Electronic Supplementary Material (ESI) for Environmental Science: Nano.
This journal is © The Royal Society of Chemistry 2017

How to judge whether QSAR/read-across predictions can be trusted? Novel approach for establishing model's applicability domain

Agnieszka Gajewicz

Laboratory of Environmental Chemometrics, Faculty of Chemistry, University of Gdansk, Gdansk, Poland

Supplementary materials

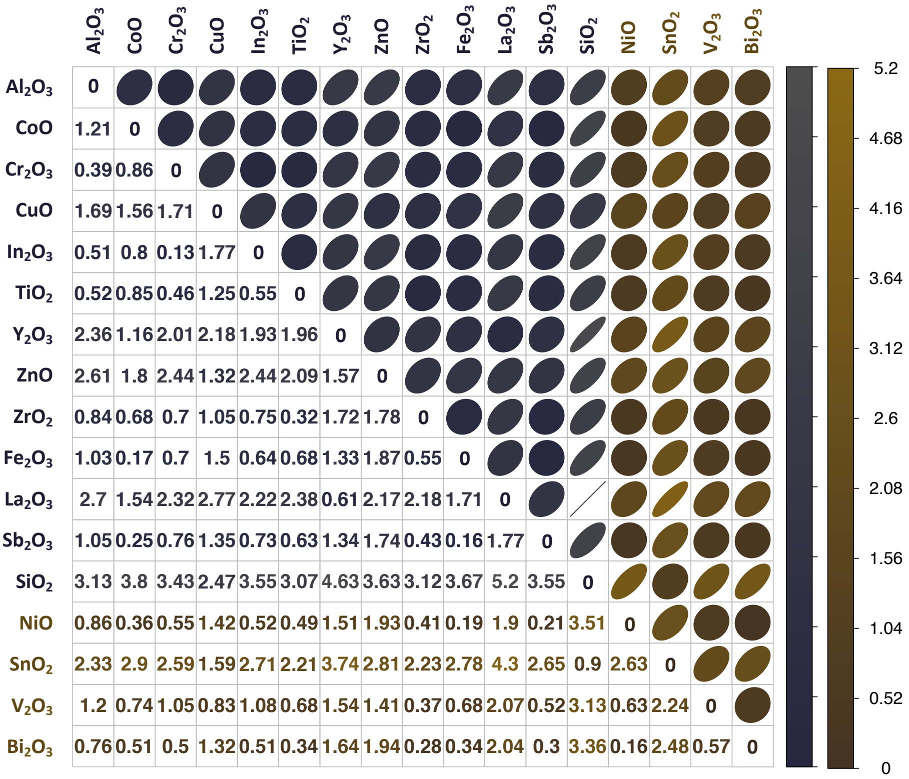
Corresponding author: A. Gajewicz, Laboratory of Environmental Chemometrics, Faculty of Chemistry, University of Gdansk, Wita Stwosza 63, 80-308 Gdansk, Poland
Phone: (+48 58) 523 52 46; fax: (+48 58) 523 50 12; e-mail: a.gajewicz@qsar.eu.org

Case Study 1

Table S1. Descriptors used for Nano-QSAR model development along with experimentally observed and predicted values of nanotoxicity to *E. coli* bacteria for training and validation compounds (MeOx NPs) reported by Pathakoti *et al.*[1]

MeOx	QMELECT	LZELEHHO	Autoscaled QMELECT	Autoscaled LZELEHHO	Set*	Experimental -log(LC_{50})	Predicted -log(LC_{50})	Residuals	Standardized cross-validated residuals	Leverages
Al_2O_3	0.10	0.21	-0.36	0.95	T	2.42	1.78	0.64	1.26	0.22
CoO	0.10	0.17	-0.49	-0.25	T	3.13	3.41	-0.28	-0.72	0.10
Cr_2O_3	0.10	0.20	-0.55	0.61	T	2.06	2.15	-0.09	-0.36	0.18
CuO	0.13	0.18	1.05	0.02	T	4.24	4.06	0.18	0.22	0.19
In_2O_3	0.10	0.20	-0.65	0.53	T	2.83	2.18	0.65	1.22	0.19
TiO_2	0.11	0.19	-0.10	0.50	T	2.14	2.6	-0.46	-1.07	0.11
Y_2O_3	0.10	0.13	-0.60	-1.40	T	5.79	4.98	0.81	1.67	0.24
ZnO	0.12	0.13	0.97	-1.30	T	5.8	5.9	-0.1	-0.50	0.49
ZrO_2	0.11	0.18	0.02	0.20	T	2.58	3.12	-0.54	-1.18	0.08
Fe_2O_3	0.10	0.17	-0.45	-0.08	T	2.4	3.19	-0.79	-1.68	0.10
La_2O_3	0.09	0.12	-1.17	-1.63	T	4.96	4.93	0.03	-0.09	0.31
Sb_2O_3	0.10	0.17	-0.29	-0.10	T	3.12	3.33	-0.21	-0.56	0.08
SiO_2	0.15	0.24	2.61	1.94	T	2.54	2.37	0.17	0.80	0.70
NiO	0.10	0.18	-0.37	0.09	V	3.79	3.01	0.78	1.19	0.10
SnO_2	0.14	0.22	1.94	1.34	V	2.53	2.77	-0.24	-0.58	0.41
V_2O_3	0.11	0.17	0.23	-0.10	V	3.48	3.68	-0.2	-0.52	0.09
Bi_2O_3	0.10	0.18	-0.26	0.20	V	3.55	2.92	0.63	0.92	0.09

** T – training set; V – validation set; QMELECT – the absolute electronegativity of the metal atom; LZELEHHO – the absolute electronegativity of the metal oxide*



