Electronic Supplementary Material (ESI) for Physical Chemistry Chemical Physics.

Supporting Information for Publication

Exploring Absolute Yield Curve of Secondary Electrons by Using Machine Learning Methods

[1]Mehnaz, [2]Bo Da[#], [1]Z.J. Ding[*]

[1]*Hefei National Research Center for Physical Sciences at Microscale and Department of Physics, University of Science and Technology of China, Hefei, Anhui 230026, P.R. China;*

[2]*Research and Services Division of Materials Data and Integrated System, National Institute for Materials Science, 1-2-1 Sengen, Tsukuba, Ibaraki 305-0047, Japan*

[#]e-mail: DA.Bo@nims.go.jp

*e-mail: zjding@ustc.edu.cn

Contents

Table S1: Descriptors used for training the ML models

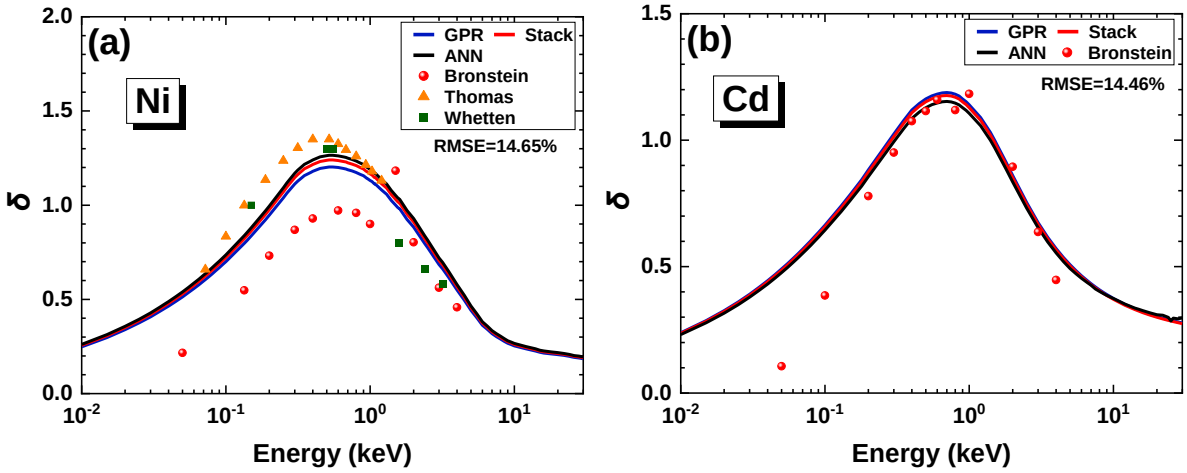
S#	Elements	*Z*	Group	Period	*N*v	At. Wt.	ρ (g/cm^3)	Work Function (eV)
1	Li	3	1	2	1	6.94	0.53	2.9
2	Be	4	2	2	2	9.01	1.85	4.98
3	B	5	13	2	3	10.81	2.34	4.45
4	C	6	14	2	4	12.01	2.26	5.0
5	Mg	12	2	3	2	24.31	1.74	3.66
6	Al	13	13	3	3	26.98	2.7	4.28
7	Si	14	14	3	4	28.09	2.33	4.85
8	K	19	1	4	1	39.10	0.86	2.3
9	Ca	20	2	4	2	40.07	1.55	2.87
10	Sc	21	3	4	3	44.96	2.99	3.5
11	Ti	22	4	4	4	47.87	4.54	4.33
12	V	23	5	4	5	50.94	6.11	4.3
13	Cr	24	6	4	6	52.00	7.19	4.5
15	Fe	26	8	4	8	55.85	7.87	4.5
16	Cu	29	11	4	11	63.55	8.96	4.65
17	Zn	30	12	4	12	65.38	7.13	4.33
18	Ga	31	13	4	3	69.72	5.91	4.2
19	Ge	32	14	4	4	72.59	5.32	5.0
20	Se	34	16	4	6	78.97	4.79	5.9
21	Sr	38	2	5	2	87.62	2.54	2.59
22	Y	39	3	5	3	88.91	4.47	3.1
23	Zr	40	4	5	4	91.22	6.51	4.05
24	Nb	41	5	5	5	92.91	8.57	4.3
25	Mo	42	6	5	6	95.94	10.28	4.6
28	Pd	46	10	5	10	106.42	12.02	5.12
29	Ag	47	11	5	11	107.87	10.50	4.26
30	In	49	13	5	3	114.82	7.31	4.12
31	Sn	50	14	5	4	118.71	7.31	4.42
32	Sb	51	15	5	5	121.76	6.68	4.7
33	Te	52	16	5	6	127.6	6.24	4.95
34	Cs	55	1	6	1	132.91	1.87	2.14
35	Ba	56	2	6	2	137.33	3.59	2.7
36	Hf	72	4	6	4	178.49	13.31	3.9
37	Ta	73	5	6	5	180.95	16.65	4.25
38	W	74	6	6	6	183.85	19.35	4.55
39	Re	75	7	6	7	186.21	21.04	4.96
41	Pt	78	10	6	10	195.08	21.45	5.65
42	Au	79	11	6	11	196.97	19.32	5.1
43	Pb	82	14	6	4	207.2	11.35	4.25
44	Bi	83	15	6	5	208.98	9.75	4.22

Table S2: LOOCV test results for each element

Element	RMSE(%)			Element	RMSE(%)		
	GPR	Stack	ANN		GPR	Stack	ANN
Li	1.96	2.05	2.65	Y	10.13	11.27	10.47
Be	7.54	7.13	8.25	Zr	3.93	4.85	4.39
B	5.64	4.36	6.32	Nb	10.63	9.67	9.67
C	6.51	7.14	7.68	Mo	10.48	10.33	11.05
Mg	9.45	8.25	9.13	Pd	7.46	8.29	7.78
Al	8.96	8.12	8.5	Ag	12.49	11.06	12.36
Si	2.26	2.79	3.56	In	3.27	3.75	4.22
K	14.5	13.6	15.42	Sn	6.43	6.13	6.12
Ca	13.21	14.05	12.35	Sb	5.54	5.28	4.97
Sc	11.94	11.11	10.8	Te	7.3	7.1	7.76
Ti	5.85	5.93	5.36	Cs	3.61	3.55	4.08
V	2.19	3.26	3.5	Ba	3.08	3.62	4.15
Cr	2.06	2.54	2.26	Hf	8.3	8.5	7.84
Fe	3.08	3.16	3.19	Ta	9.33	8.74	9.53
Cu	3.18	2.75	2.53	W	6.69	7.36	7.51
Zn	6.88	6.32	6.55	Re	2.84	3.97	3.52
Ga	9.89	9.56	9.64	Pt	14.16	13.27	12.44
Ge	5.78	5.85	5.11	Au	4.27	5.55	5.16
Se	3.54	4.66	5.3	Pb	7.85	6.17	7.13
Sr	4.06	4.51	4.82	Bi	7.24	7.7	8.08
Average	6.84	6.83	7.02				

Figure S1:

Figure S1 shows the results of ML models' validation through predictions of SEY for six unseen elements (Ni, Cd, Hg, Tl, La and Th). The experimental SEY data of these elements was initially taken aside from the training dataset and was not used to train the models. It can be observed that the ML models are capable of predicting not only the available experimental data in the high energy region, but also work well for low energies below 50 eV where there exists no experimental data for these elements. We have compared the predicted yield wherever experimental data are available. The RMSE values for the deviation of ML prediction from the measured data for each element have also been shown in Fig. S1. For Ni, there are three measured datasets. The ML predicted yield curves are in closer agreement with the measurements of Thomas [99] and Whetten [106] compared with that of Bronstein [98]. For Cd, Hg, Tl and La, the ML predicted yields are in good agreement with the measured data of Bronstein [98]. For Th, measured data are present only around the maxima, and the agreement between ML predictions and the available data is satisfactory. These results indicate that the prediction ability of our ML models for unseen data is good and we can use this model to predict SEY for other unseen elements.



