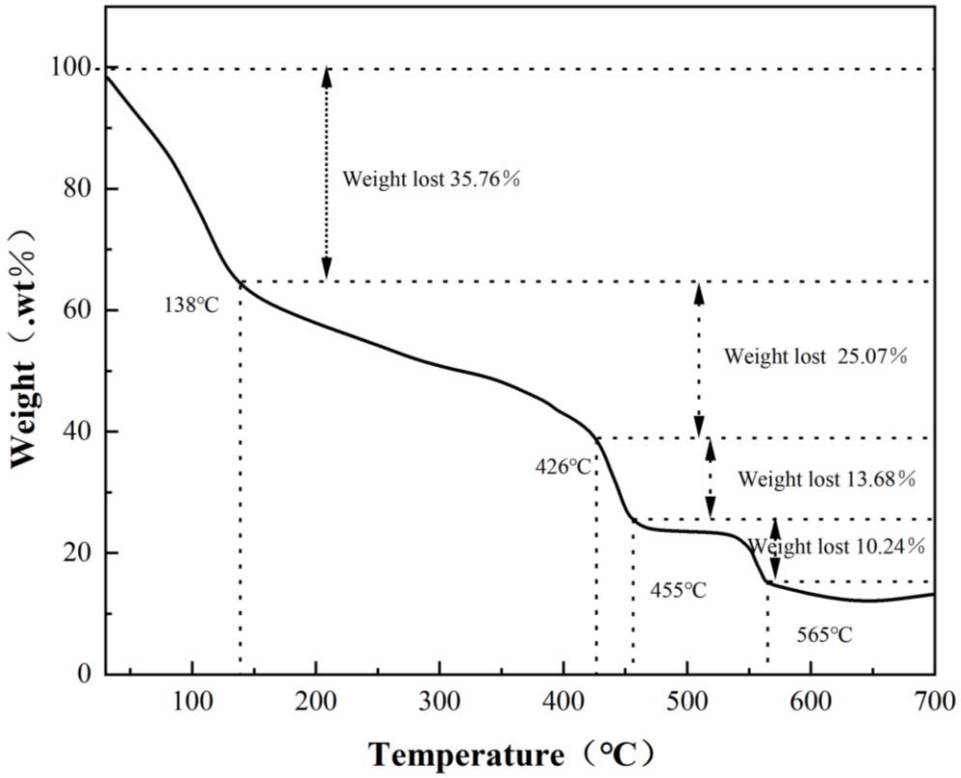
Faculty of Materials Technology, Shanghai Institute of Technology, Shanghai 201418, China.

[#]These authors contributed equally to this work.

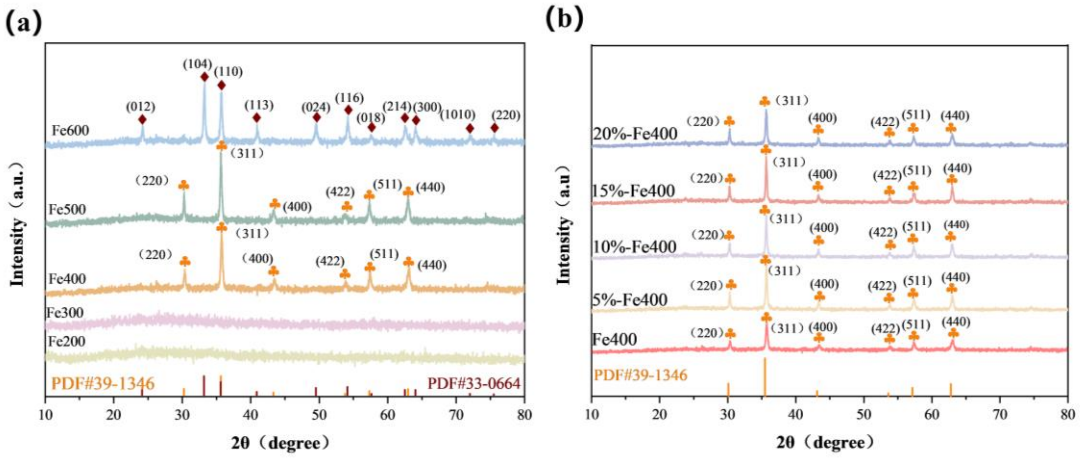
[*]**Correspondence to:** Xiao-Hong Zheng, Yu-Feng Liu, Faculty of Materials Technology, Shanghai Institute of Technology, Shanghai 201418, China. E-mail: zhengxiaohong@sit.edu.cn; yfliu@sit.edu.cn



