

Twelve different types of data normalization for the proposition of classification, univariate and multivariate regression models for the direct analyses of alloys by Laser-induced breakdown spectroscopy (LIBS)

Jeyne Pricylla Castro and Edenir Rodrigues Pereira-Filho*

Group of Applied Instrumental Analysis, Chemistry Department, Federal University of São Carlos, São Carlos – São Paulo State, Brazil, P. O. Box 676, Zip Code 13565-905

*corresponding author: erpf@ufscar.br

Phone: + 55 16 3351 8092
Fax: + 55 16 3351 8350

Supplementary Material

Table 1S

Pictorial description of the twelve normalization modes used in this study

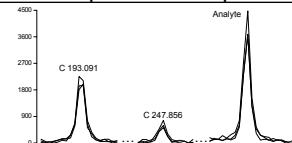
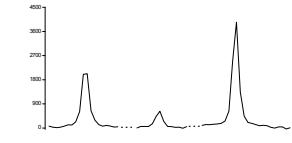
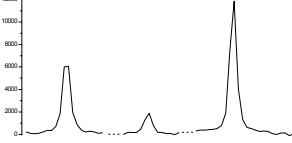
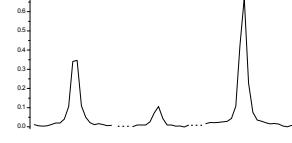
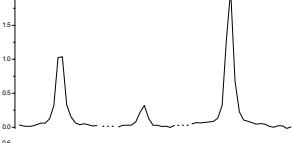
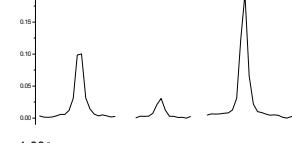
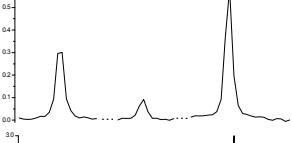
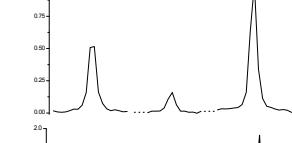
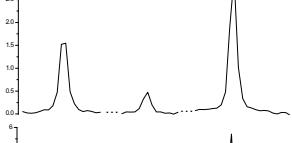
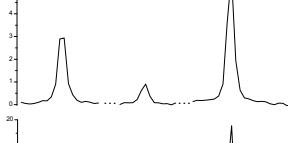
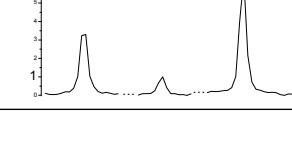
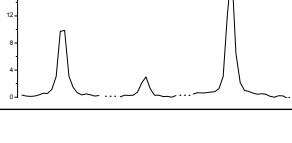
Identification	Normalization mode	Graphical description	Identification	Normalization mode	Graphical description
None	Original data composed of carbon lines and a hypothetical analyte		-	-	-
1	Average		7	Sum	
2	Normalized by the individual norm and averaged over n pulses (the final norm of each spectrum is 1)		8	Normalized by the individual norm and sum over n pulses	
3	Normalized by the individual area and averaged over n pulses		9	Normalized by the individual area and sum over n pulses	
4	Normalized by the individual maximum and averaged over n pulses (the signal of the highest emission line is 1)		10	Normalized by the individual maximum and sum over n pulses	
5	Normalized by the individual C I 193.091 nm intensity and averaged over n pulses (the signal of C I 193.091 is 1)		11	Normalized by the individual C I 193.091 nm intensity and sum over n pulses	
6	Normalized by the individual C I 247.856 nm intensity and averaged over n pulses (the signal of C I 247.856 is 1)		12	Normalized by the individual C I 247.856 nm intensity and sum over n pulses	

Table 2S: Anova tables for models calculated using signal area

Sample name and emission line	Parameter	Sum of Square (SS)	Degree of Freedom (DF)	Mean of Square (MS)	F test
Aluminum alloy (Al I 394.400 nm)	Regression (R)	4.983x10 ⁸	3	1.661x10 ⁸	Test F1: Test F1: MSR/MSres = 32.61
	Residue (res)	5.602x10 ⁷	11	5.093x10 ⁶	F tab (95%, 3, 11) = 3.587
	Total (T)	5.543x10 ⁸	14	3.959x10 ⁷	
	Pure error (PE)	6.160x10 ⁵	2	3.080x10 ⁵	Test F2: MSLoF/MSPE = 19.99
	Lack of fit (LoF)	5.541x10 ⁷	9	6.157x10 ⁶	F tab (95%, 9, 2) = 19.38
	R ²	0.8989	R	0.9481	
	Maximum R ²	0.9989	Maximum R	0.9994	
Co-Cr-Mo alloy (Cr II 283.563 nm)	Regression (R)	1.203x10 ⁹	5	2.406x10 ⁸	Test F1: MSR/MSres = 72.67
	Residue (res)	2.979x10 ⁷	9	3.311x10 ⁶	F tab (95%, 5, 9) = 3.481
	Total (T)	1.233x10 ⁹	14	8.806x10 ⁷	
	Pure error (PE)	1.140x10 ⁶	2	5.700x10 ⁵	Test F2: MSLoF/MSPE = 7.183
	Lack of fit (LoF)	2.866x10 ⁷	7	4.094x10 ⁶	F tab (95%, 7, 2) = 19.35
	R ²	0.9758	R	0.9878	
	Maximum R ²	0.9991	Maximum R	0.9995	
Unidentified alloy (Cu I 324.754 nm)	Regression (R)	1.464x10 ⁷	4	3.659x10 ⁶	Test F1: MSR/MSres = 77.60
	Residue (res)	4.716x10 ⁵	10	4.716x10 ⁴	F tab (95%, 4, 10) = 3.478
	Total (T)	1.511x10 ⁷	14	1.079x10 ⁶	
	Pure error (PE)	1.800x10 ⁴	2	9.000x10 ³	Test F2: MSLoF/MSPE = 6.300
	Lack of fit (LoF)	4.536x10 ⁵	8	5.670x10 ⁴	F tab (95%, 8, 2) = 19.37
	R ²	0.9688	R	0.9843	
	Maximum R ²	0.9988	Maximum R	0.9994	
Steel (Fe II 259.940 nm)	Regression (R)	4.184x10 ¹⁰	4	1.046x10 ¹⁰	Test F1: MSR/MSres = 221.5
	Residue (res)	4.722x10 ⁸	10	4.722x10 ⁷	F tab (95%, 4, 10) = 3.478
	Total (T)	4.232x10 ¹⁰	14	3.023x10 ⁹	
	Pure error (PE)	1.510x10 ⁷	2	7.550x10 ⁶	Test F2: MSLoF/MSPE = 7.568
	Lack of fit (LoF)	4.571x10 ⁸	8	5.714x10 ⁷	F tab (95%, 8, 2) = 19.37
	R ²	0.9888	R	0.9944	
	Maximum R ²	0.9996	Maximum R	0.9998	
Stainless steel (Fe II 259.940 nm)	Regression (R)	2.792x10 ⁹	3	9.307x10 ⁸	Test F1: MSR/MSres = 19.18
	Residue (res)	5.338x10 ⁸	11	4.853x10 ⁷	F tab (95%, 3, 11) = 3.587
	Total (T)	3.326x10 ⁹	14	2.374x10 ⁸	
	Pure error (PE)	1.710x10 ⁷	2	8.550x10 ⁶	Test F2: MSLoF/MSPE = 6.715
	Lack of fit (LoF)	5.167x10 ⁸	9	5.741x10 ⁷	F tab (95%, 9, 2) = 19.38
	R ²	0.8395	R	0.9162	
	Maximum R ²	0.9949	Maximum R	0.9974	