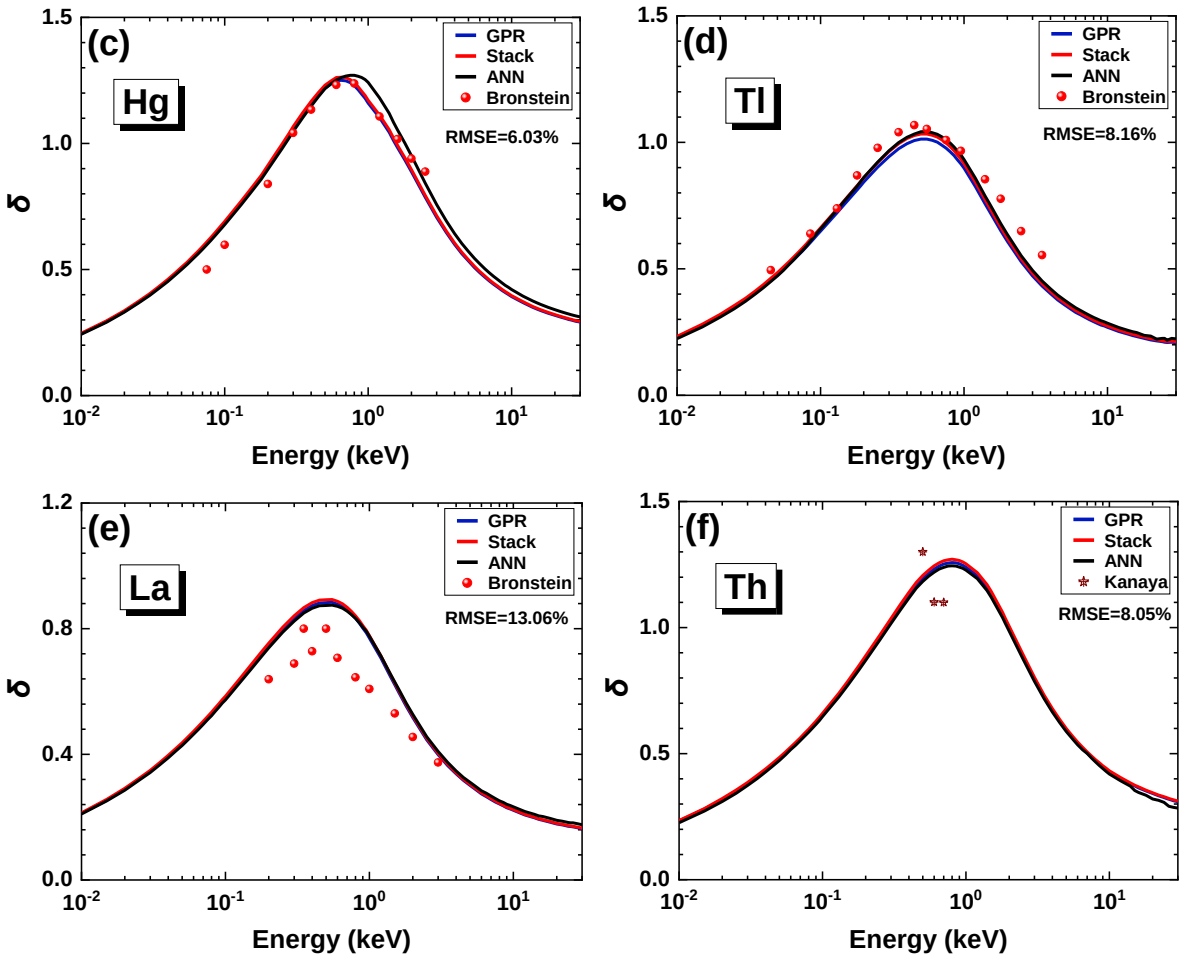
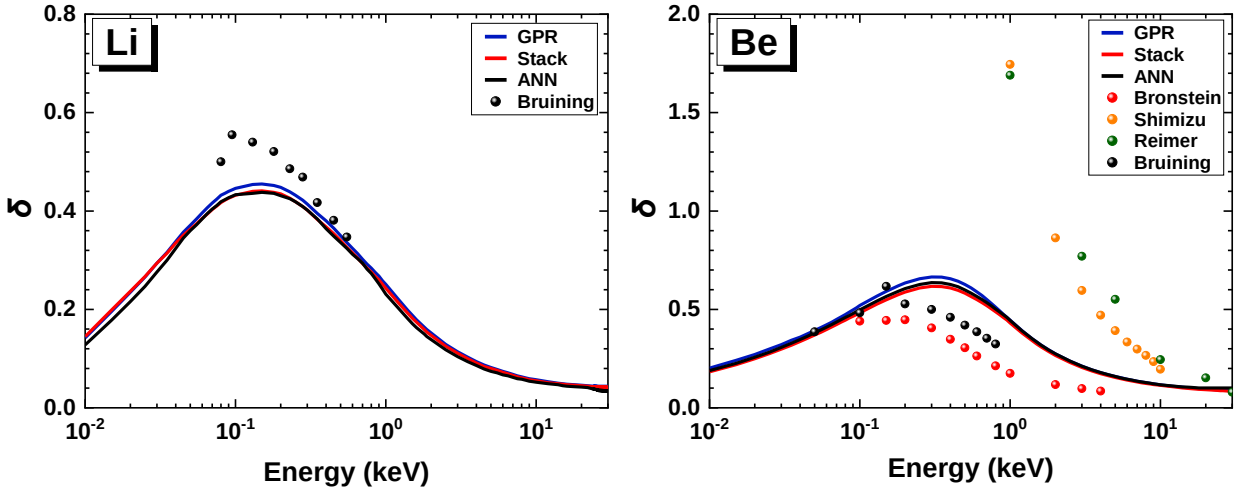


