Supplementary Information

Modeling the reactivity of ozone and sulphate radicals towards organic chemicals in water using machine learning approaches

Shikha Gupta[1a], Nikita Basant[*2]

[1]CSIR-Indian Institute of Toxicology Research,

Post Box 80, Mahatma Gandhi Marg, Lucknow-226 001, India

[2]ETRC, Gomtinagar, Lucknow-226010, India

([a]Current address: CSIR-NBRI, Lucknow)

***Author for all correspondance :** nikita_basant@yahoo.com

Table S1a: List of references for ozone data

S.No.	Ozone data Reference
1	J.L. Acero, F. Javier Benitez, F.J. Real, and C. Maya, *Ind. Eng. Chem. Res.*, 2003, **42**, 5762–5769.
2	F.J. Benitez, J. Beltran-Heredia, and T. Gonzalez, *Ozone Sci. Eng.,* 1994, **16**, 213–234.
3	J. Benner, E. Salhi, T. Ternes, and U. von Gunten, *Water Res.* 2008, **42**, 3003–3012.
4	V. Butkovic, L. Klasinc, M. Orhanovic, J. Turk, and H. Gusten, *Environ. Sci. Technol.,* 1983, **17**, 546–548.
5	P. Cheme-Ayala, M.G. El-Din, and D.W. Smith, *Chemosphere,* 2010, **78**, 557–562.
6	M. Deborde, S. Rabouan, J.P. Duguet, and B. Legube, *Environ. Sci. Technol.,* 2005, **39**, 6086–6092.
7	M.C. Dodd, M.O. Buffle, and U. von Gunten, *Environ. Sci. Technol.,* 2006, **40**, 1969–1977.
8	M.M. Huber, S. Canonica, G.Y. Park, and U. von Gunten, *Environ. Sci. Technol.,* 2003, **37**, 1016–1024.
9	X. Jin, S. Peldszus, and P.M. Huck, *Water Res.,* 2012, 46, 6519–6530.
10	C.C.D. Yao, and W.R. Haag, *Water Res.,* 1991, **25**,761–773.
11	J. Hoigne, and H. Bader, *Water Res.,* 1983, **17**, 185-194.
12	J. Hoigne, and H. Bader, *Water Res.,* 1983, **17**, 173-183.
13	J. Hoigne, H. Bader, W. Haag, and J. Staehelin, *Water Res.,* 1985, **19**, 993-1004.
14	A. Peter, and U. Von Gunten, *Environ. Sci. Technol.,* 2007, **41**, 626-631.