Table 2S: Continuation

Sample name and emission line	Parameter	Sum of Square (SS)	Degree of Freedom (DF)	Mean of Square (MS)	F test
Stainless steel (Fe II 259.940 nm)	Regression (R)	2.545x10 ⁹	5	5.091x10 ⁸	Test F1: MSR/MSres = 45.51
	Residue (res)	1.007x10 ⁸	9	1.118x10 ⁷	F tab (95%, 5, 9) = 3.481
	Total (T)	2.646x10 ⁹	14	1.890x10 ⁸	
	Pure error (PE)	8.290x10 ⁶	2	4.145x10 ⁶	Test F2: MSLoF/MSPE = 3.184
	Lack of fit (LoF)	9.240x10 ⁷	7	1.320x10 ⁷	F tab (95%, 7, 2) = 19.35
	R ²	0.9619	R	0.9808	
	Maximum R ²	0.9969	Maximum R	0.9984	
Stainless steel (Mn II 294.920 nm)	Regression (R)	3.989x10 ⁷	4	9.973x10 ⁶	Test F1: MSR/MSres = 105.4
	Residue (res)	9.459x10 ⁵	10	9.459x10 ⁴	F tab (95%, 4, 10) = 3.478
	Total (T)	4.084x10 ⁷	14	2.917x10 ⁶	
	Pure error (PE)	1.470x10 ⁴	2	7.350x10 ³	Test F2: MSLoF/MSPE = 15.84
	Lack of fit (LoF)	9.312x10 ⁵	8	1.164x10 ⁵	F tab (95%, 8, 2) = 19.37
	R ²	0.9768	R	0.9884	
	Maximum R ²	0.9996	Maximum R	0.9998	
Hastelloy (Mo I 386.411 nm)	Regression (R)	1.057x10 ⁶	1	1.057x10 ⁶	Test F1: MSR/MSres = 25.25
	Residue (res)	5.444x10 ⁵	13	4.187x10 ⁴	F tab (95%, 1, 13) = 4.667
	Total (T)	1.602x10 ⁶	14	1.144x10 ⁵	
	Pure error (PE)	1.300x10 ⁵	8	1.625x10 ⁴	Test F2: MSLoF/MSPE = 5.100
	Lack of fit (LoF)	4.144x10 ⁵	5	8.288x10 ⁴	F tab (95%, 5, 8) = 3.687
	R ²	0.6602	R	0.8125	
	Maximum R ²	0.9188	Maximum R	0.9586	
Hastelloy (Ni I 361.939 nm)	Regression (R)	2.295x10 ⁶	1	2.295x10 ⁶	Test F1: MSR/MSres = 20.08
	Residue (res)	1.485x10 ⁶	13	1.143x10 ⁵	F tab (95%, 1, 13) = 4.667
	Total (T)	3.781x10 ⁶	14	2.701x10 ⁵	
	Pure error (PE)	3.759x10 ⁵	8	4.699x10 ⁴	Test F2: MSLoF/MSPE = 4.721
	Lack of fit (LoF)	1.109x10 ⁶	5	2.219x10 ⁵	F tab (95%, 5, 8) = 3.687
	R ²	0.6071	R	0.7792	
	Maximum R ²	0.9006	Maximum R	0.9490	
6Al 4V titanium alloy (V I 437.924 nm)	Regression (R)	3.348x10 ³	1	3.348x10 ³	Test F1: MSR/MSres = 5.472
	Residue (res)	7.954x10 ³	13	6.118x10 ²	F tab (95%, 1, 13) = 4.667
	Total (T)	1.130x10 ⁴	14	8.073x10 ²	
	Pure error (PE)	4.870x10 ³	8	6.087x10 ²	Test F2: MSLoF/MSPE = 1.013
	Lack of fit (LoF)	3.084x10 ³	5	6.168x10 ²	F tab (95%, 5, 8) = 3.687
	R ²	0.2963	R	0.5443	
	Maximum R ²	0.5691	Maximum R	0.7544	
Steel - two layers of Cr (Zn I 481.053 nm)	Regression (R)	1.209x10 ¹¹	3	4.031x10 ¹⁰	Test F1: MSR/MSres = 73.78
	Residue (res)	6.009x10 ⁹	11	5.463x10 ⁸	F tab (95%, 3, 11) = 3.587
	Total (T)	1.269x10 ¹¹	14	9.067x10 ⁹	
	Pure error (PE)	4.250x10 ⁸	2	2.125x10 ⁸	Test F2: MSLoF/MSPE = 2.919
	Lack of fit (LoF)	5.584x10 ⁹	9	6.204x10 ⁸	F tab (95%, 9, 2) = 19.38
	R ²	0.9527	R	0.9760	
	Maximum R ²	0.9967	Maximum R	0.9983	