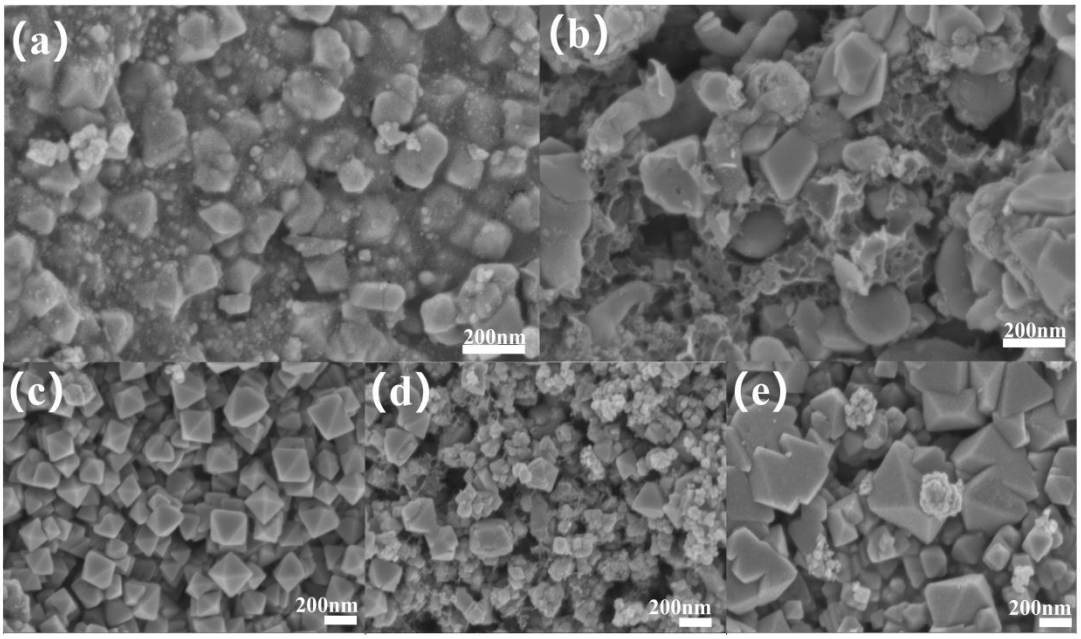
Supplementary Figure 1. Schematic depiction outlining the synthesis of Fe_2O_3.