Table S1b: List of references for sulphate radical data

S.No.	Sulphate data reference
1	X. He, S.P. Mezyk, I. Michael, D, Fatta-Kassinos, and D.D. Dionysiou, *J. Hazard. Mater.*, 2014, **279**, 375−383.
2	I. Grgic, B. Podkrajsek, P. Barzaghi, and H. Herrmann, *Atmos. Environ*., 2007, **41**, 9187−9194.
3	S. Padmaja, Z.B. Alfassi, P. Neta, and R.E. Huie, *Int. J. Chem. Kinet.,* 1993, **25**, 193−198.
4	A.B. Ross, and P. Neta, Rate constants for reactions of inorganic radicals in aqueous solution. In Commerce, Ed; National Bureau of Standards: Washington, DC, U.S., 1979.
5	M. Bonifacic, and K.D. Asmus, *J. Phys. Chem*., 1976, **80**, 2426−2430.
6	C.L. Clifton, and R.E. Huie, *Int. J. Chem. Kinet*., 1989, **21**, 677−687.
7	L. Dogliotti, and E. Hayon, *J. Phys. Chem.* 1967, **71**, 2511−2516.
8	C. George, H. El Rassy, and J.M. Chovelon, *Int. J. Chem. Kinet.,* 2001, **33**, 539−547.
9	H. Herrmann, H.W. Jacobi, G. Raabe, A. Reese, and R. Zellner, *Anal. Bioanal. Chem.,* 1996, **355**, 343−344.
10	R.E. Huie, C.L. Clifton, and S.A. Kafafi, *J. Phys. Chem.,* 1991, **95**, 9336−9340.
11	I. Kraljić, *Int. J. Radiat. Phys. Chem.*, 1970, **2**, 59−68.
12	P. Maruthamuthu, and P. Neta, *J. Phys. Chem.,* 1977, **81**, 1622−1625.
13	P.G. Tratnyek, T.J. Grundl, and S.B. Haderlein, Aquatic Redox Chemistry; American Chemical Society: Washington, DC, 2011.
14	P. Neta, R.E. Huie, and A.B. Ross, *J. Phys. Chem. Ref. Data*, 1990, 19, 413−513.
15	P. O'Neill, S. Steenken, and D. Schulte-Frohlinde, *J. Phys. Chem*., 1975, **79**, 2773−2779.
16	J. Paul, D.B. Naik, Y.K. Bhardwaj, and L. Varshney, *Radiat. Phys. Chem.*, 2014, **100**, 38−44.
17	J.L. Redpath, and R.L. Willson, *Int. J. Radiat. Biol*., 1975, **27**, 389−398.
18	S. Steenken, P. O'Neill, and D. Schulte-Frohlinde, *J. Phys. Chem*., 1977, **81**, 26−30.
19	M.M. Ahmed, S. Barbati, P. Doumenq, and S. Chiron, *Chem. Eng. J.* 2012, **197**, 440−447.

Table S2: Data set used in QSRR modeling

S.No.	Compound	log*k* (experimental)	Dependent variable
1	Equilenin	7.000	log k_{O3}
2	Butylated hydroxyanisole	6.520	log k_{O3}
3	Fenoterol	6.450	log k_{O3}
4	Tetracycline	6.400	log k_{O3}
5	Triclosan	6.400	log k_{O3}
6	Phenol	6.260	log k_{O3}
7	17α-Ethinylestradiol	6.260	log k_{O3}
8	Gemfibrozil	5.690	log k_{O3}
9	Methicillin	4.590	log k_{O3}
10	Benzo[a]pyrene	3.880	log k_{O3}
11	Clofibric acid	3.700	log k_{O3}
12	Trifluralin	3.280	log k_{O3}
13	Methoxychlor	2.400	log k_{O3}
14	Tri(2-butoxyethyl)phosphate	0.660	log k_{O3}
15	Butylbenzyl phthalate	0.150	log k_{O3}
16	Metformin	0.080	log k_{O3}
17	Tris(chloroethyl)phosphate	-0.100	log k_{O3}
18	Iomeprol	-0.700	log k_{O3}
19	Dicamba	-1.000	log k_{O3}
20	Dicofol	-1.000	log k_{O3}
21	Di(2-ethylhexyl)phthalate	-1.000	log k_{O3}
22	Hexachlorobenzene	-2.000	log k_{O3}
23	Bisphenol A	6.430	log k_{O3}
24	Sulfamethoxazole	6.300	log k_{O3}
25	Estrone	6.260	log k_{O3}
26	17 beta-Estradiol	6.230	log k_{O3}
27	Estriol	6.220	log k_{O3}
28	Amoxicillin	6.180	log k_{O3}
29	4-Nonylphenol	6.150	log k_{O3}
30	Lincomycin	5.830	log k_{O3}
31	Carbamazepine	5.480	log k_{O3}
32	Trimethoprim	5.430	log k_{O3}
33	Naproxen	5.410	log k_{O3}
34	Enrofloxacin	5.180	log k_{O3}
35	Pyrene	4.560	log k_{O3}
36	Ciprofloxacin	4.280	log k_{O3}
37	Phenantlene	4.200	log k_{O3}
38	Penicillin G	3.680	log k_{O3}
39	Metoprolol	3.226	log k_{O3}