Table 3S: Anova tables for models calculated using signal height

Sample name and emission line	Parameter	Sum of Square (SS)	Degree of Freedom (DF)	Mean of Square (MS)	F test
Aluminum alloy (Al I 394.400 nm)	Regression (R)	4.486x10 ⁷	2	2.243x10 ⁷	Test F1: MSR/MSres = 21.97
	Residue (res)	1.225x10 ⁷	12	1.021x10 ⁶	F tab (95%, 2, 12) = 3.885
	Total (T)	5.711x10 ⁷	14	4.079x10 ⁶	
	Pure error (PE)	9.530x10 ³	2	4.765x10 ³	Test F2: MSLoF/MSPE = 256.9
	Lack of fit (LoF)	1.224x10 ⁷	10	1.224x10 ⁶	F tab (95%, 10, 2) = 19.39
	R ²	0.7855	R	0.8863	
	Maximum R ²	0.9998	Maximum R	0.9999	
Co-Cr-Mo alloy (Cr II 283.563 nm)	Regression (R)	2.363x10 ⁸	4	5.909x10 ⁷	Test F1: MSR/MSres = 42.50
	Residue (res)	1.390x10 ⁷	10	1.390x10 ⁶	F tab (95%, 4, 10) = 3.478
	Total (T)	2.502x10 ⁸	14	1.787x10 ⁷	
	Pure error (PE)	2.850x10 ⁵	2	1.425x10 ⁵	Test F2: MSLoF/MSPE = 11.94
	Lack of fit (LoF)	1.362x10 ⁷	8	1.702x10 ⁶	F tab (95%, 8, 2) = 19.37
	R ²	0.9444	R	0.9718	
	Maximum R ²	0.9989	Maximum R	0.9994	
Unidentified alloy (Cu I 324.754 nm)	Regression (R)	1.941x10 ⁶	3	6.469x10 ⁵	Test F1: MSR/MSres = 83.48
	Residue (res)	8.525x10 ⁴	11	7.750x10 ³	F tab (95%, 3, 11) = 3.587
	Total (T)	2.026x10 ⁶	14	1.447x10 ⁵	
	Pure error (PE)	3.800x10 ³	2	1.900x10 ³	Test F2: MSLoF/MSPE = 4.763
	Lack of fit (LoF)	8.145x10 ⁴	9	9.050x10 ³	F tab (95%, 9, 2) = 19.38
	R ²	0.9579	R	0.9787	
	Maximum R ²	0.9981	Maximum R	0.9991	
Steel (Fe II 259.940 nm)	Regression (R)	3.235x10 ⁹	4	8.087x10 ⁸	Test F1: MSR/MSres = 314.9
	Residue (res)	2.568x10 ⁷	10	2.568x10 ⁶	F tab (95%, 4, 10) = 3.478
	Total (T)	3.261x10 ⁹	14	2.329x10 ⁸	
	Pure error (PE)	6.190x10 ⁵	2	3.095x10 ⁵	Test F2: MSLoF/MSPE = 10.12
	Lack of fit (LoF)	2.506x10 ⁷	8	3.132x10 ⁶	F tab (95%, 8, 2) = 19.37
	R ²	0.9921	R	0.9961	
	Maximum R ²	0.9998	Maximum R	0.9999	
Stainless steel (Fe II 259.940 nm)	Regression (R)	2.809x10 ⁸	3	9.363x10 ⁷	Test F1: MSR/MSres = 20.06
	Residue (res)	5.134x10 ⁷	11	4.667x10 ⁶	F tab (95%, 3, 11) = 3.587
	Total (T)	3.322x10 ⁸	14	2.373x10 ⁷	
	Pure error (PE)	1.790x10 ⁶	2	8.950x10 ⁵	Test F2: MSLoF/MSPE = 6.151
	Lack of fit (LoF)	4.955x10 ⁷	9	5.505x10 ⁶	F tab (95%, 9, 2) = 19.38
	R ²	0.8455	R	0.9195	
	Maximum R ²	0.9946	Maximum R	0.9973	