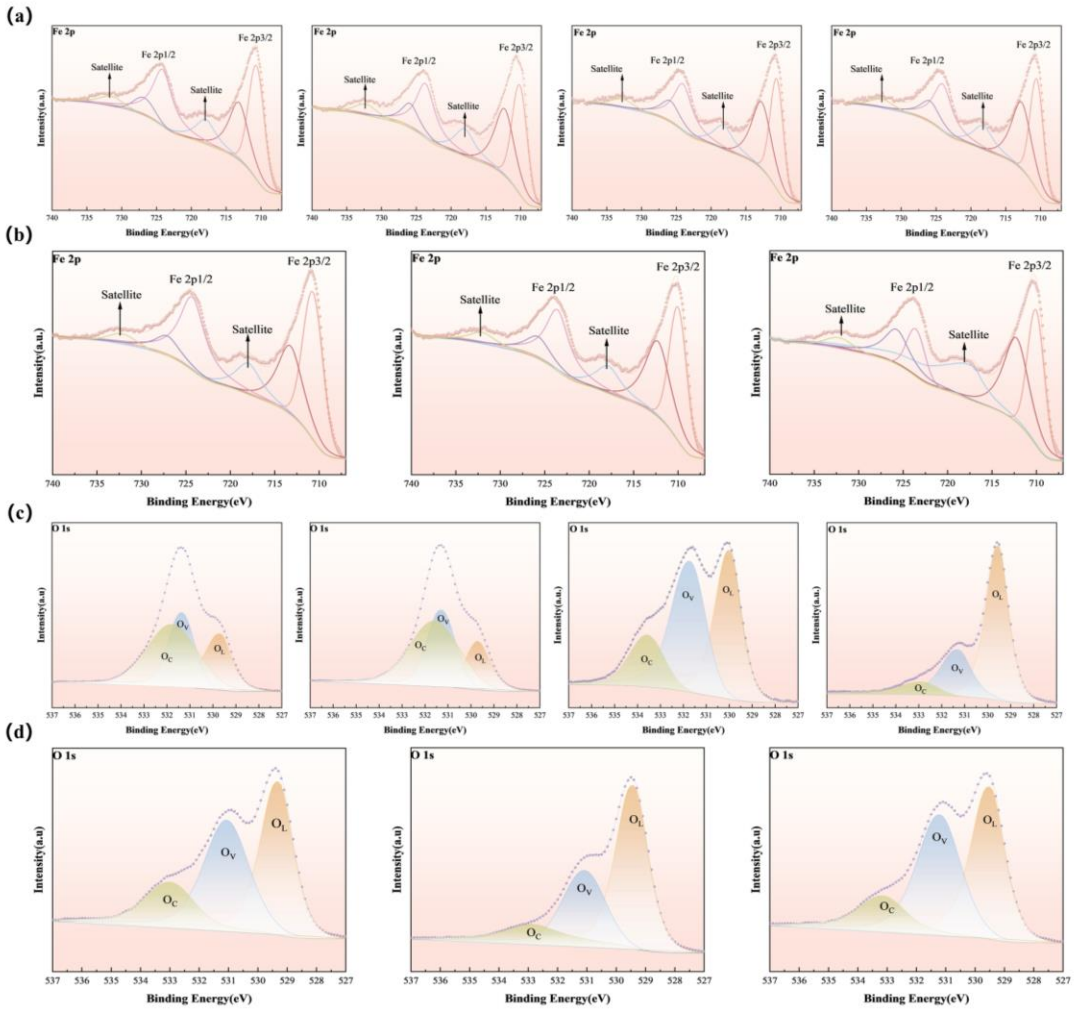
Supplementary Figure 2. The thermogravimetric analysis (TGA) curve of MIL-100(Fe).