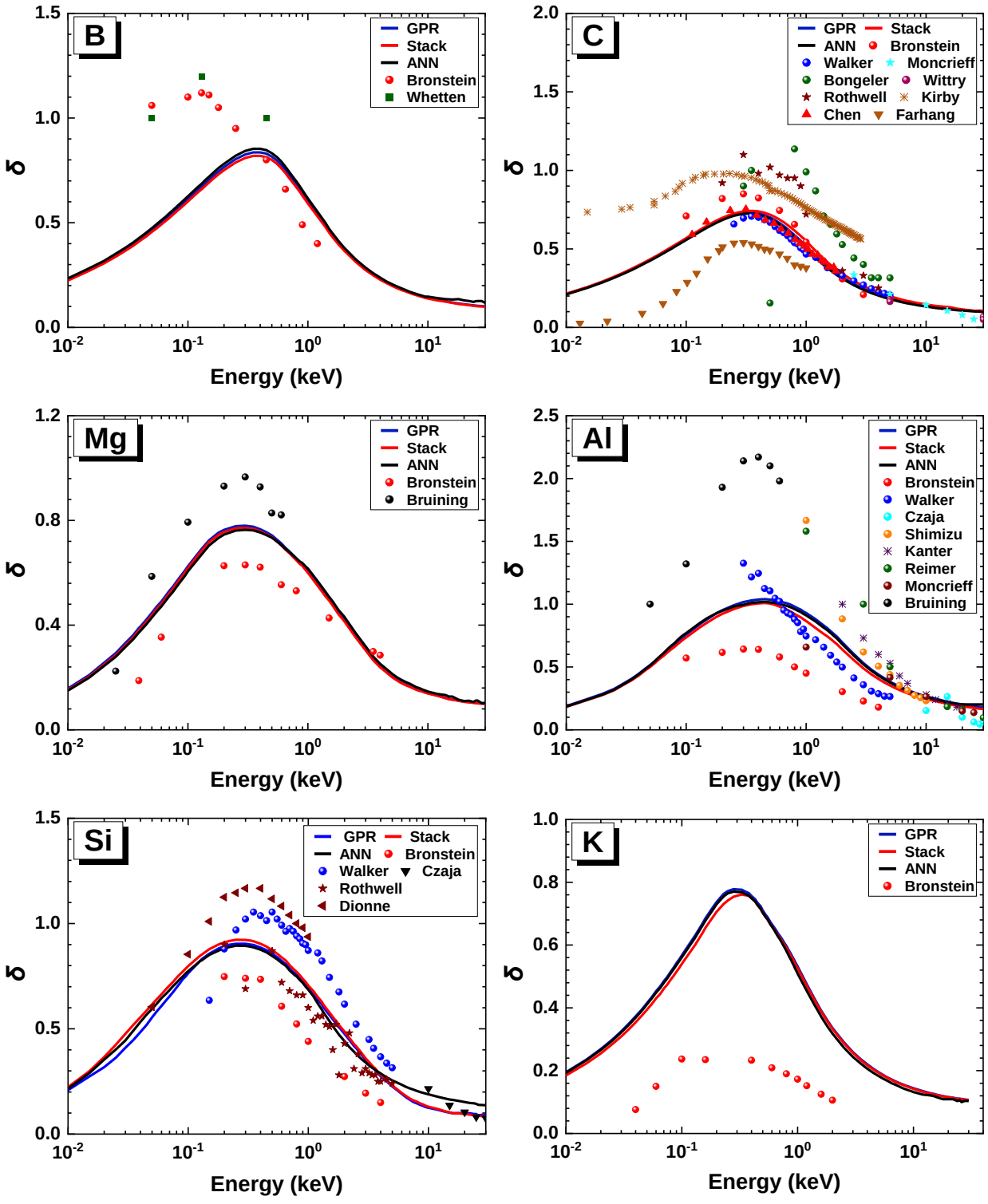
Figure S1: ML predicted yields for six unseen elements showing generalization ability of the ML models used. The solid lines represent the ML predicted yield curves whereas the measured data are shown by dots. The RMSE value is for the deviation of ML prediction (Stack generalization model) from the measured data.

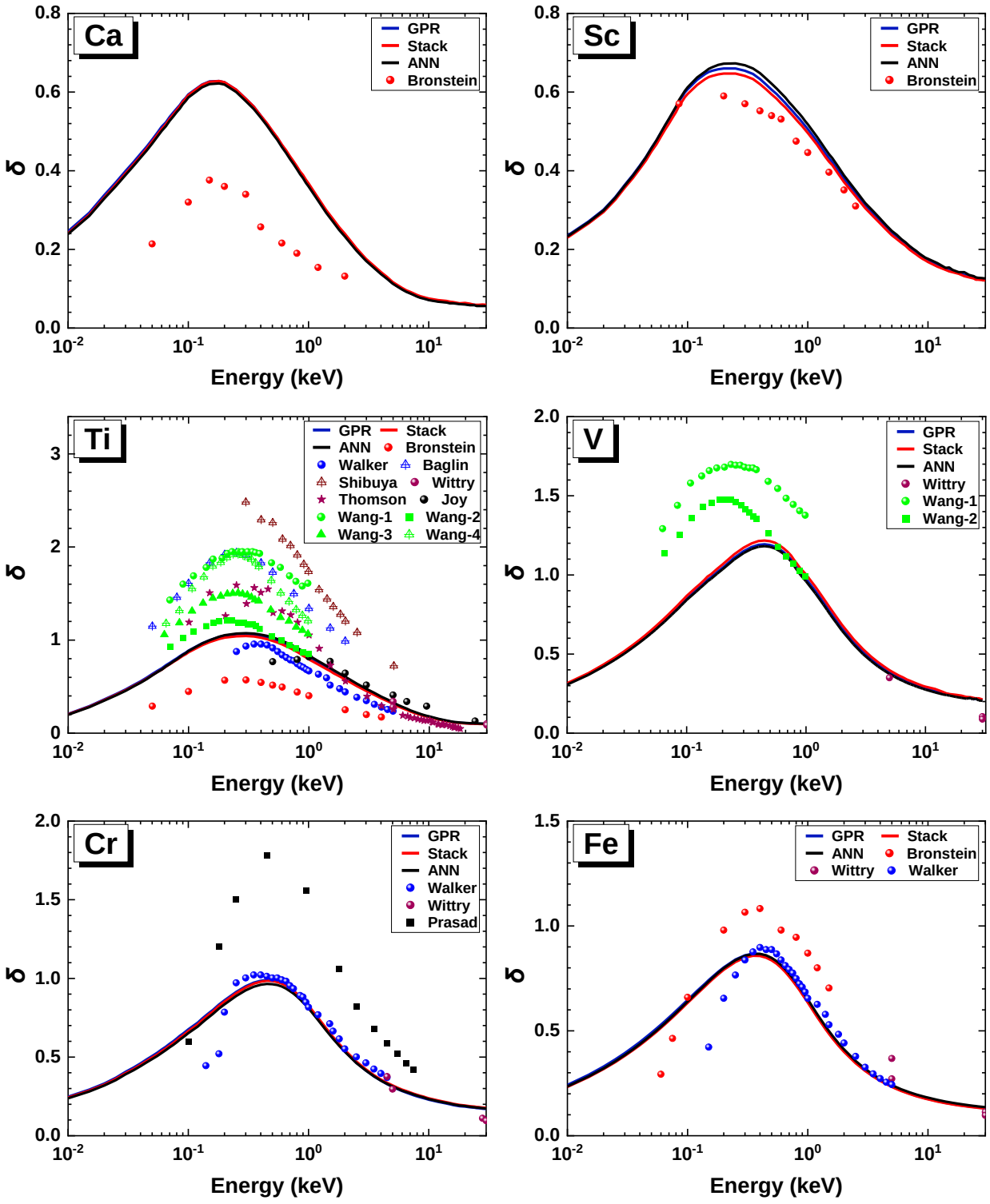
Table S3: ML predicted yield for 40 elements in the training database (Excel File).