	Al_2O_3	CoO	Cr_2O_3	CuO	In_2O_3	TiO_2	Y_2O_3	ZnO	ZrO_2	Fe_2O_3	La_2O_3	Sb_2O_3	SiO_2	NiO	SnO_2	V_2O_3	Bi_2O_3
Al_2O_3	0																
CoO	1.21	0															
Cr_2O_3	0.39	0.86	0														
CuO	1.69	1.56	1.71	0													
In_2O_3	0.51	0.8	0.13	1.77	0												
TiO_2	0.52	0.85	0.46	1.25	0.55	0											
Y_2O_3	2.36	1.16	2.01	2.18	1.93	1.96	0										
ZnO	2.61	1.8	2.44	1.32	2.44	2.09	1.57	0									
ZrO_2	0.84	0.68	0.7	1.05	0.75	0.32	1.72	1.78	0								
Fe_2O_3	1.03	0.17	0.7	1.5	0.64	0.68	1.33	1.87	0.55	0							
La_2O_3	2.7	1.54	2.32	2.77	2.22	2.38	0.61	2.17	2.18	1.71	0						
Sb_2O_3	1.05	0.25	0.76	1.35	0.73	0.63	1.34	1.74	0.43	0.16	1.77	0					
SiO_2	3.13	3.8	3.43	2.47	3.55	3.07	4.63	3.63	3.12	3.67	5.2	3.55	0				
NiO	0.86	0.36	0.55	1.42	0.52	0.49	1.51	1.93	0.41	0.19	1.9	0.21	3.51	0			
SnO_2	2.33	2.9	2.59	1.59	2.71	2.21	3.74	2.81	2.23	2.78	4.3	2.65	0.9	2.63	0		
V_2O_3	1.2	0.74	1.05	0.83	1.08	0.68	1.54	1.41	0.37	0.68	2.07	0.52	3.13	0.63	2.24	0	
Bi_2O_3	0.76	0.51	0.5	1.32	0.51	0.34	1.64	1.94	0.28	0.34	2.04	0.3	3.36	0.16	2.48	0.57	0

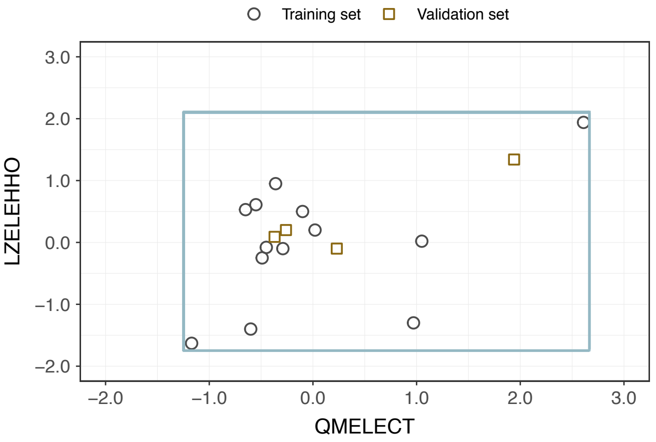
Figure S1. Euclidean distance matrix between the training (indicated in gray color) and validation (indicated in goldenrod color) compounds. The color intensity and shape of the ellipses displayed in the upper triangular part of a distance matrix reflect the values of the corresponding Euclidean distances expressed in numerical forms in the lower triangular part of this matrix. The smallest Euclidean distances (indicated as big circles) are considered as significant. In contrast, the highest distances that reflect the dissimilarity between compounds are considered as insignificant and shown as a thin ellipse.

Table S2. Input matrix data for 95% and 99% confidence interval boundaries estimation using the Student's *t*-distribution and developed in-house *R*-code

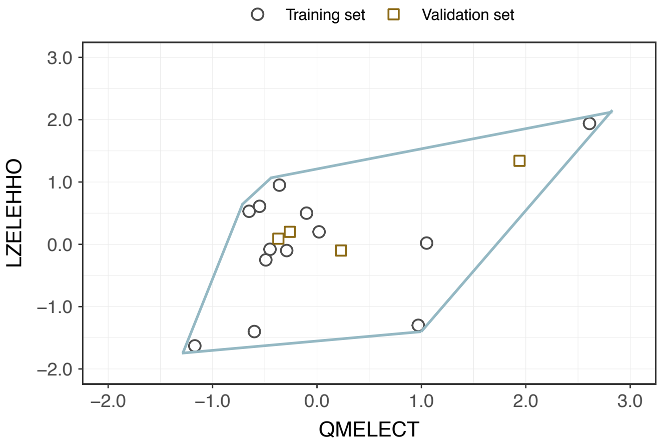
MeOx	Average Euclidean distance	Standardized cross-validated residuals	Set
Al_2O_3	1.50	1.26	1
CoO	1.22	-0.72	1
Cr_2O_3	1.33	-0.36	1
CuO	1.72	0.22	1
In_2O_3	1.34	1.22	1
TiO_2	1.23	-1.07	1
Y_2O_3	1.90	1.67	1
ZnO	2.12	-0.50	1
ZrO_2	1.18	-1.18	1
Fe_2O_3	1.17	-1.68	1
La_2O_3	2.30	-0.09	1
Sb_2O_3	1.14	-0.56	1
SiO_2	3.60	0.80	1
NiO	1.07	1.19	2
SnO_2	2.60	-0.58	2
V_2O_3	1.18	-0.52	2
Bi_2O_3	1.06	0.92	2

Table S3. Results on the applicability domain determination using: range-based method, geometrical-based method, distance-based method and standardization approach

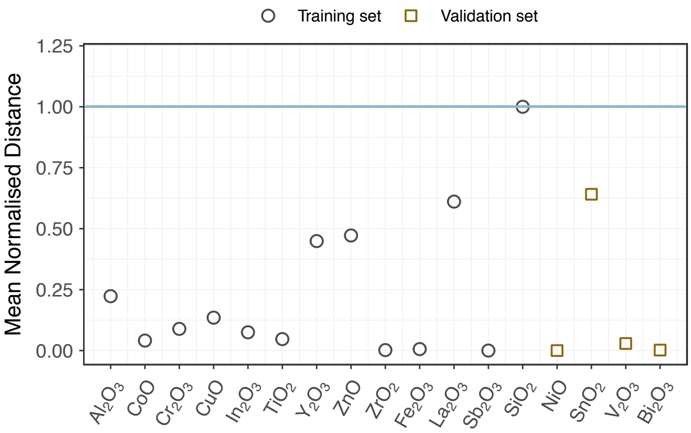
Range-based approach



Geometric approach



Basic theory of the standardization approach		
MeOx	**Set**	**AD Info.**
Al_2O_3	T	not an X-outlier
CoO	T	not an X-outlier
Cr_2O_3	T	not an X-outlier
CuO	T	not an X-outlier
In_2O_3	T	not an X-outlier
TiO_2	T	not an X-outlier
Y_2O_3	T	not an X-outlier
ZnO	T	not an X-outlier
ZrO_2	T	not an X-outlier
Fe_2O_3	T	not an X-outlier
La_2O_3	T	not an X-outlier