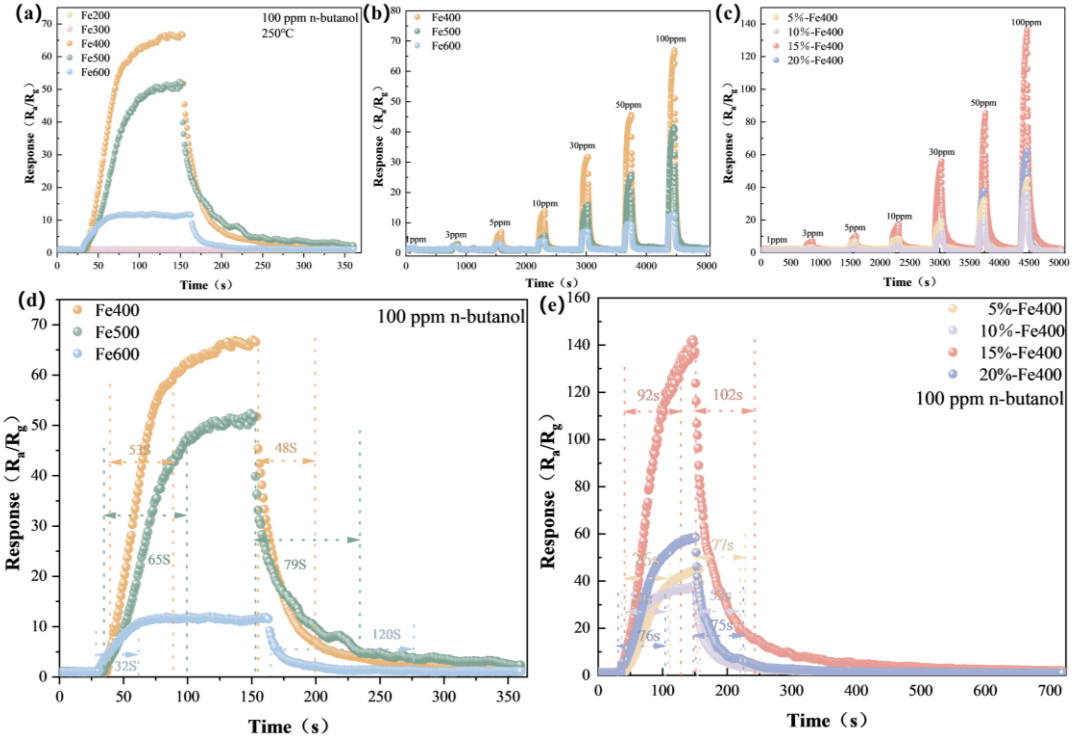
Supplementary Figure 3. (a, b) XRD patterns of Fex(x=200-600) and x%-Fe400 (x=0,5,10,15,20) samples.



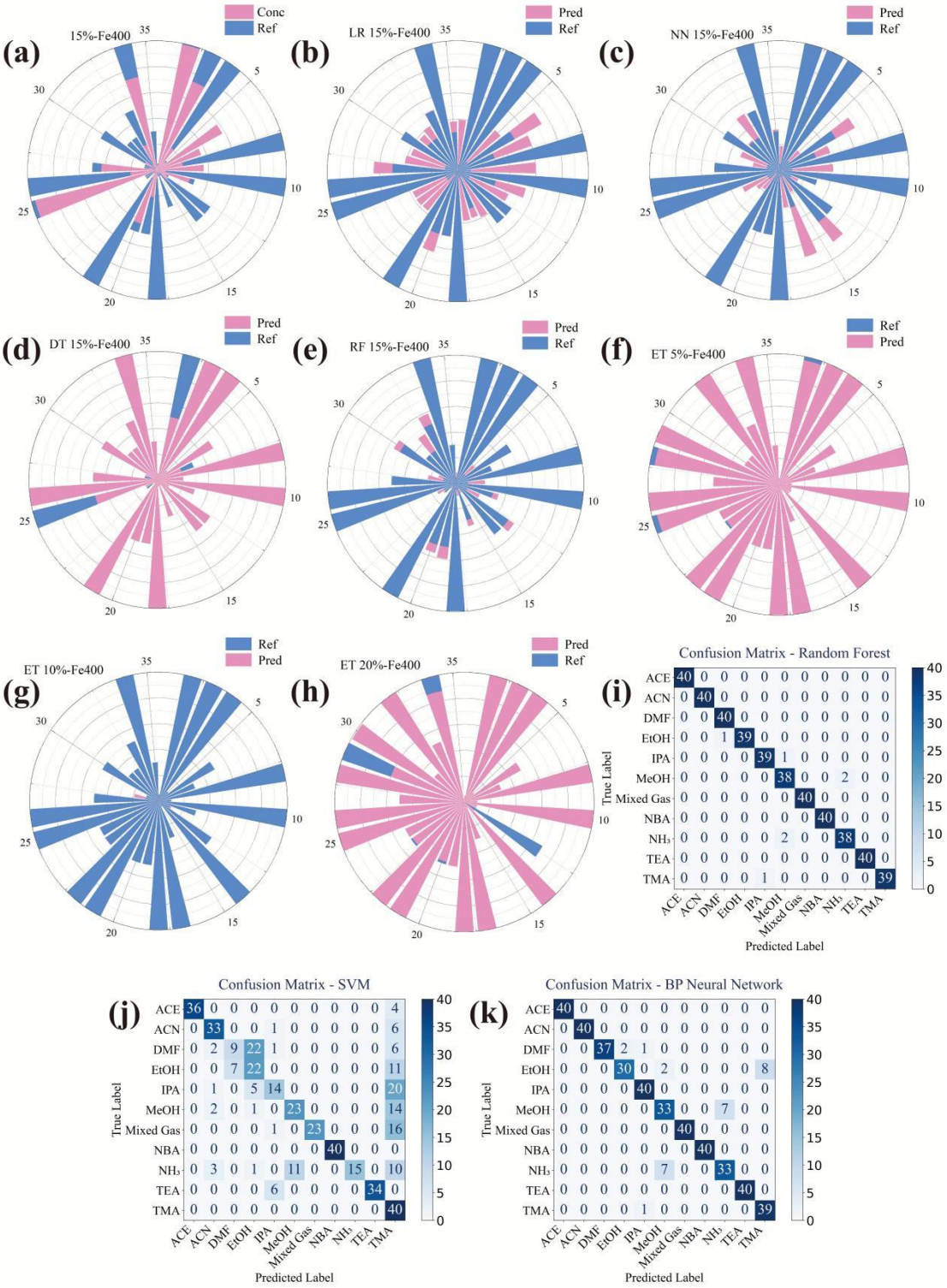
Supplementary Figure 4. (a, b) SEM images of Fex (x = 500, 600); (c-e) SEM images of x-Fe400 (x = 5%, 10%, 20%).