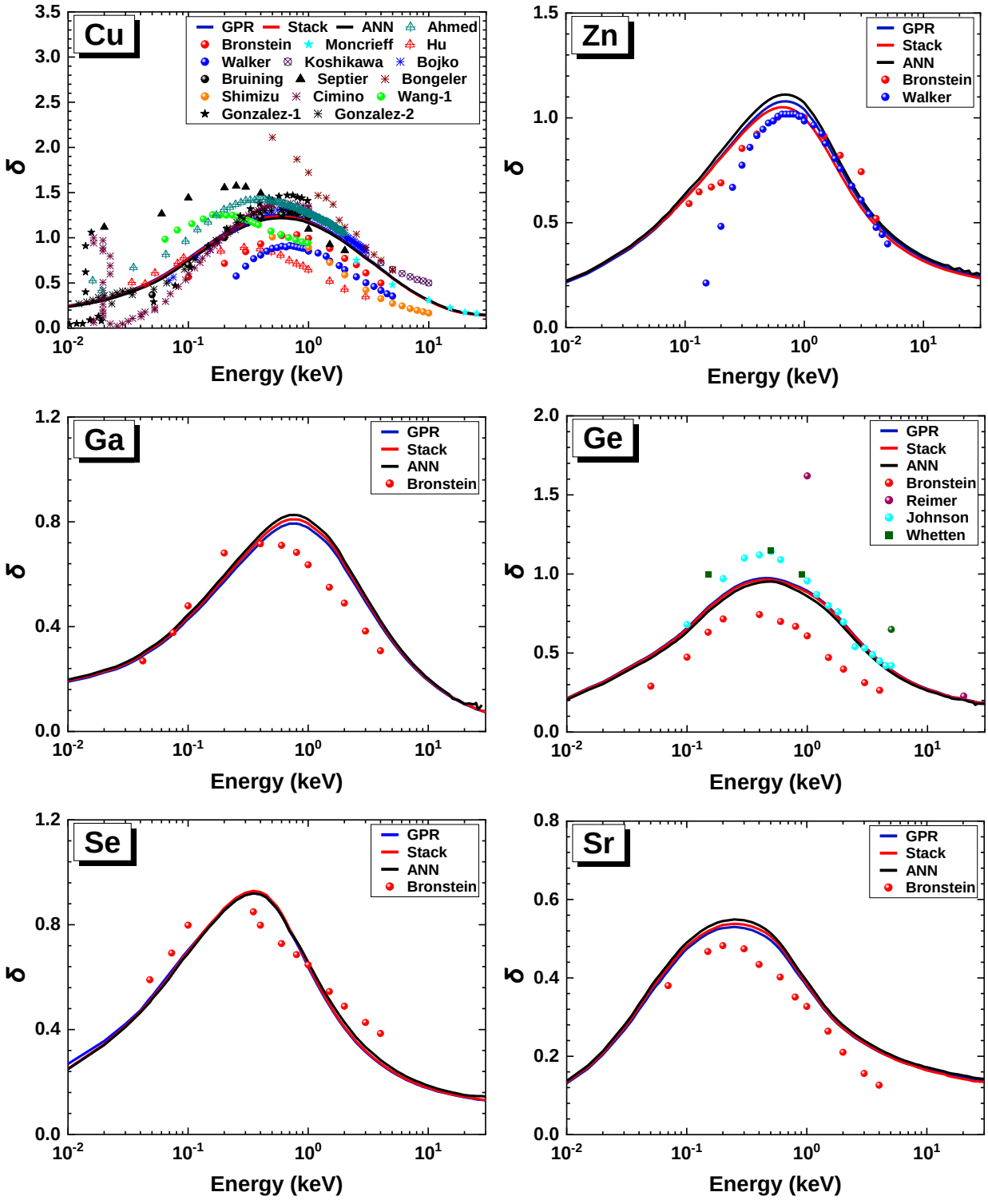
Figure S2:

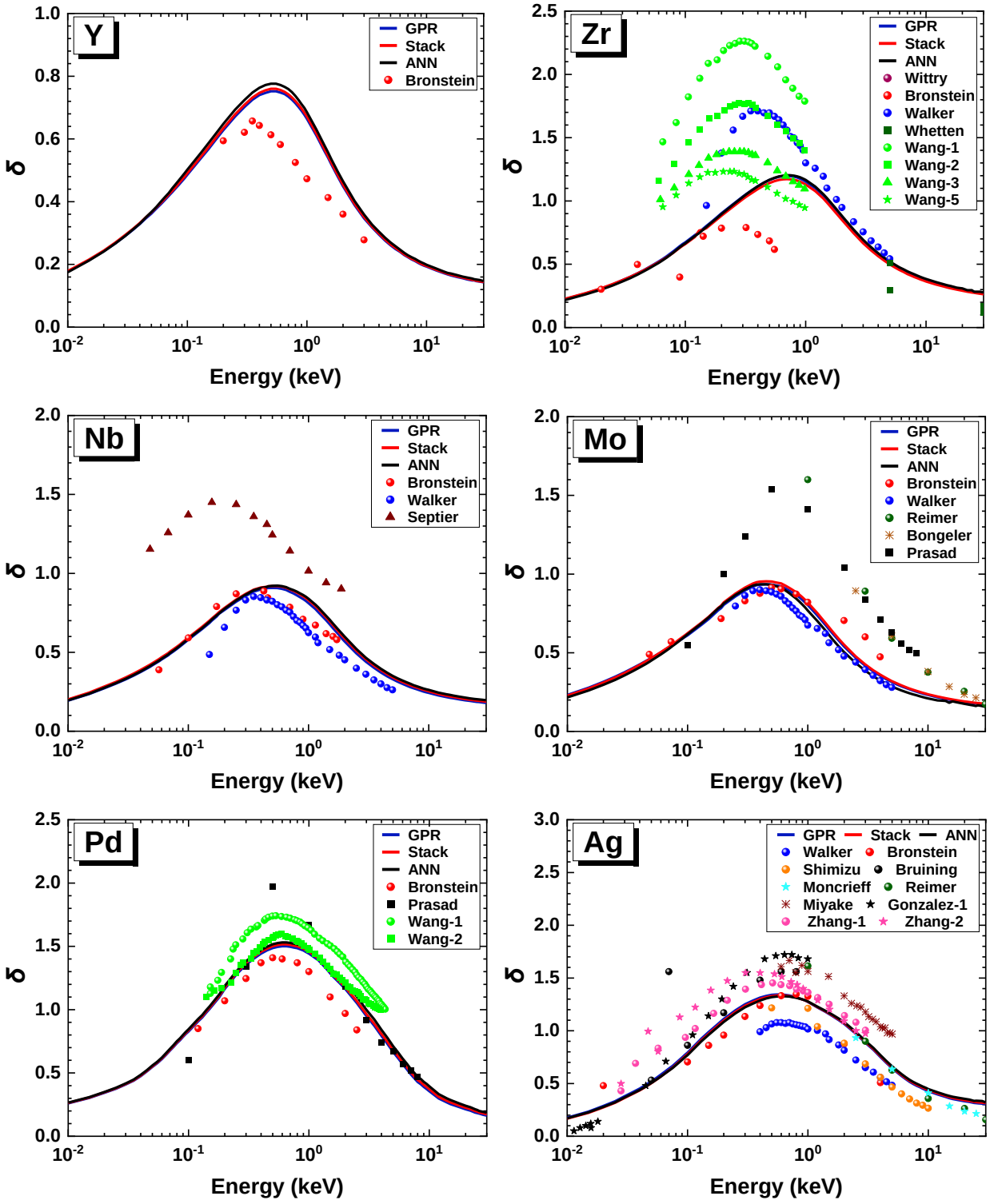
Figure S2 represents the predicted yields by the three models for all the 40 elements in the input database on a log scale for energy. We have presented the SEY data for the same element by various techniques with numeric identifiers. Sihui [18, 19] have measured SEY for transition metals, Ti, V, Zr and Hf, for as-received samples and after electron conditioning with various electron doses. In Fig. 6 Wang-1, Wang-2, Wang-3, Wang-4, Wang-5 and Wang-6 represent the measured yield for as-received samples (Ti, V, Zr and Hf) and electron conditioned samples with electron doses of 3.8×10^{-3}, 4.5×10^{-4}, 1.7×10^{-4}, 1.1×10^{-4} and 2×10^{-4} C/mm^2, respectively. For Cu and Au, Gonzalez-1 and Gonzalez-2 represent respectively the SEY measured by Gonzalez et al. [23] for Ar^+ cleaned and thermal treated samples. Similarly, for Pd, Wang-1 and Wang-2 represent SEY measured by Wang et al. [20] for electron conditioned samples with electron doses of 1×10^{-8} and 1×10^{-7} C/mm^2, respectively. For Ag, Zhang-1 and Zhang-2 represents the SEY measured by Zhang et al. [15] for Ar^+ cleaned and heat treated samples, respectively.