40	Bromoxynil	3.060	log k_{O3}
41	Bezafibrate	2.770	log k_{O3}
42	Ibuprofen	0.980	log k_{O3}
43	Alachlor	0.580	log k_{O3}
44	Diazepam	-0.120	log k_{O3}
45	Dimethyl phthalate	-0.700	log k_{O3}
46	Diethylphthalate	-0.850	log k_{O3}
47	Gamma-HCH	-1.400	log k_{O3}
48	2,4-D	0.360	log k_{O3}
49	Atrazine	0.780	log k_{O3}
50	Simazine	0.680	log k_{O3}
51	Acetochlor	0.380	log k_{O3}
52	Cyanazine	0.870	log k_{O3}
53	Iopromide	-0.100	log k_{O3}
54	Metolachlor	0.040	log k_{O3}
55	Propachlor	-0.050	log k_{O3}
56	Dalapon	-3.301	log k_{O3}
57	Propionic acid	-2.699	log k_{O3}
58	Adipic acid monoethyl ester	-2.523	log k_{O3}
59	Succinic acid	-2.523	log k_{O3}
60	Tert-butanol	-2.523	log k_{O3}
61	Methanol	-1.620	log k_{O3}
62	Cyanic acid	-2.000	log k_{O3}
63	Glyoxylic acid	0.279	log k_{O3}
64	Propan-2-ol	0.279	log k_{O3}
65	Endothall	0.301	log k_{O3}
66	Pyridine	0.477	log k_{O3}
67	Octanal	0.903	log k_{O3}
68	Fenoprop	0.949	log k_{O3}
69	Nitrosodimethylamine	1.000	log k_{O3}
70	2-isopropyl-3-methoxypyrazine	1.701	log k_{O3}
71	Picloram	2.000	log k_{O3}
72	Methylamine	2.447	log k_{O3}
73	Beta-alanine	2.491	log k_{O3}
74	1-phenoxy-2-propanol	2.505	log k_{O3}
75	Oxamyl	2.792	log k_{O3}
76	Alpha-alanine	2.806	log k_{O3}
77	Atenolol	3.230	log k_{O3}
78	Beta-cyclocitral	3.590	log k_{O3}
79	Dimethylamine	3.903	log k_{O3}
80	Trimethylamine	5.097	log k_{O3}
81	Propanolol	5.000	log k_{O3}
82	1,1-dimethylguanidine	8.663	log k_{SO4}

83	1,3,5-trimethoxy-benzene	9.255	log k_{SO4}
84	1,3-dimethoxy-benzene	9.845	log k_{SO4}
85	1,4-dimethoxy-benzene	9.857	log k_{SO4}
86	1,4-dioxane	7.612	log k_{SO4}
87	1-heptanol	8.398	log k_{SO4}
88	1-hexanol	8.204	log k_{SO4}
89	1-octanol	8.505	log k_{SO4}
90	1-pentanol	8.114	log k_{SO4}
91	1-propanol	7.728	log k_{SO4}
92	2,3,4-trimethoxy-benzoate	9.398	log k_{SO4}
93	2,3-dimethyl-2,3-butanediol	8.415	log k_{SO4}
94	2,4,6-trimethoxy-benzoate	9.415	log k_{SO4}
95	2-methoxybenzoate	9.845	log k_{SO4}
96	2-propanol	7.922	log k_{SO4}
97	3,5-dimethoxy-benzoate	9.643	log k_{SO4}
98	3-methoxybenzoate	9.881	log k_{SO4}
99	acetanilide	9.556	log k_{SO4}
100	acetate	6.672	log k_{SO4}
101	amoxicillin	9.462	log k_{SO4}
102	ampicillin	9.301	log k_{SO4}
103	anisole	9.690	log k_{SO4}
104	benzene	8.826	log k_{SO4}
105	benzoic acid	9.079	log k_{SO4}
106	carboxyformate	6.230	log k_{SO4}
107	cephalothin	9.542	log k_{SO4}
108	cyclopentanol	8.447	log k_{SO4}
109	diclofenac	9.964	log k_{SO4}
110	dimethyl disulfide	8.580	log k_{SO4}
111	dipropyl ether	8.342	log k_{SO4}
112	estradiol	9.083	log k_{SO4}
113	ethanol	7.720	log k_{SO4}
114	ethinylestradiol	9.479	log k_{SO4}
115	formic acid	5.699	log k_{SO4}
116	fumarate	7.204	log k_{SO4}
117	glycine	6.954	log k_{SO4}
118	glycolate	7.204	log k_{SO4}
119	ibuprofen	9.580	log k_{SO4}
120	isoborneol	8.723	log k_{SO4}
121	lactate	7.204	log k_{SO4}
122	lactic acid	7.000	log k_{SO4}
123	malate	7.176	log k_{SO4}
124	malic acid	6.886	log k_{SO4}
125	malonic acid	6.431	log k_{SO4}