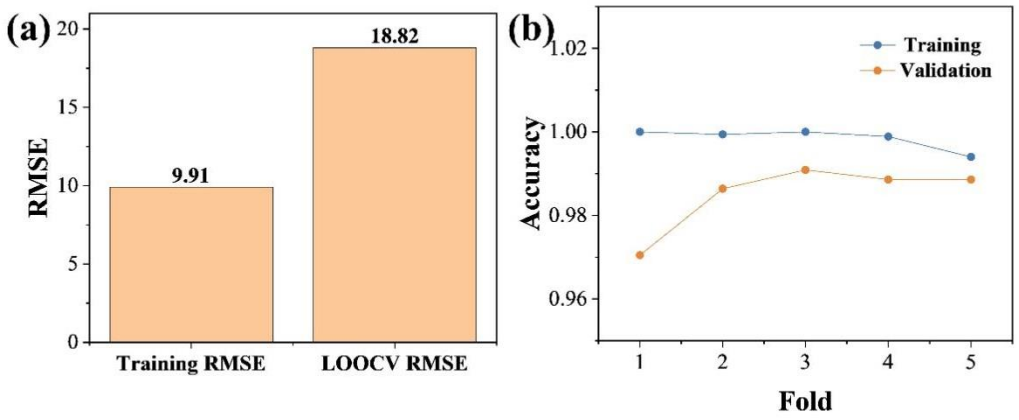
Supplementary Figure 5. Fe 2p spectra of (a)Fex(x=200, 300, 500, 600) and (b)x-Fe400(x=5%,10%,20%); O 1s spectra of (c)Fex(x=200, 300, 500, 600) and (d)x-Fe400(x=5%,10%,20%).



Supplementary Figure 6. (a) The real-time response curves of the sensors subjected to heat treatment at different temperatures to 100 ppm n-butanol;(b) Dynamic responses to 1–100 ppm n-butanol at 250 ℃ for sensors calcined at various temperatures;(c) Dynamic responses to 1–100 ppm n-butanol at the optimal temperature for sensors with different H_2BDC contents; (d) Response/recovery time curves of Fex (x=400, 500, 600) sensors to 100 ppm n-butanol; (e) Response/recovery time curves of x-Fe400 (x=5%, 10%, 15%, 20%) sensors to 100 ppm n-butanol. All subfigures are original and were plotted using Origin 2024.