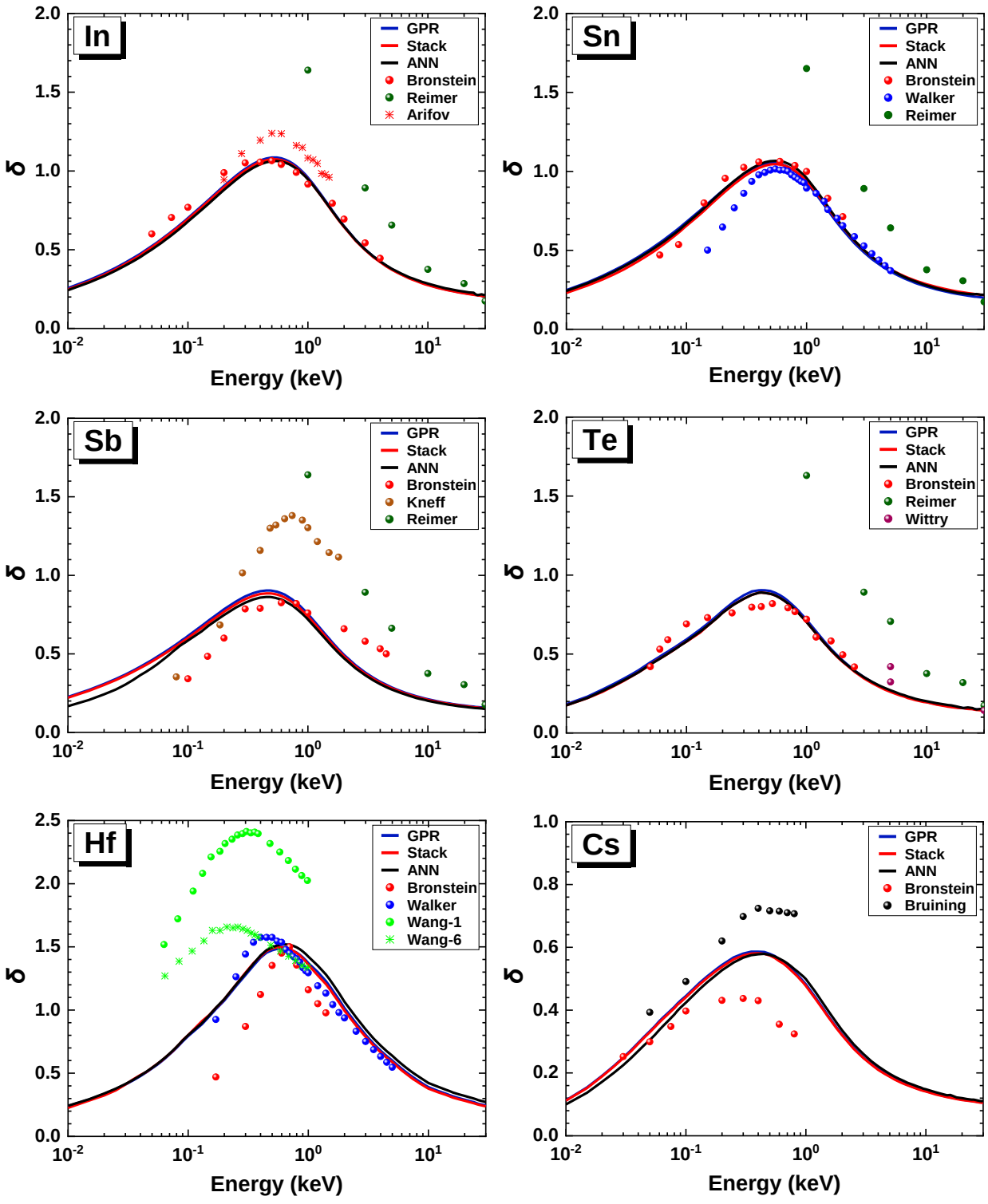


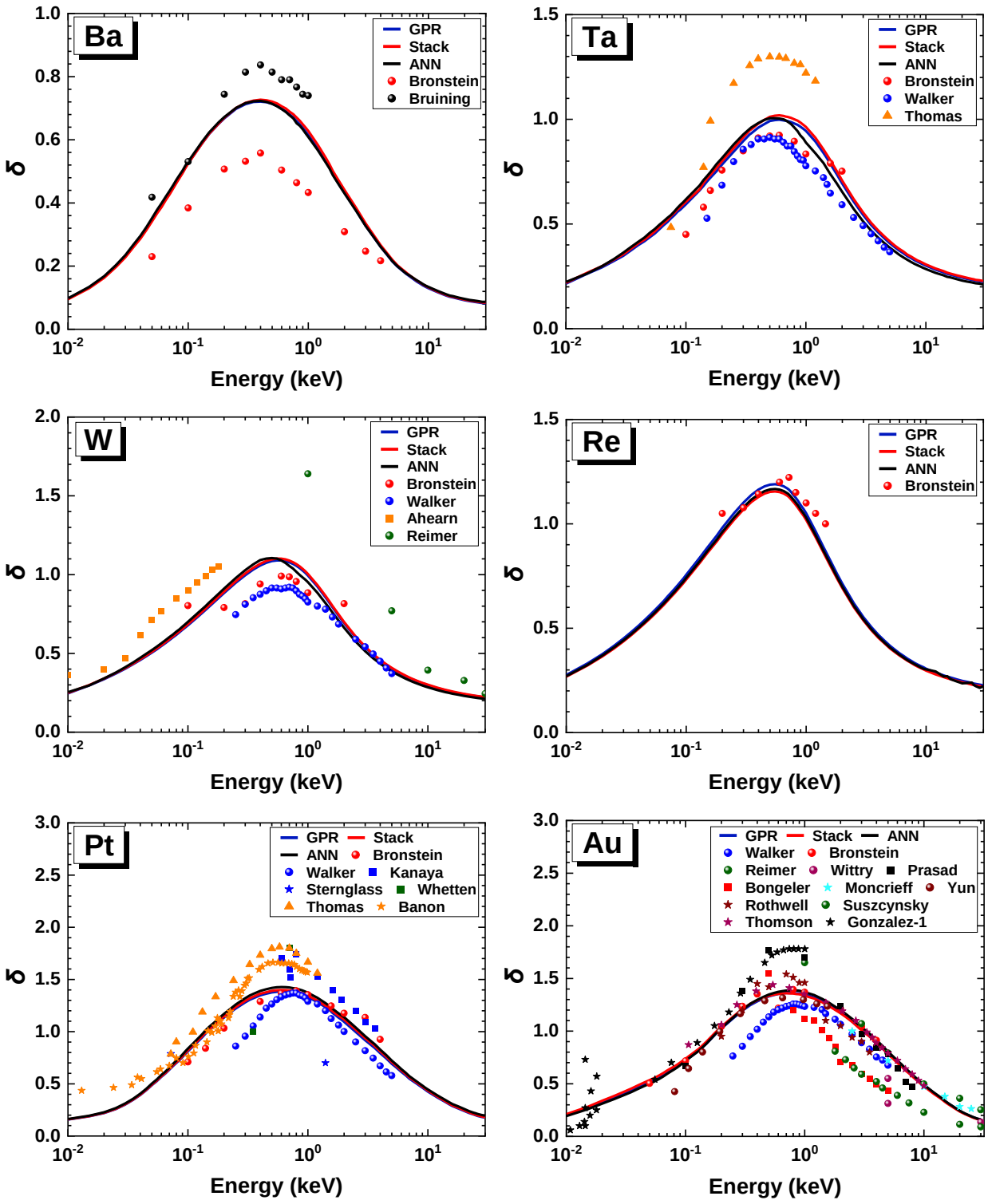

B
GPR
Stack
ANN
Bronstein
Whetten
C
GPR
Stack
ANN
Bronstein
Walker
Moncrieff
Bongeler
Wittry
Rothwell
Kirby
Chen
Farhang
Mg
GPR
Stack
ANN
Bronstein
Bruining
Al
GPR
Stack
ANN
Bronstein
Walker
Czaja
Shimizu
Kanter
Reimer
Moncrieff
Bruining
Si
GPR
Stack
ANN
Bronstein
Walker
Czaja
Rothwell
Dionne
K
GPR
Stack
ANN
Bronstein
δ
Energy (keV)