126	mesitylene	9.114	log k_{SO4}
127	methanol	7.021	log k_{SO4}
128	methionine	9.041	log k_{SO4}
129	m-toluic acid	9.301	log k_{SO4}
130	o-bromobenzoic acid	8.940	log k_{SO4}
131	oxetane	8.053	log k_{SO4}
132	p-bromobenzoic acid	9.000	log k_{SO4}
133	p-chlorobenzoic acid	8.556	log k_{SO4}
134	p-cresol	9.447	log k_{SO4}
135	penicillinV	9.378	log k_{SO4}
136	p-hydroxybenzoic acid	9.398	log k_{SO4}
137	piperacillin	9.079	log k_{SO4}
138	p-methoxybenzoic acid	9.544	log k_{SO4}
139	progesterone	9.076	log k_{SO4}
140	p-toluic acid	9.255	log k_{SO4}
141	serine	7.362	log k_{SO4}
142	succinic acid	6.851	log k_{SO4}
143	terephthalic acid	8.230	log k_{SO4}
144	tetrahydrofuran	8.369	log k_{SO4}
145	tetrahydropyran	8.057	log k_{SO4}
146	tyrosine	9.505	log k_{SO4}
147	1,2,4-trimethoxy-benzene	9.892	log k_{SO4}
148	1,2-dimethoxy-benzene	9.699	log k_{SO4}
149	1,3-dioxane	7.535	log k_{SO4}
150	1-butanol	7.903	log k_{SO4}
151	2,6-dimethoxy-benzoate	9.398	log k_{SO4}
152	3-pentanol	8.204	log k_{SO4}
153	diethyl ether	8.114	log k_{SO4}
154	formaldehyde	7.041	log k_{SO4}
155	glycolic acid	6.623	log k_{SO4}
156	histidine	9.398	log k_{SO4}
157	hydrogen malate	6.929	log k_{SO4}
158	o-toluic acid	9.146	log k_{SO4}
159	oxalic acid	6.041	log k_{SO4}
160	oxepane	8.342	log k_{SO4}
161	penicillinG	9.146	log k_{SO4}
162	p-nitrosodimethylaniline	9.362	log k_{SO4}
163	p-xylene	9.431	log k_{SO4}
164	ticarcillin	8.903	log k_{SO4}
165	toluene	9.114	log k_{SO4}
166	tryptophan	9.362	log k_{SO4}

Table S3a: Correlation matrix of selected descriptors in $\log k_{O3}$ data set

Descriptors	AMR	nX	minHBa	MDEC-24
AMR	1.00			
nX	0.20	1.00		
minHBa	-0.29	-0.37	1.00	
MDEC-24	0.30	0.07	0.31	1.00