Supplementary Figure 7.(a) Measured vs. actual NBA concentration under different conditions; Corrected concentration using: (b) Linear Regression (LR), (c) Neural Network (NN), (d) Decision Tree (DT), (e) Random Forest (RF);(f-h) Humidity correction plots using the ET algorithm for x-Fe400 (x = 5%, 10%, 20%); (i-k) Confusion matrix heatmaps for gas classification using four sensors, corresponding to the Random Forest, SVM, and BP algorithms, respectively.



Supplementary Figure 8. (a) Training Vs LOOCV RMSE - ET; (b) 5-fold cross-validation accuracy - CatBoost.

Supplementary Table 1. Crystal plane peak intensity

Sample	Fe400	5%-Fe400	10%-Fe400	15%-Fe400	20%-Fe400
Intensity（a.u）	552(311)	1120(311)	1062(311)	1172(311)	1014(311)

Supplementary Table 2. Calculate the average crystallite size

Sample	5%-Fe400	10%-Fe400	15%-Fe400	20%-Fe400
Average size/nm	55.6	58.7	60.8	64.5

$$\text{Size}=\frac{K\times\lambda}{B\times\text{Cos}\theta} \quad (1)$$

K is the Scherrer constant, B is the full width at half maximum (FWHM) of the diffraction peak from the measured sample, θ is the Bragg angle, and λ is the X-ray wavelength[][1].

Supplementary Table 3. 15%-Fe400 Humidity-corrected data

Day	RH	T	Res	Conc	ref
1	25	22	10.46174	3.497597043	5

2	25	22	142.30725	99.06898626	100
3	35	22	107.35	73.72938277	100
4	55	23	33.69575	20.33932804	100
5	25	22	17.47356	8.580283426	10
6	25	22	85.53275	57.91462433	50
7	25	22	55.56336	36.19059113	30
8	55	23	33.69575	20.33932804	100
9	25	22	55.56336	36.19059113	10
10	65	20	14.77887	6.626972564	100
11	35	24	41.91441	26.29682143	30
12	25	23	17.47356	8.580283426	10
13	55	24	20.25259	10.59473017	50
14	65	24	8.883	2.35320938	50
15	25	22	10.46174	3.497597043	5
16	65	24	5.8703	0.16938132	30
17	35	25	7.893	1.635584067	5

18	65	2 1	14.79884	6.641448298	100
19	45	2 3	34.50765	20.92785329	50
20	35	2 3	64.986	43.02081838	50
21	65	2 4	14.669	6.547330651	100
22	25	2 2	6.82328	0.860171795	3
23	35	2 4	7.8805	1.626523142	10
24	45	2 3	7.107	1.065833062	10
25	25	2 4	139.461	97.00581349	100
26	55	2 3	34.019	20.57364358	100
27	35	2 4	64.891	42.95195535	50
28	35	2 4	14	5.972505527	10
29	45	2 4	7.0184	1.00160922	10
30	55	2 4	20.245	10.58922837	50
31	45	2 3	7.112	1.069457432	30
32	55	2 3	13.1651	5.457192563	30
33	45	2 2	34.906	21.21660686	50

34	35	23	106.761	73.30243195	100
35	45	24	7.908	1.646457178	30

Supplementary Table 4. 5%-Fe400 Humidity-corrected data

Day	RH	T	Res	Conc	ref
1	25	22	6.42461	5.04207164	5
2	25	22	44.53121	91.98389231	100
3	35	22	22.710917	42.19997057	100
4	55	23	4.453123	0.544040611	100
5	25	22	8.42841	9.613826146	10
6	25	22	32.14813	63.73137121	50
7	25	22	19.37951	34.59922428	30
8	55	23	4.319527	0.239235683	100
9	25	22	8.42841	9.613826146	10
10	65	20	4.75	1.221378052	100
11	35	24	4.32865	0.260050194	10
12	25	23	7.654	7.846976956	10