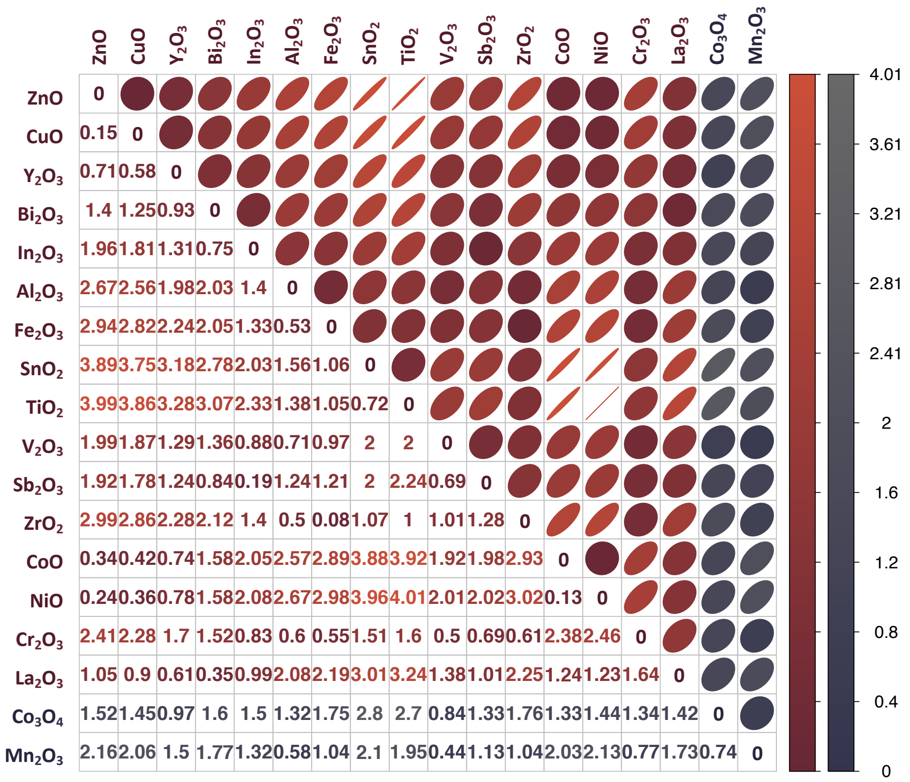
Distance-based approach (based on distance scores calculated by the Euclidean distance norm)		**Sb_2O_3**	T	not an X-outlier
		SiO_2	T	not an X-outlier
		NiO	V	within AD
		SnO_2	V	within AD
		V_2O_3	V	within AD
		Bi_2O_3	V	within AD

Case Study 2

Table S4. Descriptors used for Nano-QSAR model development along with experimentally observed and predicted values of nanotoxicity to *E. coli* bacteria for training and validation compounds (MeOx NPs) reported by Mu *et al.*[2]

MeOx	ΔH_{me+}	Z/r	Autoscaled ΔH_{me+}	Autoscaled Z/r	Set*	Experimental $\log(1/EC_{50})$	Predicted $\log(1/EC_{50})$	Residuals	Standardized residuals	Leverages
ZnO	662.4	2.70	-1.29	-1.12	T	3.45	3.39	0.06	0.31	0.17
CuO	713.7	2.74	-1.14	-1.09	T	3.20	3.35	-0.15	-0.88	0.15
Y_2O_3	837.2	3.33	-0.77	-0.64	T	2.87	3.13	-0.26	-1.51	0.10
Bi_2O_3	1137.4	2.91	0.10	-0.96	T	2.82	2.87	-0.05	-0.32	0.31
In_2O_3	1271.1	3.75	0.50	-0.32	T	2.81	2.62	0.19	1.05	0.20
Al_2O_3	1187.8	5.56	0.25	1.06	T	2.49	2.46	0.03	0.14	0.23
Fe_2O_3	1363.4	5.46	0.77	0.98	T	2.29	2.25	0.04	0.19	0.13
SnO_2	1717.3	5.80	1.80	1.24	T	2.01	1.89	0.12	0.65	0.29
TiO_2	1575.7	6.56	1.39	1.83	T	1.74	1.95	-0.21	-1.22	0.29
V_2O_3	1097.7	4.69	-0.01	0.40	T	3.14	2.69	0.45	2.52	0.10
Sb_2O_3	1233.1	3.95	0.38	-0.17	T	2.64	2.62	0.02	0.08	0.13
ZrO_2	1357.7	5.56	0.75	1.06	T	2.15	2.31	-0.16	-0.94	0.14
CoO	594.6	3.08	-1.48	-0.84	T	3.51	3.39	0.12	0.65	0.25
NiO	596.9	2.90	-1.48	-0.97	T	3.45	3.42	0.03	0.14	0.22
Cr_2O_3	1266.6	4.84	0.48	0.51	T	2.51	2.48	0.03	0.14	0.08
La_2O_3	1017.2	2.91	-0.25	-0.96	T	2.87	3.04	-0.17	-1.00	0.19
Co_3O_4	811.1	4.60	-0.85	0.33	V	3.00	2.94	0.06	0.31	0.36
Mn_2O_3	1018.0	5.17	-0.25	0.77	V	3.08	2.77	0.36	2.01	0.28

* *T – training set; V – validation set; ΔH_{me+} – the enthalpy of formation of a gaseous cation; Z/r – polarization force parameters.*