Ca
GPR
Stack
ANN
Bronstein
δ
Energy (keV)
Sc
GPR
Stack
ANN
Bronstein
δ
Energy (keV)
Ti
GPR
Stack
ANN
Bronstein
Walker
Baglin
Shibuya
Wittry
Thomson
Joy
Wang-1
Wang-2
Wang-3
Wang-4
δ
Energy (keV)
V
GPR
Stack
ANN
Wittry
Wang-1
Wang-2
δ
Energy (keV)
Cr
GPR
Stack
ANN
Walker
Wittry
Prasad
δ
Energy (keV)
Fe
GPR
Stack
ANN
Bronstein
Wittry
Walker
δ
Energy (keV)


Cu
GPR
Stack
ANN
Ahmed
Bronstein
Moncrieff
Hu
Walker
Koshikawa
Bojko
Bruining
Septier
Bongeler
Shimizu
Cimino
Wang-1
Gonzalez-1
Gonzalez-2
δ
Energy (keV)
Zn
GPR
Stack
ANN
Bronstein
Walker
δ
Energy (keV)
Ga
GPR
Stack
ANN
Bronstein
δ
Energy (keV)
Ge
GPR
Stack
ANN
Bronstein
Reimer
Johnson
Whetten
δ
Energy (keV)
Se
GPR
Stack
ANN
Bronstein
δ
Energy (keV)
Sr
GPR
Stack
ANN
Bronstein
δ
Energy (keV)


Y
GPR
Stack
ANN
Bronstein
δ
Energy (keV)
Zr
GPR
Stack
ANN
Wittry
Bronstein
Walker
Whetten
Wang-1
Wang-2
Wang-3
Wang-5
δ
Energy (keV)
Nb
GPR
Stack
ANN
Bronstein
Walker
Septier
δ
Energy (keV)
Mo
GPR
Stack
ANN
Bronstein
Walker
Reimer
Bongeler
Prasad
δ
Energy (keV)
Pd
GPR
Stack
ANN
Bronstein
Prasad
Wang-1
Wang-2
δ
Energy (keV)
Ag
GPR
Stack
ANN
Walker
Bronstein
Shimizu
Bruining
Moncrieff
Reimer
Miyake
Gonzalez-1
Zhang-1
Zhang-2
δ
Energy (keV)


In
GPR
Stack
ANN
Bronstein
Reimer
Arifov
Sn
GPR
Stack
ANN
Bronstein
Walker
Reimer
Sb
GPR
Stack
ANN
Bronstein
Kneff
Reimer
Te
GPR
Stack
ANN
Bronstein
Reimer
Wittry
Hf
GPR
Stack
ANN
Bronstein
Walker
Wang-1
Wang-6
Cs
GPR
Stack
ANN
Bronstein
Bruining
δ
Energy (keV)


Ba
GPR
Stack
ANN
Bronstein
Bruining
Ta
GPR
Stack
ANN
Bronstein
Walker
Thomas
W
GPR
Stack
ANN
Bronstein
Walker
Ahearn
Reimer
Re
GPR
Stack
ANN
Bronstein
Pt
GPR
Stack
ANN
Bronstein
Walker
Kanaya
Sternglass
Whetten
Thomas
Banon
Au
GPR
Stack
ANN
Walker
Bronstein
Reimer
Wittry
Prasad
Bongeler
Moncrieff
Yun
Rothwell
Suszcynsky
Thomson
Gonzalez-1
S
Energy (keV)

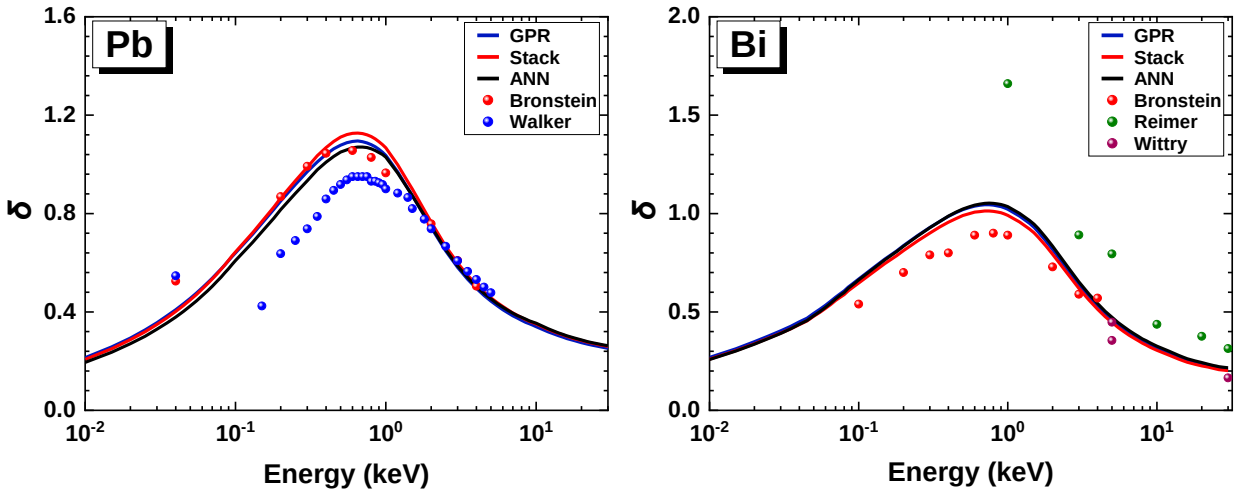


Figure S2: ML predicted yields (solid lines) for all the 40 elements included in the training dataset. Experimental data measured by different research groups are represented by marks.

It may be noted that we have not included the experimental data of compounds as well as the data of as-received samples if they are unusually high. It is quite noticeable that the ML predicted yields by the three models for all the elements are in very good agreement with an average RMSE deviation of 1.47%. It can be observed that the measured data among different research groups for an element usually have closer agreement at high energies above 1 keV than at lower energies and, consequently, ML predicted yields show also better agreement with the measured data at high energies. Wittry [104], Reimer [79], Thomson [8] and Whetten [103] have measured SEY at high energies; a good agreement has been found between their measured data and the ML predicted yield. At low energies, no such a good agreement can be found even for the measurements performed with very good experimental setups. As an example, for Sn, the ML predicted yield show a deviation of 17.56% (RMSE) from Walker's measurement [55]. At energies below 500 eV, the ML predicted yields have a large variation (~30%) with the Walker's data. However, at energies above 500 eV the RMSE deviation is only 3%.

For Li, B and V the ML predicted yields are in good agreement with the available experimental data at high energies; however, at low energies below the maximum energy the measured yields [18, 20] are usually higher than the predicted value. For Si, ML predicted yields show a good agreement with the available data except the slightly lower than the data of Dione [30] and Walker [12] around the peak energy. For K and Ca, the ML predicted yields are comparatively higher than

the only available experimental data of Bronstein [93]. For Ti, the ML prediction just lie in a very wide range of experimental data.