13	35	24	4.281	0.151334702	10
14	65	24	4.567	0.803855807	100
15	25	22	6.42461	5.04207164	5
16	35	24	8.59632	9.996919918	30
17	35	25	23.9878	45.11323295	100
18	65	21	4.63	0.947592973	100
19	45	23	11.924	17.58916267	50
20	35	23	14.6398	23.78537531	50
21	65	24	4.643	0.977253023	100
22	25	22	44.98743	93.02477755	100
23	35	24	14.5236	23.5202601	50
24	45	23	11.9478	17.64346338	50
25	25	24	44.3532	91.57775496	100
26	55	23	4.43531	0.503399498	100
27	35	24	16.935	29.02197125	50
28	35	24	23.358	43.67631759	100

29	45	24	16.5334	28.10570386	100
30	35	24	9.8846	12.93618526	30
31	45	23	6.066172	4.224280173	30
32	55	23	4.3171	0.23369838	100
33	45	22	11.9359	17.61631303	50
34	35	23	23.8915	44.89352042	100
35	45	24	6.65236	5.561692904	30

Supplementary Table 5. 10%-Fe400 Humidity-corrected data

Day	RH	T	Res	Conc	ref
1	25	22	2.58153	5.045210323	5
2	25	22	36.18821	98.60240528	100
3	35	22	22.710917	61.08314663	100
4	55	23	3.6185	7.932017483	100
5	25	22	3.76259	8.333147741	10
6	25	22	19.89716	53.2499652	50
7	25	22	11.47141	29.79363047	30

8	55	23	3.608	7.902786671	100
9	25	22	3.76259	8.333147741	10
10	65	20	3.86	8.60432616	100
11	35	24	7.1992	17.90028117	30
12	25	23	3.6723	8.081790596	10
13	55	24	1.9896	3.397344172	50
14	65	24	4.242	9.667770942	100
15	25	22	2.58153	5.045210323	5
16	45	24	4.25438	9.702235461	30
17	35	25	23.9878	64.63784416	100
18	65	21	3.884	8.671139445	100
19	45	23	7.379	18.40082403	50
20	35	23	13.198	34.60026169	50
21	65	24	3.895	8.7017622	100
22	25	22	36.633	99.84065032	100
23	35	24	13.2113	34.63728738	50

24	45	23	7.4516	18.60293422	50
25	25	24	36.549	99.60680382	100
26	55	23	3.611	7.911138331	100
27	35	24	13.012	34.08245873	50
28	35	24	2.4983	4.813507419	10
29	45	24	1.42206	1.817377022	10
30	55	24	1.98954	3.397177139	50
31	45	23	4.29352	9.811196793	30
32	55	23	1.147	1.051641101	30
33	45	22	7.4505	18.59987194	50
34	35	23	22.7166	61.09896718	100
35	45	24	4.288	9.795829737	30

Supplementary Table 6. 20%-Fe400 Humidity-corrected data

Day	RH	T	Res	Conc	ref
1	25	22	4.75658	4.964502881	5
2	25	22	62.2925	96.55012575	100

3	35	22	31.769175	47.96312597	100
4	55	23	6.23	7.309891439	100
5	25	22	7.08107	8.664623858	10
6	25	22	38.17832	58.16519691	50
7	25	22	19.50246	28.43698068	30
8	55	23	6.1578	7.194963548	100
9	25	22	7.336	9.07042119	10
10	65	20	6.654	7.984814237	100
11	35	24	3.5764	3.085893477	10
12	25	23	7.134	8.748877782	10
13	35	24	3.357	2.736652765	10
14	65	24	6.2387	7.323740091	100
15	25	22	4.75658	4.964502881	5
16	35	24	9.96425	13.25406705	30
17	35	25	31.9885	48.3122473	100
18	65	21	6.87456	8.335901436	100