	ZnO	CuO	Y_2O_3	Bi_2O_3	In_2O_3	Al_2O_3	Fe_2O_3	SnO_2	TiO_2	V_2O_3	Sb_2O_3	ZrO_2	CoO	NiO	Cr_2O_3	La_2O_3	Co_3O_4	Mn_2O_3
ZnO	0																	
CuO	0.15	0																
Y_2O_3	0.71	0.58	0															
Bi_2O_3	1.4	1.25	0.93	0														
In_2O_3	1.96	1.81	1.31	0.75	0													
Al_2O_3	2.67	2.56	1.98	2.03	1.4	0												
Fe_2O_3	2.94	2.82	2.24	2.05	1.33	0.53	0											
SnO_2	3.89	3.75	3.18	2.78	2.03	1.56	1.06	0										
TiO_2	3.99	3.86	3.28	3.07	2.33	1.38	1.05	0.72	0									
V_2O_3	1.99	1.87	1.29	1.36	0.88	0.71	0.97	2	2	0								
Sb_2O_3	1.92	1.78	1.24	0.84	0.19	1.24	1.21	2	2.24	0.69	0							
ZrO_2	2.99	2.86	2.28	2.12	1.4	0.5	0.08	1.07	1	1.01	1.28	0						
CoO	0.34	0.42	0.74	1.58	2.05	2.57	2.89	3.88	3.92	1.92	1.98	2.93	0					
NiO	0.24	0.36	0.78	1.58	2.08	2.67	2.98	3.96	4.01	2.01	2.02	3.02	0.13	0				
Cr_2O_3	2.41	2.28	1.7	1.52	0.83	0.6	0.55	1.51	1.6	0.5	0.69	0.61	2.38	2.46	0			
La_2O_3	1.05	0.9	0.61	0.35	0.99	2.08	2.19	3.01	3.24	1.38	1.01	2.25	1.24	1.23	1.64	0		
Co_3O_4	1.52	1.45	0.97	1.6	1.5	1.32	1.75	2.8	2.7	0.84	1.33	1.76	1.33	1.44	1.34	1.42	0	
Mn_2O_3	2.16	2.06	1.5	1.77	1.32	0.58	1.04	2.1	1.95	0.44	1.13	1.04	2.03	2.13	0.77	1.73	0.74	0

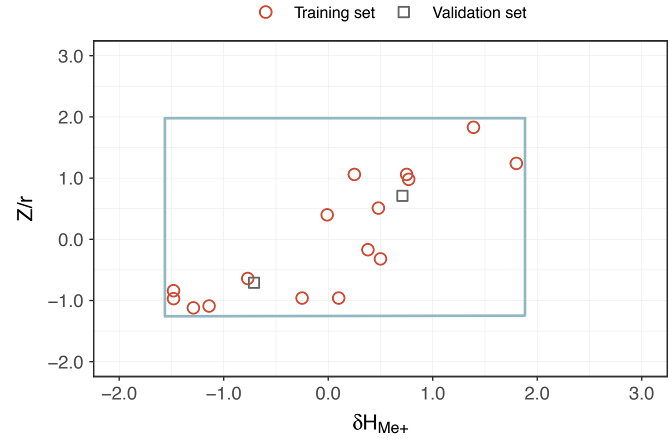
Figure S2. Euclidean distance matrix between the training (indicated in tomato color) and validation (indicated in gray color) compounds. The color intensity and shape of the ellipses displayed in the upper triangular part of a distance matrix reflect the values of the corresponding Euclidean distances expressed in numerical forms in the lower triangular part of this matrix. The smallest Euclidean distances (indicated as big circles) are considered as significant. In contrast, the highest distances that reflect the dissimilarity between compounds are considered as insignificant and shown as a thin ellipse.

Table S5. Input matrix data for 95% and 99% confidence interval boundaries estimation using the Student's *t*-distribution and developed in-house *R*-code

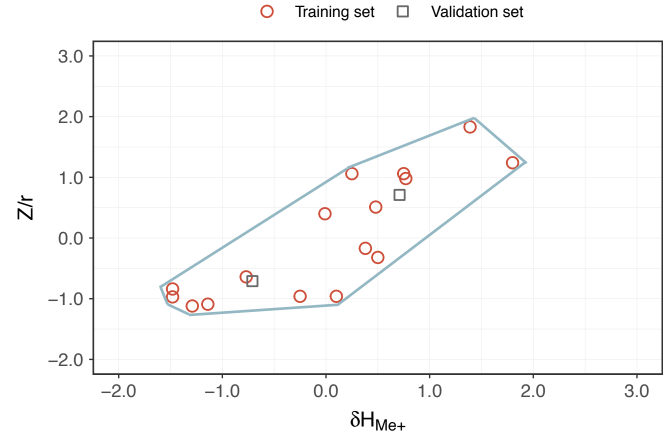
MeOx	Average Euclidean distance	Standardized residuals	Set
ZnO	1.91	0.31	1
CuO	1.82	-0.88	1
Y_2O_3	1.52	-1.51	1
Bi_2O_3	1.57	-0.32	1
In_2O_3	1.42	1.05	1
Al_2O_3	1.63	0.14	1
Fe_2O_3	1.66	0.19	1
SnO_2	2.43	0.65	1
TiO_2	2.51	-1.22	1
V_2O_3	1.37	2.52	1
Sb_2O_3	1.36	0.08	1
ZrO_2	1.69	-0.94	1
CoO	1.93	0.65	1
NiO	1.97	0.14	1
Cr_2O_3	1.42	0.14	1
La_2O_3	1.54	-1.00	1
Co_3O_4	1.57	0.31	2
Mn_2O_3	1.49	2.01	2

Table S6. Results on the applicability domain determination using: range-based method, geometrical-based method, distance-based method and standardization approach

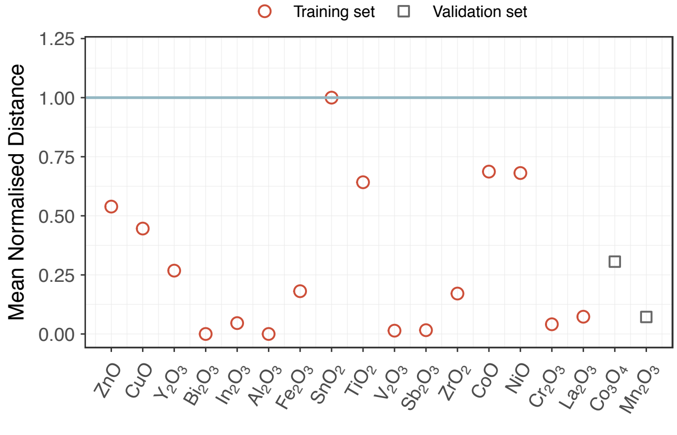
Range-based approach



Geometric approach



Distance-based approach (based on distance scores calculated by the Euclidean distance norm)