Table 3S: Continuation

Sample name and emission line	Parameter	Sum of Square (SS)	Degree of Freedom (DF)	Mean of Square (MS)	F test
Stainless steel (Fe II 259.940 nm)	Regression (R)	2.907x10 ⁸	6	4.844x10 ⁷	Test F1: MSR/MSres = 63.64
	Residue (res)	6.089x10 ⁶	8	7.612x10 ⁵	F tab (95%, 6, 8) = 3.580
	Total (T)	2.967x10 ⁸	14	2.119x10 ⁷	
	Pure error (PE)	1.030x10 ⁶	2	5.150x10 ⁵	Test F2: MSLoF/MSPE = 1.637
	Lack of fit (LoF)	5.059x10 ⁶	6	8.432x10 ⁵	F tab (95%, 6, 2) = 19.33
	R ²	0.9795	R	0.9897	
	Maximum R ²	0.9965	Maximum R	0.9983	
Stainless steel (Mn II 294.920 nm)	Regression (R)	1.153x10 ⁷	4	2.884x10 ⁶	Test F1: MSR/MSres = 102.7
	Residue (res)	2.808x10 ⁵	10	2.808x10 ⁴	F tab (95%, 4, 10) = 3.478
	Total (T)	1.181x10 ⁷	14	8.439x10 ⁵	
	Pure error (PE)	5.450x10 ³	2	2.725x10 ³	Test F2: MSLoF/MSPE = 12.63
	Lack of fit (LoF)	2.754x10 ⁵	8	3.442x10 ⁴	F tab (95%, 8, 2) = 19.37
	R ²	0.9762	R	0.9880	
	Maximum R ²	0.9995	Maximum R	0.9998	
Hastelloy (Mo I 386.411 nm)	Regression (R)	2.003x10 ⁵	1	2.003x10 ⁵	Test F1: MSR/MSres = 25.86
	Residue (res)	1.006x10 ⁵	13	7.744x10 ³	F tab (95%, 1, 13) = 4.667
	Total (T)	3.010x10 ⁵	14	2.150x10 ⁴	
	Pure error (PE)	3.080x10 ⁴	8	3.850x10 ³	Test F2: MSLoF/MSPE = 3.629
	Lack of fit (LoF)	6.987x10 ⁴	5	1.397x10 ⁴	F tab (95%, 5, 8) = 3.687
	R ²	0.6655	R	0.8158	
	Maximum R ²	0.8977	Maximum R	0.9475	
Hastelloy (Ni I 361.939 nm)	Regression (R)	3.977x10 ⁵	1	3.977x10 ⁵	Test F1: MSR/MSres = 18.05
	Residue (res)	2.864x10 ⁵	13	2.203x10 ⁴	F tab (95%, 1, 13) = 4.667
	Total (T)	6.841x10 ⁵	14	4.886x10 ⁴	
	Pure error (PE)	1.939x10 ⁵	8	2.424x10 ⁴	Test F2: MSLoF/MSPE = 0.7630
	Lack of fit (LoF)	9.248x10 ⁴	5	1.849x10 ⁴	F tab (95%, 5, 8) = 3.687
	R ²	0.5813	R	0.7624	
	Maximum R ²	0.7165	Maximum R	0.8465	
6Al 4V titanium alloy (V I 437.924 nm)	Regression (R)	8.601x10 ³	1	8.601x10 ³	Test F1: MSR/MSres = 9.621
	Residue (res)	1.162x10 ⁴	13	8.939x10 ²	F tab (95%, 1, 13) = 4.667
	Total (T)	2.022x10 ⁴	14	1.444x10 ³	
	Pure error (PE)	5.727x10 ³	10	5.727x10 ²	Test F2: MSLoF/MSPE = 3.431
	Lack of fit (LoF)	5.894x10 ³	3	1.965x10 ³	F tab (95%, 3, 10) = 3.708
	R ²	0.4253	R	0.6522	
	Maximum R ²	0.7168	Maximum R	0.8466	
Steel - two layers of Cr (Zn I 481.053 nm)	Regression (R)	4.129x10 ⁹	3	1.376x10 ⁹	Test F1: MSR/MSres = 63.93
	Residue (res)	2.368x10 ⁸	11	2.153x10 ⁷	F tab (95%, 3, 11) = 3.587
	Total (T)	4.366x10 ⁹	14	3.118x10 ⁸	
	Pure error (PE)	2.740x10 ⁷	2	1.370x10 ⁷	Test F2: MSLoF/MSPE = 1.698
	Lack of fit (LoF)	2.094x10 ⁸	9	2.327x10 ⁷	F tab (95%, 9, 2) = 19.38
	R ²	0.9458	R	0.9725	
	Maximum R ²	0.9937	Maximum R	0.9969	

Table 4S

Anova table parameters (Tables 2S and 3S) identification and equations for the models calculated at Table 3.

Parameters	Sum of Squares (SS)	Degree of freedom (D.F.)	Remarks
Regression, SSR	$\sum_{i=1}^n (\hat{y}_i - \bar{y})^2$	p-1	
residue, SSres	$\sum_{i=1}^n (y_i - \hat{y}_i)^2$	n-p	
Total, SST	$\sum_{i=1}^n (y_i - \bar{y})^2$	n-1	m = number of independent experiments
Pure Error, SSPE	$\sum_{i=1}^m \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2$	n-m	\hat{y}_i : predicted response (area or height) \bar{y} : average of experimental responses (area or height)
Lack of Fit, SSLoF	$\sum_{i=1}^m \sum_{j=1}^{n_i} (\hat{y}_i - \bar{y}_i)^2$	m-p	
R ²	$\frac{SSR}{SST}$		
Maximum R ²	$\frac{SST - SSPE}{SST}$		

Fig. 1S

Pictorial description used for standard deviation calculation of the signal background.