Table S3b: Correlation matrix of selected descriptors in $\log k_{SO4}$ data set

Descriptors	AMR	SssO	meanI
AMR	1.00		
SssO	0.10	1.00	
meanI	-0.37	-0.27	1.00

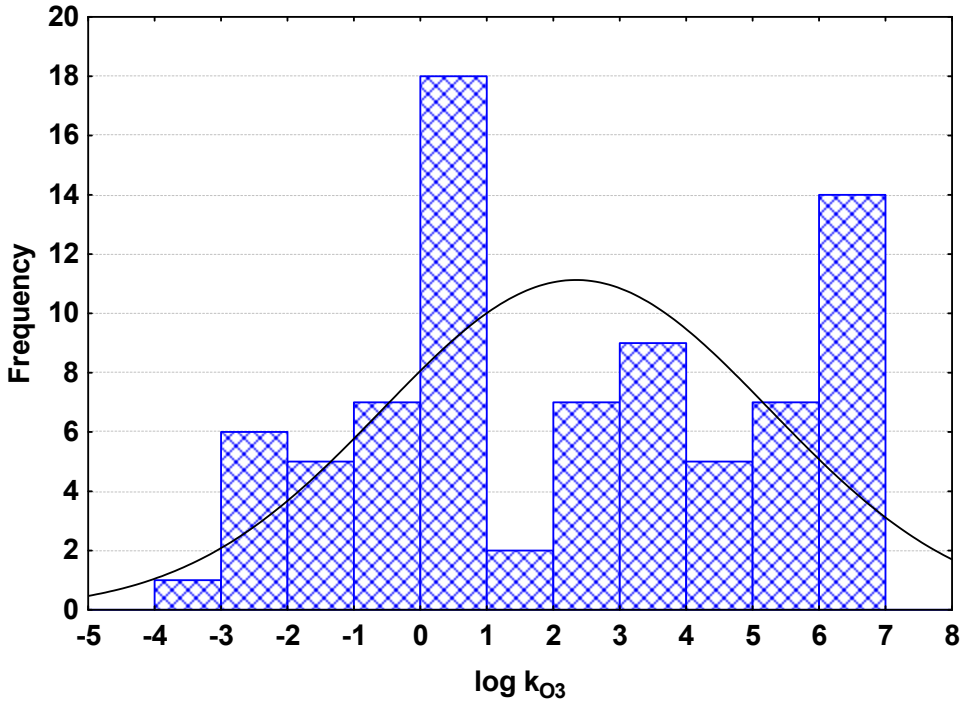

Figure S1a: Histogram of $\log k_{O3}$ values

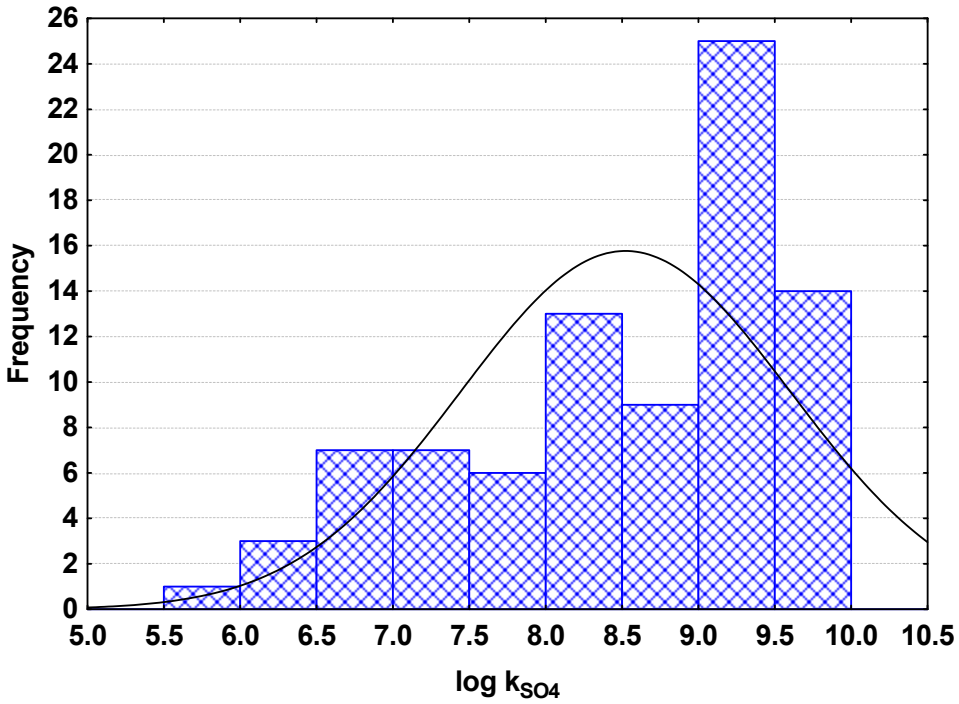

Figure S1b: Histogram of logk_{SO4} values