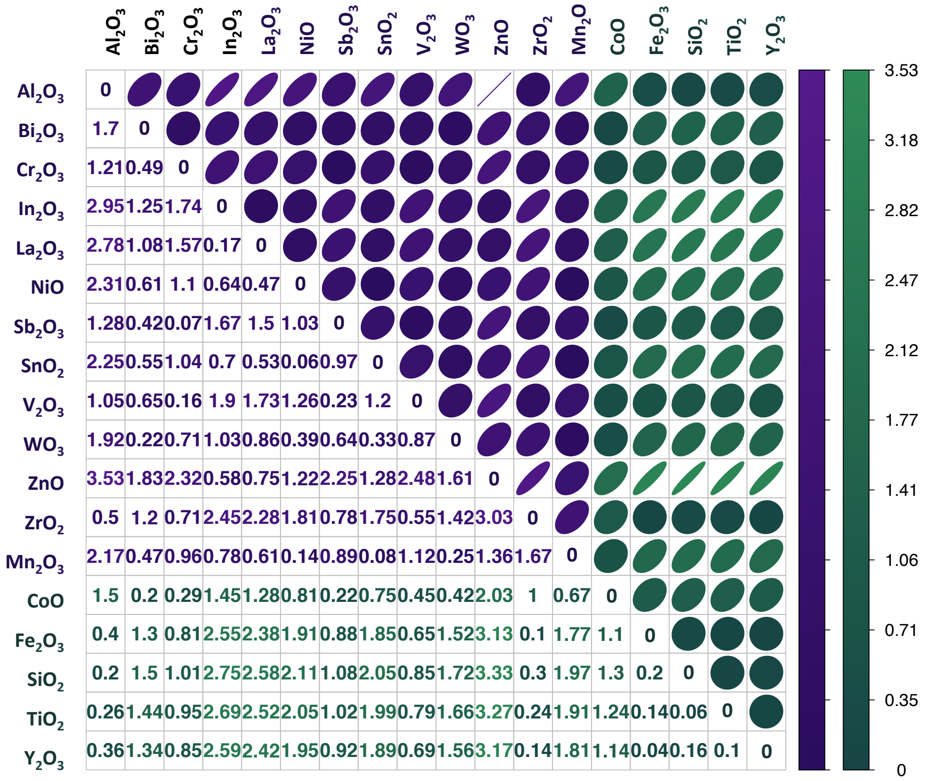
Basic theory of the standardization approach		
MeOx	**Set**	**AD Info.**
ZnO	T	not an X-outlier
CuO	T	not an X-outlier
Y_2O_3	T	not an X-outlier
Bi_2O_3	T	not an X-outlier
In_2O_3	T	not an X-outlier
Al_2O_3	T	not an X-outlier
Fe_2O_3	T	not an X-outlier
SnO_2	T	not an X-outlier
TiO_2	T	not an X-outlier
V_2O_3	T	not an X-outlier
Sb_2O_3	T	not an X-outlier
ZrO_2	T	not an X-outlier
CoO	T	not an X-outlier
NiO	T	not an X-outlier
Cr_2O_3	T	not an X-outlier
La_2O_3	T	not an X-outlier
Co_3O_4	V	within AD
Mn_2O_3	V	within AD

Case Study 3

Table S7. Descriptors used for Nano-QSAR model development along with experimentally observed and predicted values of nanotoxicity to HaCaT cells for training and validation compounds (MeOx NPs) reported by Pan *et al.*[3]

MeOx	DCW (1,3)	Autoscaled DCW(1,3)	Set*	Experimental $\log(1/LC_{50})$	Predicted $\log(1/LC_{50})$	Residuals	Standardized cross-validated residuals	Leverages
Al_2O_3	20.32	-1.82	T	1.85	1.82	0.03	-0.01	0.35
Bi_2O_3	26.58	-0.12	T	2.50	2.47	0.03	-0.10	0.08
Cr_2O_3	24.77	-0.61	T	2.30	2.28	0.02	-0.18	0.11
In_2O_3	31.20	1.13	T	2.92	2.95	-0.03	-0.54	0.18
La_2O_3	30.57	0.96	T	2.87	2.88	-0.01	-0.42	0.15
NiO	28.85	0.49	T	2.49	2.70	-0.21	-1.86	0.10
Sb_2O_3	25.06	-0.54	T	2.31	2.31	0.00	-0.32	0.10
SnO_2	28.62	0.43	T	2.67	2.68	-0.01	-0.40	0.09
V_2O_3	24.20	-0.77	T	2.24	2.22	0.02	-0.19	0.13
WO_3	27.41	0.10	T	2.56	2.55	0.01	-0.28	0.08
ZnO	33.34	1.71	T	3.32	3.17	0.15	1.11	0.32
ZrO_2	22.16	-1.32	T	2.02	2.01	0.01	-0.24	0.22
Mn_2O_3	28.31	0.35	T	2.64	2.65	-0.01	-0.38	0.09
CoO	25.85	-0.32	V	2.83	2.39	0.44	2.51	0.09
Fe_2O_3	21.79	-1.42	V	2.05	1.97	0.08	0.18	0.24
SiO_2	21.06	-1.62	V	2.12	1.90	0.23	1.13	0.30
TiO_2	21.28	-1.56	V	1.76	1.92	-0.16	-1.35	0.28
Y_2O_3	21.64	-1.46	V	2.21	1.96	0.26	1.33	0.26

** T – training set; V – validation set; DCW(1,3) – SMILES-based optimal molecular descriptor which includes: aggregation size; individual size; mass percentage of metal elements; cationic charge and molecular weight.*