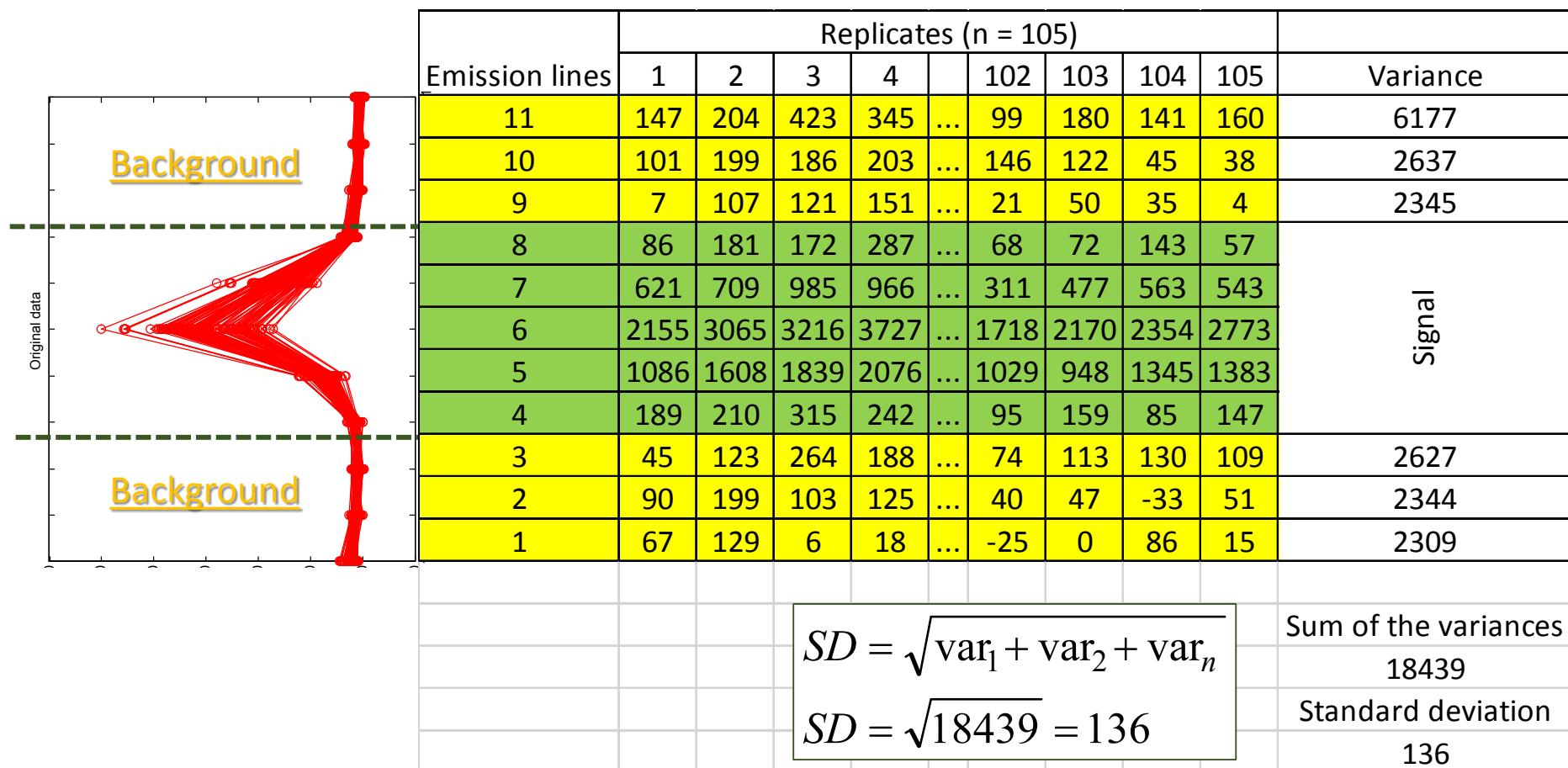


Fig. 2Sa Emission spectra for some samples and selected emission lines: steel with C and Fe lines (a), stainless steel with C and Fe lines (b), stainless steel with Mn lines (c), steel – two layers of Cr with Zn lines (d), aluminum alloy with Al lines (e), unidentified sample with Cu lines (f), Co-Cr-Mo alloy sample with Co and Cr lines (g), 6Al 4V titanium alloy with Ti and V lines (h), and hastelloy sample with Ni and Mo lines (i).