For C, the ML predicted yields are in good agreement with the measured yield of Bronstein [93], Walker [12] and Chen [107] as well as with that of Moncrieff [80], Bongeler [81] and Wittry [104] at higher energies. For Al, the ML predicted yields are also in good agreement with the measured data of Walker [12]. It may be noted that the ML predicted δ_{m} for Al is well below that of as-received sample (> 3) and is in the range for pure Al (~1) [97]. For Ag, the ML predictions are in close agreement with the measured data except those of Miyake [11] and Hai [15] which are exceptionally higher than the other measured data. For Pd, In, Sn, W, Pt and Au good agreement can also be found between the measured data and the ML predictions.

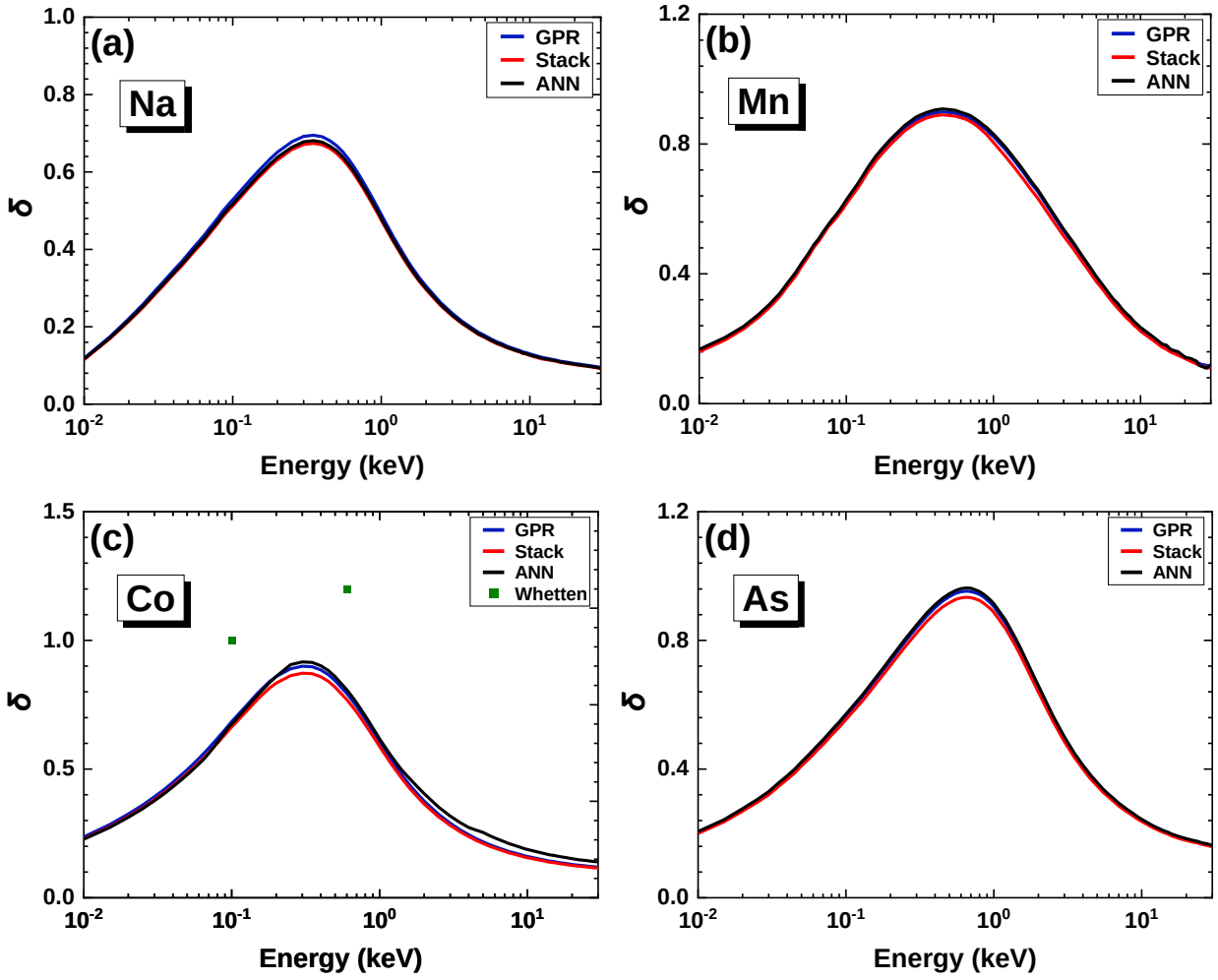
Table S4: RMSE (%) between ML models in the predicted yield

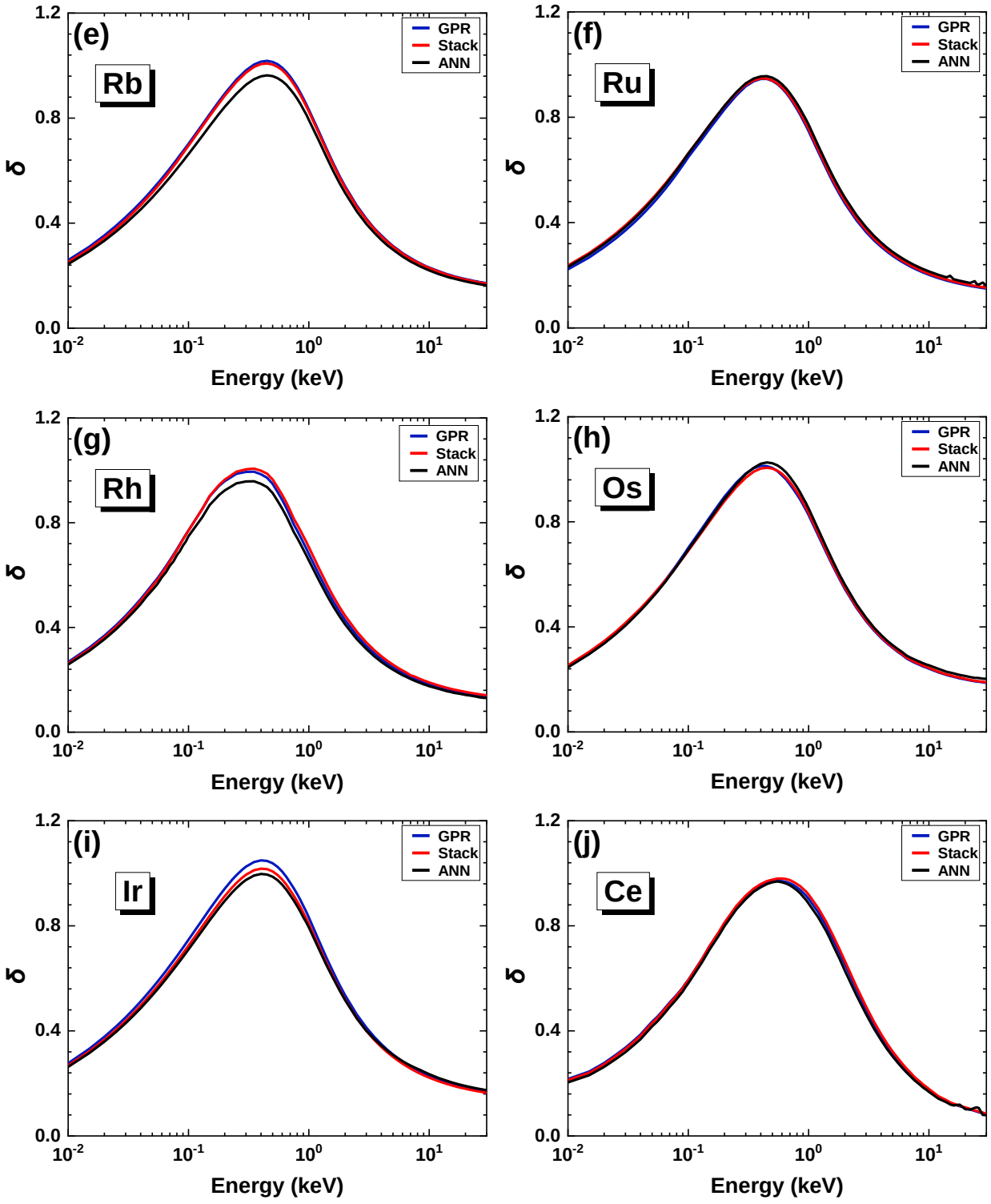
Element	RMSE (%)			Element	RMSE (%)		
	GPR-Stack	GPR-ANN	Stack-ANN		GPR-Stack	GPR-ANN	Stack-ANN
Li	0.81	0.73	1.22	Ru	1.10	1.54	0.91
Be	2.43	1.18	1.44	Rh	1.24	2.11	3.04
B	1.10	0.90	1.93	Pd	1.11	2.74	1.67
C	0.83	1.63	2.45	Ag	1.15	1.87	0.87
Na	0.41	1.22	0.81	Cd	0.63	1.90	1.32
Mg	0.60	1.04	1.28	In	0.66	1.45	0.97
Al	1.25	3.42	2.72	Sn	1.67	1.28	1.27
Si	3.51	2.59	3.32	Sb	1.06	3.12	2.22
K	1.16	1.56	1.41	Te	1.06	1.03	0.48
Ca	0.35	0.56	0.53	Cs	0.37	1.43	1.48
Sc	0.80	0.73	1.46	Ba	0.84	0.58	0.42
Ti	1.32	1.34	2.25	La	0.49	0.75	0.90
V	2.44	0.91	1.54	Ce	1.83	0.77	1.32
Cr	1.09	1.46	0.76	Nd	1.39	0.44	1.17
Mn	0.55	0.92	1.42	Sm	2.24	0.73	2.94
Fe	0.92	0.92	0.66	Eu	1.00	0.90	1.72
Co	1.56	2.12	2.94	Lu	1.67	0.64	1.15
Ni	2.45	4.08	1.63	Hf	1.10	2.73	3.51
Cu	1.06	2.35	1.44	Ta	2.26	2.69	1.17
Zn	3.35	1.70	1.82	W	0.64	2.04	2.18
Ga	0.92	2.02	1.24	Re	2.00	1.34	0.72
Ge	0.55	1.89	1.36	Os	0.61	1.27	1.03
As	1.16	0.58	1.74	Ir	1.92	3.03	1.32
Se	0.51	1.09	0.85	Pt	1.16	2.93	1.81
Rb	0.70	3.20	2.54	Au	1.45	1.27	2.41
Sr	0.78	0.59	1.23	Hg	0.67	2.98	2.64
Y	0.41	1.25	0.86	Tl	1.13	1.77	0.96
Zr	1.35	0.91	1.91	Pb	1.57	1.94	2.77
Nb	1.24	1.63	0.75	Bi	1.80	0.93	2.32
Mo	1.00	1.85	2.46	Th	0.71	1.02	1.64
Average	1.22	1.59	1.60				