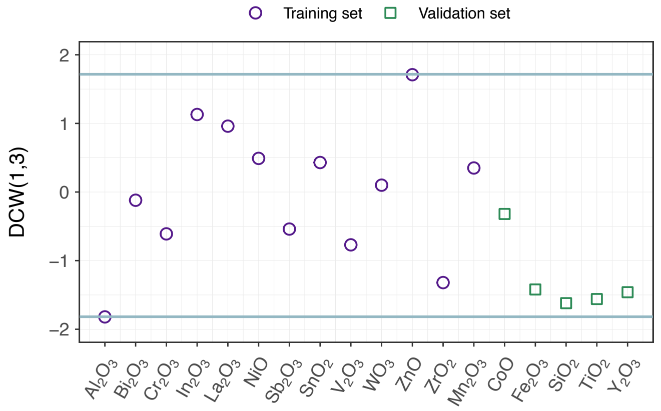
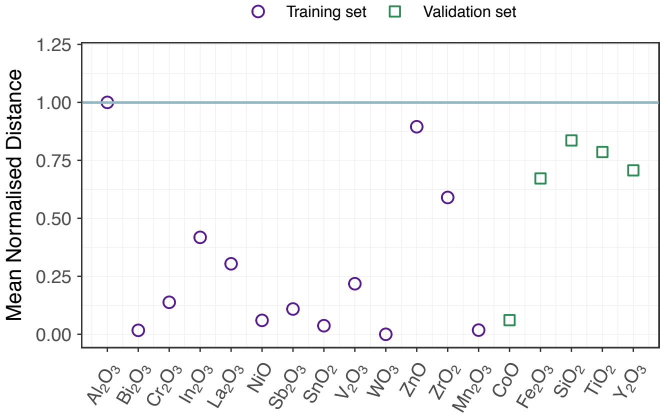
	Al_2O_3	Bi_2O_3	Cr_2O_3	In_2O_3	La_2O_3	NiO	Sb_2O_3	SnO_2	V_2O_3	WO_3	ZnO	ZrO_2	Mn_2O	CoO	Fe_2O_3	SiO_2	TiO_2	Y_2O_3
Al_2O_3	0																	
Bi_2O_3	1.7	0																
Cr_2O_3	1.21	0.49	0															
In_2O_3	2.95	1.25	1.74	0														
La_2O_3	2.78	1.08	1.57	0.17	0													
NiO	2.31	0.61	1.1	0.64	0.47	0												
Sb_2O_3	1.28	0.42	0.07	1.67	1.5	1.03	0											
SnO_2	2.25	0.55	1.04	0.7	0.53	0.06	0.97	0										
V_2O_3	1.05	0.65	0.16	1.9	1.73	1.26	0.23	1.2	0									
WO_3	1.92	0.22	0.71	1.03	0.86	0.39	0.64	0.33	0.87	0								
ZnO	3.53	1.83	2.32	0.58	0.75	1.22	2.25	1.28	2.48	1.61	0							
ZrO_2	0.5	1.2	0.71	2.45	2.28	1.81	0.78	1.75	0.55	1.42	3.03	0						
Mn_2O_3	2.17	0.47	0.96	0.78	0.61	0.14	0.89	0.08	1.12	0.25	1.36	1.67	0					
CoO	1.5	0.2	0.29	1.45	1.28	0.81	0.22	0.75	0.45	0.42	2.03	1	0.67	0				
Fe_2O_3	0.4	1.3	0.81	2.55	2.38	1.91	0.88	1.85	0.65	1.52	3.13	0.1	1.77	1.1	0			
SiO_2	0.2	1.5	1.01	2.75	2.58	2.11	1.08	2.05	0.85	1.72	3.33	0.3	1.97	1.3	0.2	0		
TiO_2	0.26	1.44	0.95	2.69	2.52	2.05	1.02	1.99	0.79	1.66	3.27	0.24	1.91	1.24	0.14	0.06	0	
Y_2O_3	0.36	1.34	0.85	2.59	2.42	1.95	0.92	1.89	0.69	1.56	3.17	0.14	1.81	1.14	0.04	0.16	0.1	0

Figure S3. Euclidean distance matrix between the training (indicated in purple color) and validation (indicated in seagreen color) compounds. The color intensity and shape of the ellipses displayed in the upper triangular part of a distance matrix reflect the values of the corresponding Euclidean distances expressed in numerical forms in the lower triangular part of this matrix. The smallest Euclidean distances (indicated as big circles) are considered as significant. In contrast, the highest distances that reflect the dissimilarity between compounds are considered as insignificant and shown as a thin ellipse.

Table S8. Input matrix data for 95% and 99% confidence interval boundaries estimation using the Student's *t*-distribution and developed in-house *R*-code

MeOx	Average Euclidean distance	Standardized cross-validated residuals	Set
Al_2O_3	1.97	-0.01	1
Bi_2O_3	0.87	-0.10	1
Cr_2O_3	1.01	-0.18	1
In_2O_3	1.32	-0.54	1
La_2O_3	1.19	-0.42	1
NiO	0.92	-1.86	1
Sb_2O_3	0.98	-0.32	1
SnO_2	0.90	-0.40	1
V_2O_3	1.10	-0.19	1
WO_3	0.85	-0.28	1
ZnO	1.85	1.11	1
ZrO_2	1.51	-0.24	1
Mn_2O_3	0.88	-0.38	1
CoO	0.85	2.51	2
Fe_2O_3	1.48	0.18	2
SiO_2	1.65	1.13	2
TiO_2	1.60	-1.35	2
Y_2O_3	1.51	1.33	2

Table S9. Results on the applicability domain determination using: range-based method, geometrical-based method, distance-based method and standardization approach

Range-based approach	
Distance-based approach (based on distance scores calculated by the Euclidean distance norm)	