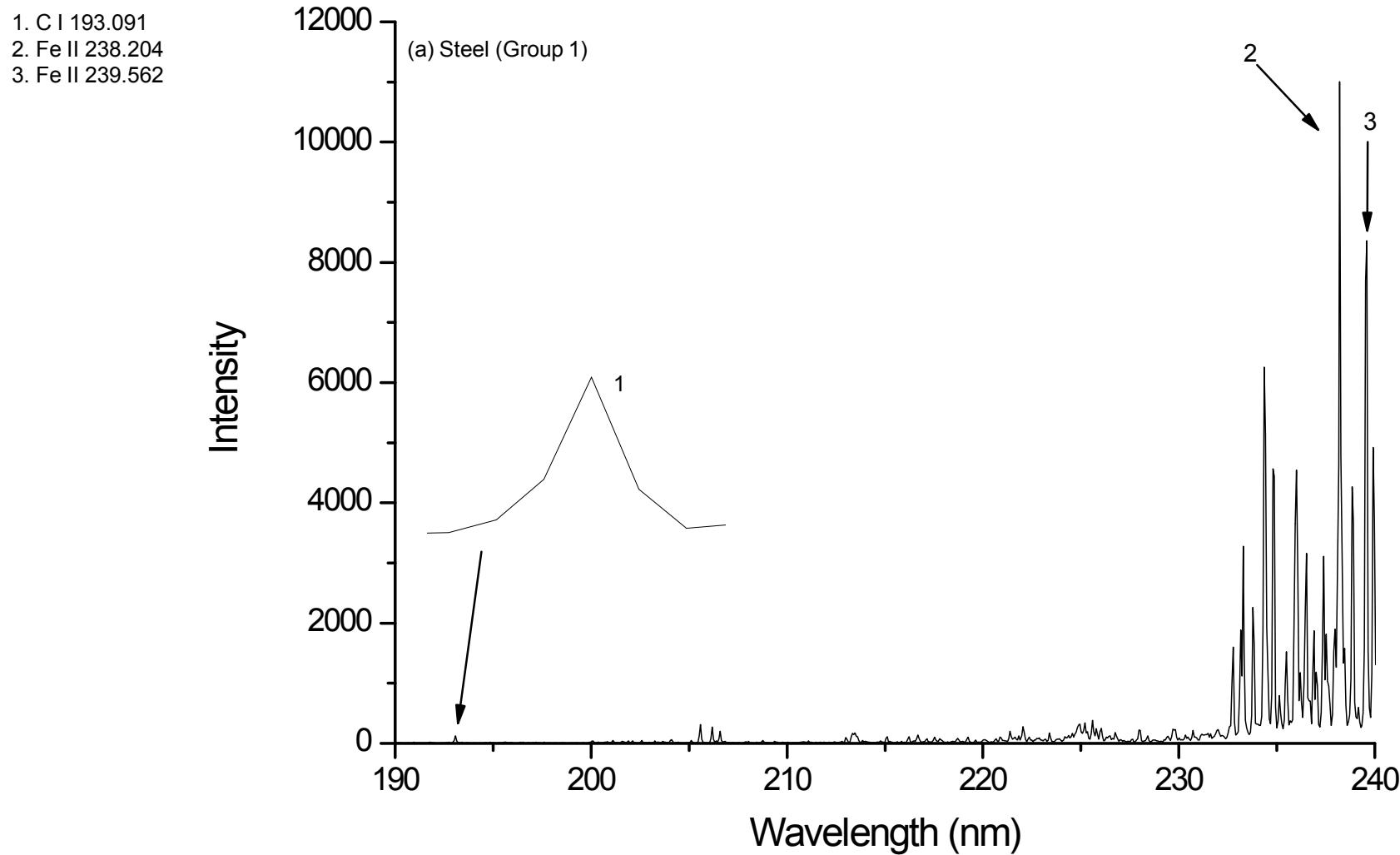


Fig. 2Sb

- 4. C I 247.856
- 5. Fe I 248.327
- 6. Fe II 259.940
- 7. Fe II 273.955
- 8. Fe II 274.648
- 9. Fe II 274.948
- 10. Fe II 275.573