19	45	2 3	15.5367	22.12428767	50
20	35	2 3	19.6054	28.60084047	50
21	65	2 4	7.39	9.156378339	100
22	25	2 2	62.7818	97.32899303	100
23	35	2 4	19.5426	28.50087549	50
24	45	2 3	15.93451	22.75752125	50
25	25	2 4	62.043	96.15297189	100
26	55	2 3	6.5206	7.772468244	100
27	35	2 4	20.0843	29.36315304	50
28	35	2 4	9.946	13.22501671	100
29	45	2 4	25.34996	37.74502563	100
30	35	2 4	9.8922	13.13937792	30
31	45	2 3	7.9365	10.02629652	30
32	55	2 3	6.4831	7.712775779	100
33	45	2 2	15.7718	22.49851963	50
34	35	2 3	9.893	13.14065136	100

35	25	2/4	3.18697	2.465999172	3

Model performance was evaluated using statistical metrics derived from a comparison of the algorithm-corrected concentrations and the actual reference values. These parameters include the Mean Absolute Error (MAE), calculated as shown in Eq. (2)[2]:

$$MAE=\frac{\Sigma_{i=1}^{n}|y_i-\hat{y}_i|}{n} \quad (2)$$

where n is the total number of data points, $\hat{y}_i$ is the actual concentration of the i-th data point, and $\hat{y}_i$ is the corrected concentration of the i-th data point.

The Root Mean Square Error (RMSE), calculated as shown in Eq. (3), using the same data variables as defined for MAE[2].

$$RMSE=\sqrt{\frac{\Sigma_{i=1}^{n}(y_i-\hat{y}_i)^2}{n}} \quad (3)$$

The Pearson correlation coefficient (r), calculated as shown in Eq. (4), where $\bar{y}$ is the mean of the actual concentrations, $\hat{\bar{y}}$ is the mean of the corrected concentrations, and the other data variables are the same as those defined for MAE[2].

$$r=\frac{\Sigma_{i=1}^{n}(y_i-\bar{y})(\hat{y}_i-\hat{\bar{y}})}{\sqrt{\Sigma_{i=1}^{n}(y_i-\bar{y})^2\,\Sigma_{i=1}^{n}(\hat{y}_i-\hat{\bar{y}})^2}} \quad (4)$$

The Spearman's rank correlation coefficient (ρ), calculated as shown in Eq. (5), where D_i represents the difference between the ranks of each data point after sorting, and n is the total number of data points[2].

$$\rho=1-\frac{6\Sigma_{i=1}^{n}D_i^2}{n(n^2-1)} \quad (5)$$

For detailed analysis, the cumulative distribution function (CDF) of the absolute errors was plotted. The y-axis represents the normalized mean absolute error (NMAE), calculated using Eq. (6)[3].

Target diagrams were used to visualize calibration performance. The x-axis represents the normalized root mean square error (NRMSE), calculated using Eq. (7)[3]. The same data variables as defined for the Pearson correlation coefficient (R) were used in this

calculation.

$$\mathrm{NMAE}=\frac{\mathrm{MAE}}{\frac{1}{n}\sum_{i}^{n} y_i} \tag{6}$$

$$\mathrm{NRMSE}=\frac{\mathrm{RMSE}}{\max(y_i)-\min(y_i)} \tag{7}$$

REFERENCES

1. Zahra S, Naz U, Irshad M, et al. Effect of calcination temperature on structural and magnetic properties of polypropylene glycol stabilized nickel ferrite nanoparticles. *BMC Chemistry* 2025; 19: 106.[DOI:10.1186/s13065-025-01454-w]
2. Zhang Z, Chen S, Yang R, et al. Modeling enzyme temperature stability from sequence segment perspective. *Journal of chemical information and modeling* 2025;65:20[DOI:10.1021/acs.jcim.5c01674]
3. Prasianakis NI, Laloy E, Jacques D, et al. Geochemistry and machine learning: methods and benchmarking. *Environmental Earth Sciences* 2025; 84: 121.[DOI:10.1007/s12665-024-12066-3]