Table S5: ML predicted yield for unseen elements (Excel file)

Figure S3:

Figure S3 represents the ML predicted SEY for 14 elements including Na, Mn, Co, As, Rb, Ru, Rh, Os, Ir, Ce, Nd, Sm, Eu and Lu. For these elements no experimental data exist. This confirms the robustness of the ML models used in this study.




(e)
Rb
(f)
Ru
(g)
Rh
(h)
Os
(i)
Ir
(j)
Ce
GPR
Stack
ANN
δ
Energy (keV)
1.2
0.8
0.4
0.0
10^-2
10^-1
10^0
10^1

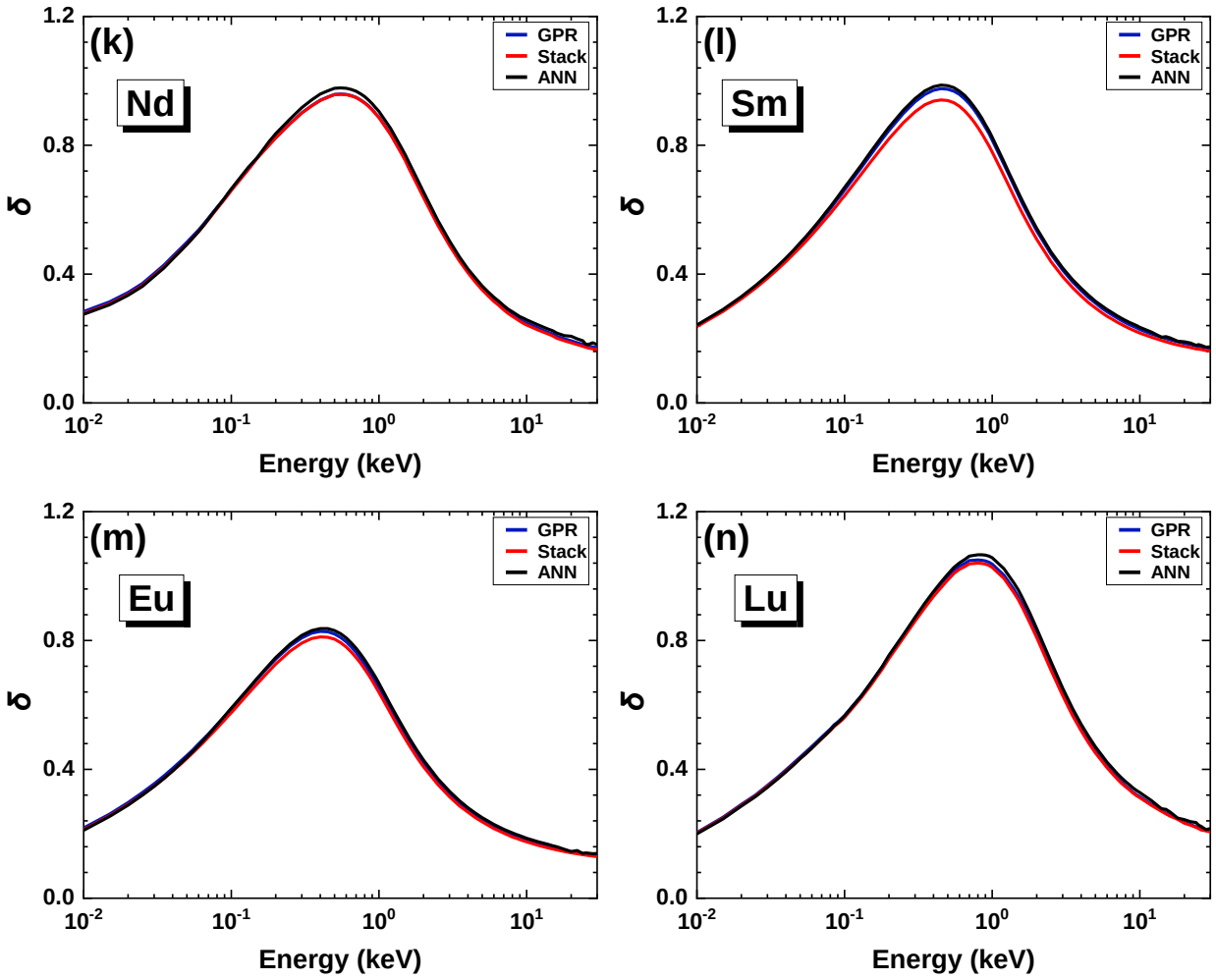


Figure S3: ML predicted yields (solid curve) for 14 unseen elements used to check the generalization ability of the model.