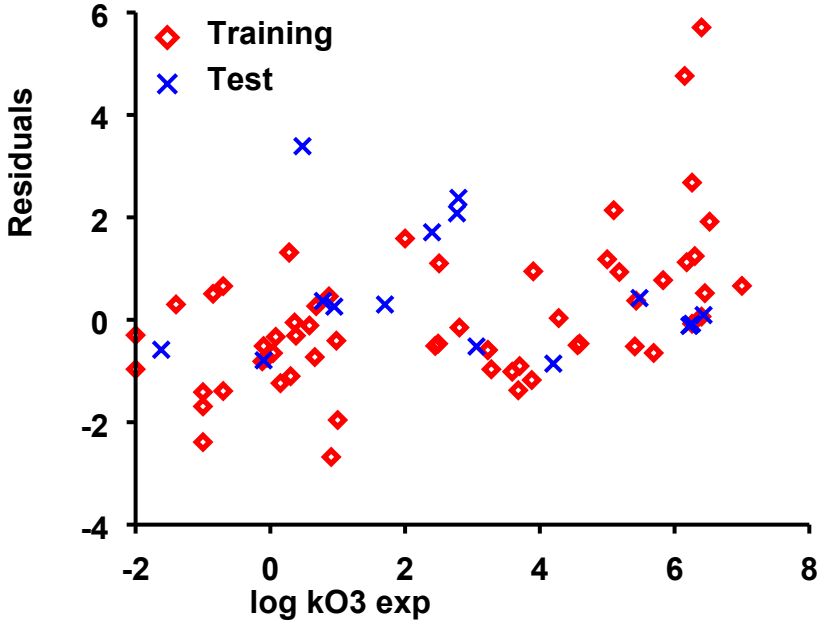

Figure S2a: Plot showing the distribution of experimental and residual values of logk_{O3} in training and test sets by SDT-QSRR model

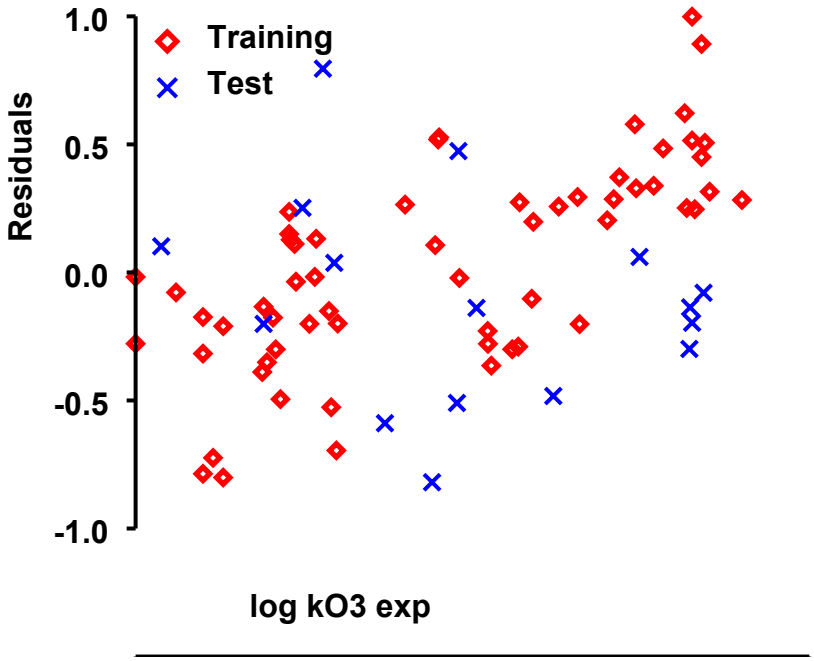

Figure S2b: Plot showing the distribution of experimental and residual values of $\log k_{O3}$ in training and test sets by DTB-QSRR model