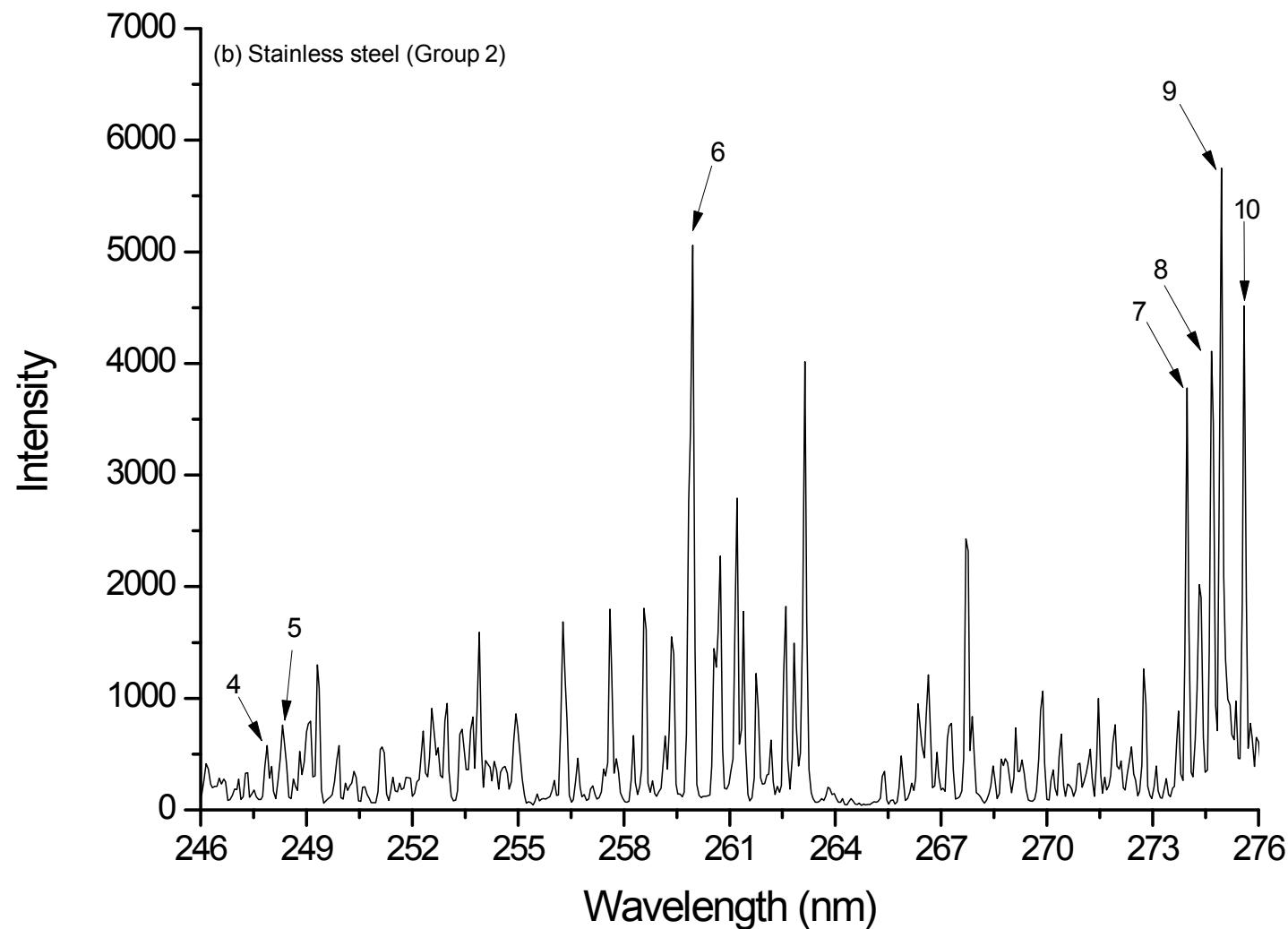


Fig. 2Sc

11. Mn II 257.610
12. Mn II 259.373
13. Mn II 294.920

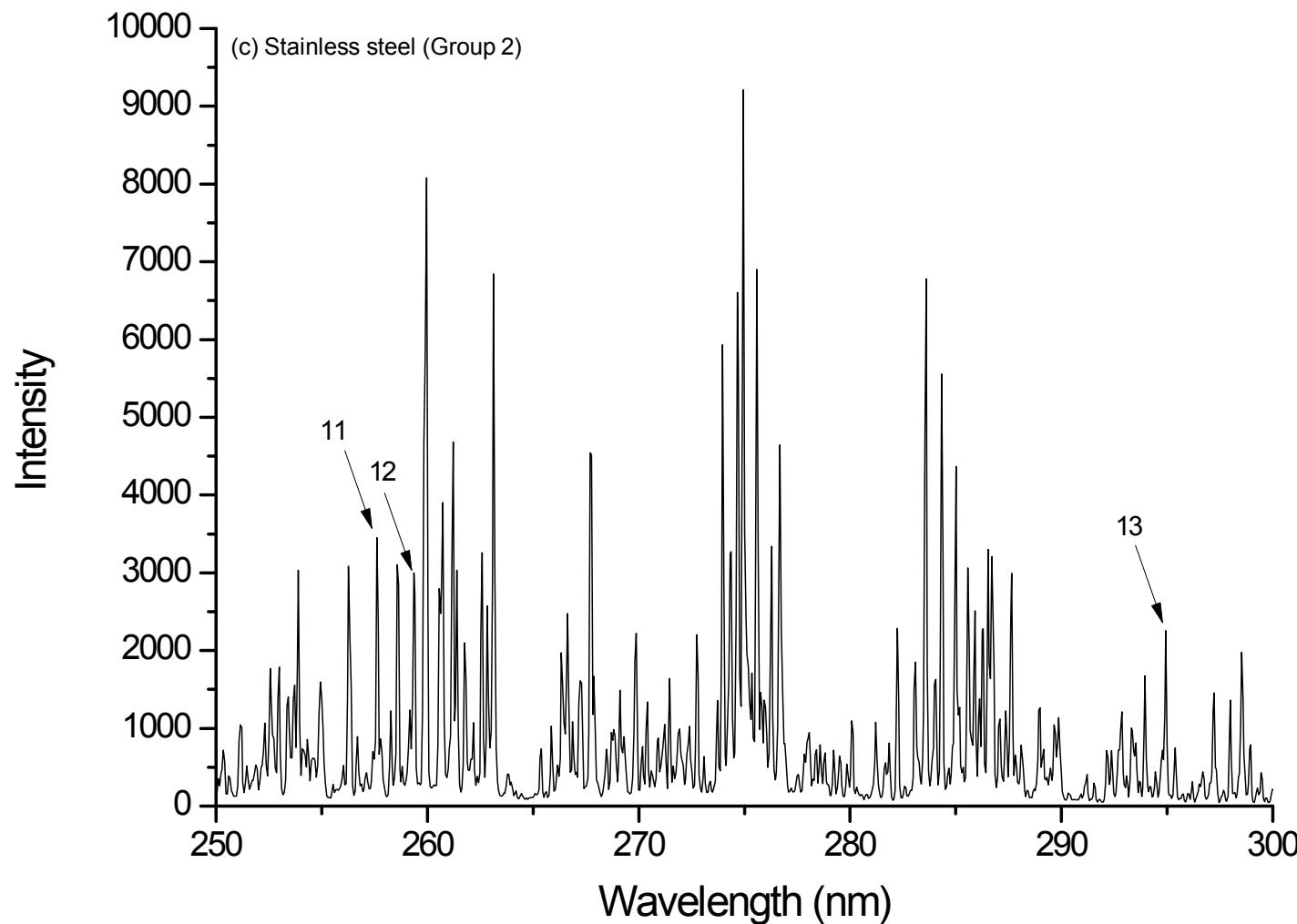


Fig. 2Sd

- 14. Zn II 255.795
- 15. Zn I 328.233
- 16. Zn I 330.258
- 17. Zn I 334.502
- 18. Zn I 468.014
- 19. Zn I 472.215
- 20. Zn I 481.053