Basic theory of the standardization approach

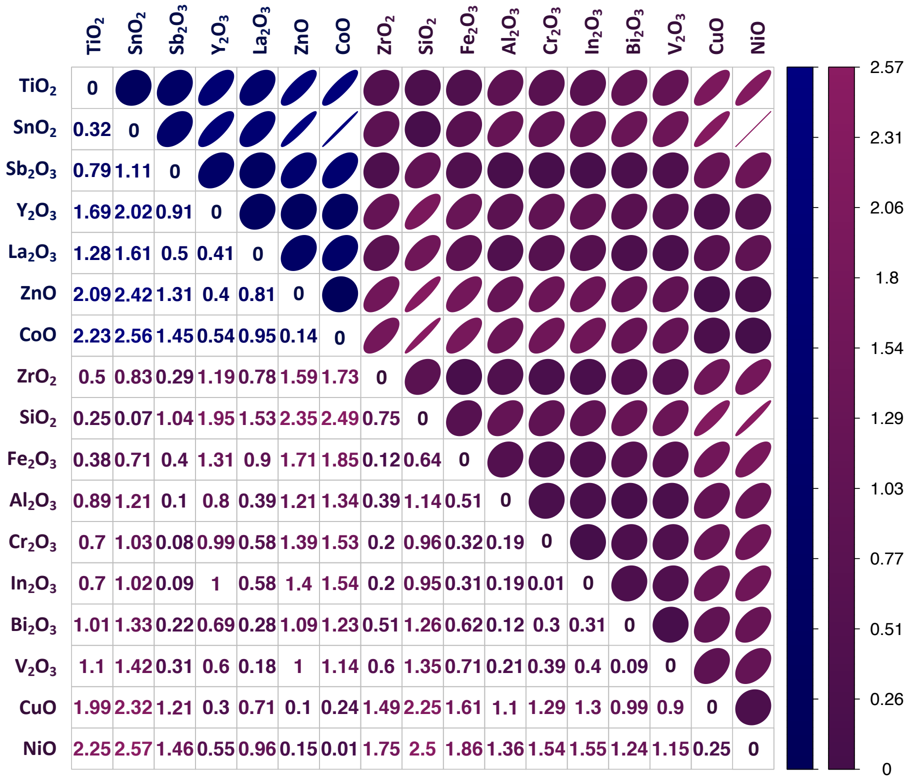
MeOx	Set	AD Info.
Al_2O_3	T	not an X-outlier
Bi_2O_3	T	not an X-outlier
Cr_2O_3	T	not an X-outlier
In_2O_3	T	not an X-outlier
La_2O_3	T	not an X-outlier
NiO	T	not an X-outlier
Sb_2O_3	T	not an X-outlier
SnO_2	T	not an X-outlier
V_2O_3	T	not an X-outlier
WO_3	T	not an X-outlier
ZnO	T	not an X-outlier
ZrO_2	T	not an X-outlier
Mn_2O_3	T	not an X-outlier
CoO	V	within AD
Fe_2O_3	V	within AD
SiO_2	V	within AD
TiO_2	V	within AD
Y_2O_3	V	within AD

Case Study 4

Table S10. Descriptors used for read-across model development along with experimentally observed and predicted values of nanotoxicity to *E. coli* bacteria for training and validation compounds (MeOx NPs) reported by Gajewicz[4]

MeOx	ΔH_{Me+} [kcal/mol]	Autoscaled ΔH_{Me+}	Set*	Experimental $\log(EC_{50})^{-1}$	Predicted $\log(EC_{50})^{-1}$	Residuals	Standardized residuals	Leverages
TiO_2	1575.73	1.11	T	1.74	2.10	-0.36	-1.56	0.35
SnO_2	1717.32	1.43	T	2.01	1.81	0.20	1.06	0.49
Sb_2O_3	1233.06	0.32	T	2.64	2.56	0.08	0.51	0.16
Y_2O_3	837.15	-0.58	T	2.87	3.16	-0.29	-1.24	0.20
La_2O_3	1017.22	-0.17	T	2.87	2.78	0.09	0.56	0.15
ZnO	662.44	-0.99	T	3.45	3.45	0.00	0.11	0.30
CoO	601.80	-1.12	T	3.51	3.42	0.09	0.55	0.35
ZrO_2	1357.66	0.61	V	2.15	2.41	-0.26	-1.12	0.20
SiO_2	1686.38	1.36	V	2.20	1.99	0.21	1.10	0.45
Fe_2O_3	1408.29	0.73	V	2.29	2.16	0.13	0.75	0.23
Al_2O_3	1187.83	0.22	V	2.49	2.65	-0.16	-0.65	0.15
Cr_2O_3	1268.70	0.41	V	2.51	2.65	-0.14	-0.51	0.17
In_2O_3	1271.13	0.41	V	2.81	2.65	0.16	0.89	0.17
Bi_2O_3	1137.40	0.10	V	2.82	2.73	0.09	0.53	0.14
V_2O_3	1097.73	0.01	V	3.14	2.81	0.33	1.66	0.14
CuO	706.25	-0.89	V	3.20	3.46	-0.26	-1.10	0.27
NiO	596.70	-1.14	V	3.45	3.51	-0.06	-0.16	0.36

** T – training set; V – validation set; ΔH_{Me+} – the enthalpy of the formation of gaseous cations having the same oxidation state as that in the metal oxide structure.*