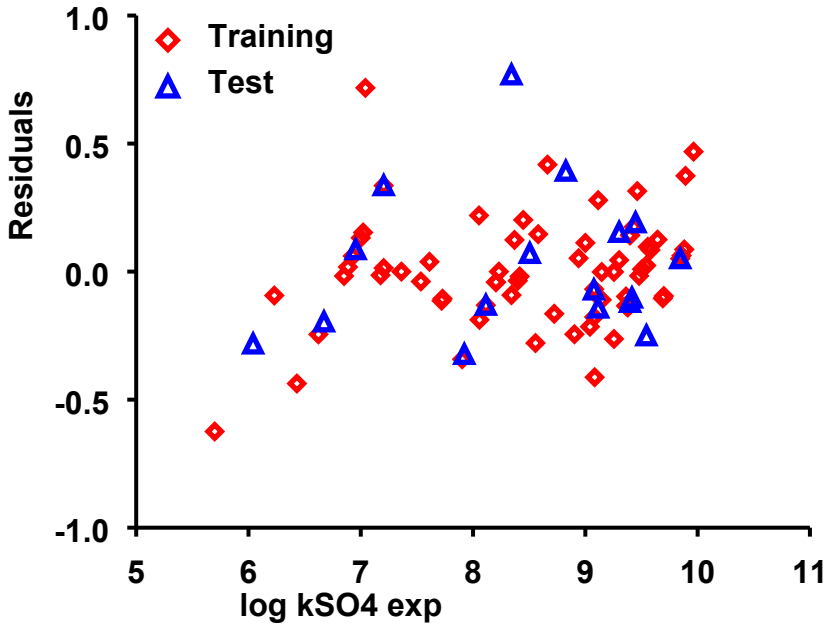

Figure S3a: Plot showing the distribution of experimental and residual values of $\log k_{SO4}$ in training and test sets by SDT-QSRR model

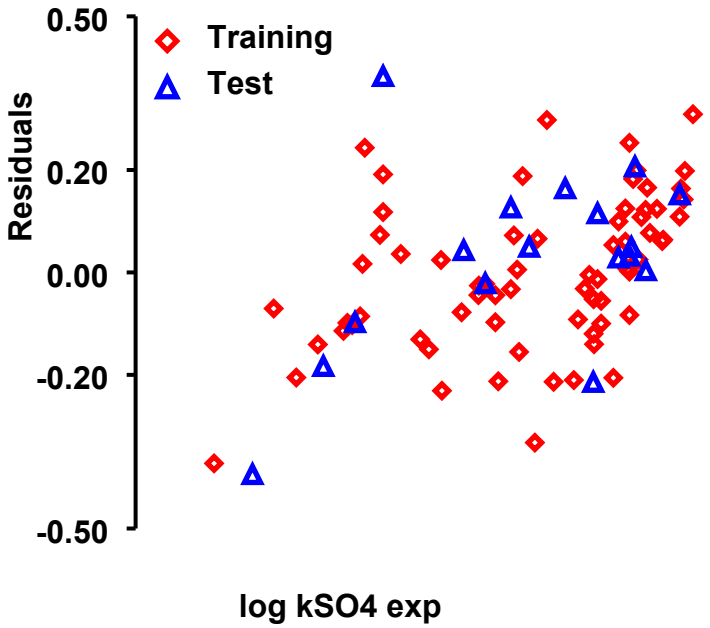

Figure S3b: Plot showing the distribution of experimental and residual values of $\log k_{SO4}$ in training and test sets by DTB-QSRR model