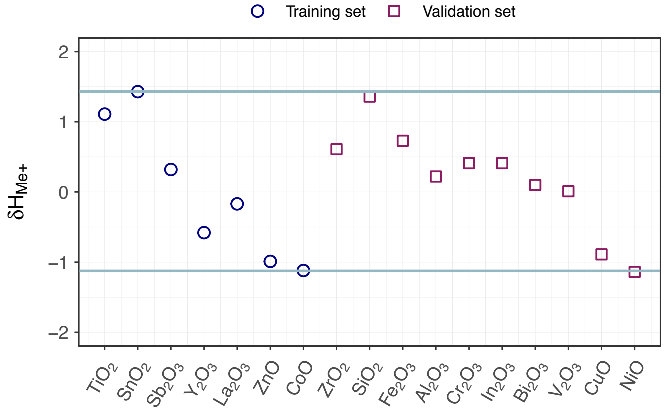
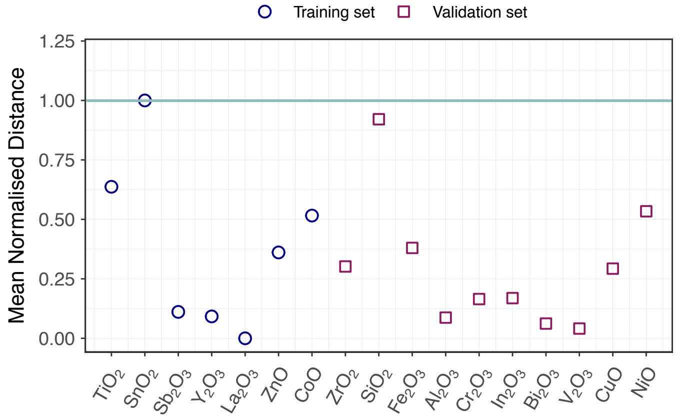
	TiO_2	SnO_2	Sb_2O_3	Y_2O_3	La_2O_3	ZnO	CoO	ZrO_2	SiO_2	Fe_2O_3	Al_2O_3	Cr_2O_3	In_2O_3	Bi_2O_3	V_2O_3	CuO	NiO
TiO_2	0																
SnO_2	0.32	0															
Sb_2O_3	0.79	1.11	0														
Y_2O_3	1.69	2.02	0.91	0													
La_2O_3	1.28	1.61	0.5	0.41	0												
ZnO	2.09	2.42	1.31	0.4	0.81	0											
CoO	2.23	2.56	1.45	0.54	0.95	0.14	0										
ZrO_2	0.5	0.83	0.29	1.19	0.78	1.59	1.73	0									
SiO_2	0.25	0.07	1.04	1.95	1.53	2.35	2.49	0.75	0								
Fe_2O_3	0.38	0.71	0.4	1.31	0.9	1.71	1.85	0.12	0.64	0							
Al_2O_3	0.89	1.21	0.1	0.8	0.39	1.21	1.34	0.39	1.14	0.51	0						
Cr_2O_3	0.7	1.03	0.08	0.99	0.58	1.39	1.53	0.2	0.96	0.32	0.19	0					
In_2O_3	0.7	1.02	0.09	1	0.58	1.4	1.54	0.2	0.95	0.31	0.19	0.01	0				
Bi_2O_3	1.01	1.33	0.22	0.69	0.28	1.09	1.23	0.51	1.26	0.62	0.12	0.3	0.31	0			
V_2O_3	1.1	1.42	0.31	0.6	0.18	1	1.14	0.6	1.35	0.71	0.21	0.39	0.4	0.09	0		
CuO	1.99	2.32	1.21	0.3	0.71	0.1	0.24	1.49	2.25	1.61	1.1	1.29	1.3	0.99	0.9	0	
NiO	2.25	2.57	1.46	0.55	0.96	0.15	0.01	1.75	2.5	1.86	1.36	1.54	1.55	1.24	1.15	0.25	0

Figure S4. Euclidean distance matrix between the training (indicated in navy color) and validation (indicated in maroon color) compounds. The color intensity and shape of the ellipses displayed in the upper triangular part of a distance matrix reflect the values of the corresponding Euclidean distances expressed in numerical forms in the lower triangular part of this matrix. The smallest Euclidean distances (indicated as big circles) are considered as significant. In contrast, the highest distances that reflect the dissimilarity between compounds are considered as insignificant and shown as a thin ellipse.

Table S11. Input matrix data for 95% and 99% confidence interval boundaries estimation using the Student's *t*-distribution and developed in-house *R*-code

MeOx	Average Euclidean distance	Standardized residuals	Set
TiO_2	1.40	-1.56	1
SnO_2	1.67	1.06	1
Sb_2O_3	1.01	0.51	1
Y_2O_3	0.99	-1.24	1
La_2O_3	0.93	0.56	1
ZnO	1.20	0.11	1
CoO	1.31	0.55	1
ZrO_2	0.99	-1.12	2
SiO_2	1.38	1.10	2
Fe_2O_3	1.04	0.75	2
Al_2O_3	0.85	-0.65	2
Cr_2O_3	0.90	-0.51	2
In_2O_3	0.90	0.89	2
Bi_2O_3	0.83	0.53	2
V_2O_3	0.82	1.66	2
CuO	0.98	-1.10	2
NiO	1.14	-0.16	2

Table S12. Results on the applicability domain determination using: range-based method, geometrical-based method, distance-based method and standardization approach

Range-based approach	
Distance-based approach (based on distance scores calculated by the Euclidean distance norm)	

Basic theory of the standardization approach		
MeOx	**Set**	**AD Info.**
TiO_2	T	not an X-outlier
SnO_2	T	not an X-outlier
Sb_2O_3	T	not an X-outlier
Y_2O_3	T	not an X-outlier
La_2O_3	T	not an X-outlier
ZnO	T	not an X-outlier
CoO	T	not an X-outlier
ZrO_2	V	within AD
SiO_2	V	within AD
Fe_2O_3	V	within AD
Al_2O_3	V	within AD
Cr_2O_3	V	within AD
In_2O_3	V	within AD
Bi_2O_3	V	within AD
V_2O_3	V	within AD
CuO	V	within AD
NiO	V	within AD

References

1. Pathakoti, K.; Huang, M. J.; Watts, J. D.; He, X.; Hwang, H. M., Using experimental data of Escherichia coli to develop a QSAR model for predicting the photo-induced cytotoxicity of metal oxide nanoparticles. *Journal of photochemistry and photobiology. B, Biology* **2014,** *130*, 234-40.

2. Mu, Y.; Wu, F.; Zhao, Q.; Ji, R.; Qie, Y.; Zhou, Y.; Hu, Y.; Pang, C.; Hristozov, D.; Giesy, J. P.; Xing, B., Predicting Toxic Potencies of Metal Oxide Nanoparticles by Means of Nano-QSARs. *Nanotoxicology* **2016,** *10*, 1207-1214.

3. Pan, Y.; Li, T.; Cheng, J.; Telesca, D.; Zink, J. I.; Jiang, J., Nano-QSAR modeling for predicting the cytotoxicity of metal oxide nanoparticles using novel descriptors. *RSC Advances* **2016,** *6*, 25766-25775.

4. Gajewicz, A., What if the number of nanotoxicity data is too small for developing predictive Nano-QSAR models? An alternative read-across based approach for filling data gaps. *Nanoscale* **2017,** *9*, 8435-8448.