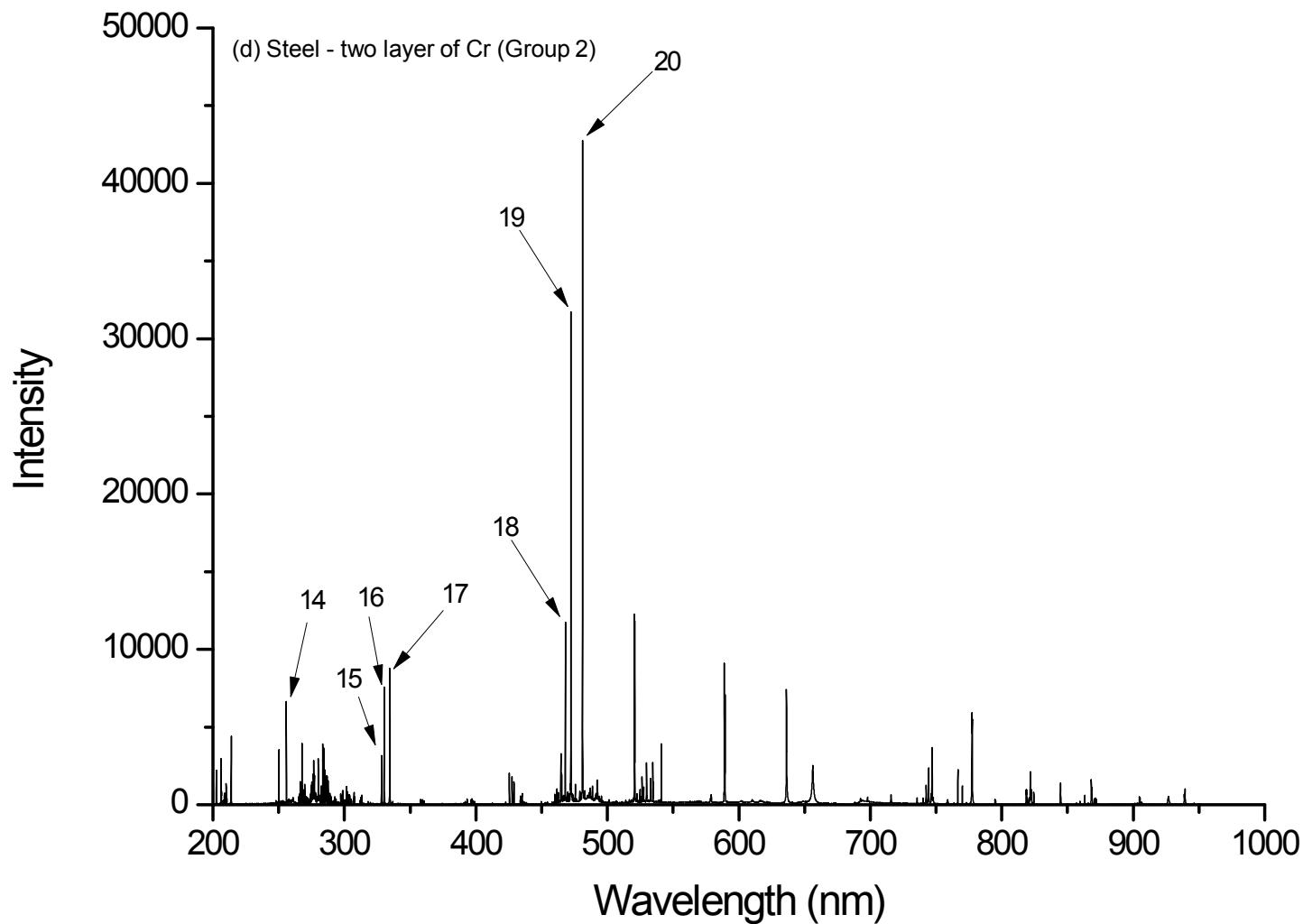


Fig. 2Se

21. Al I 308.215
22. Al I 309.271
23. Al I 394.400
24. Al I 396.152

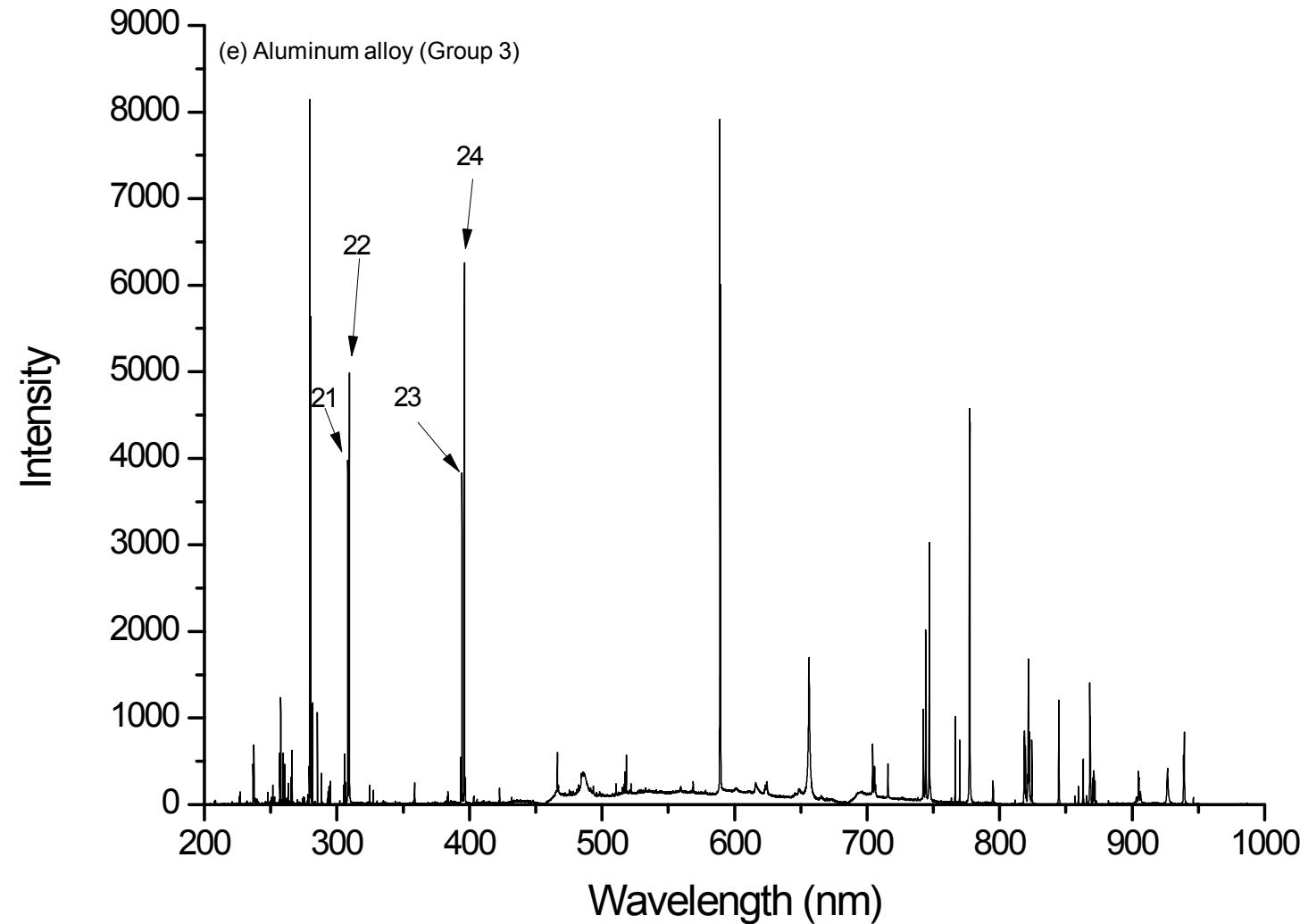


Fig. 2Sf

25. Cu II 224.700
26. Cu I 324.754
27. Cu I 327.396
28. Cu I 510.554
29. Cu I 515.324
30. Cu I 521.820
31. Cu I 578.213

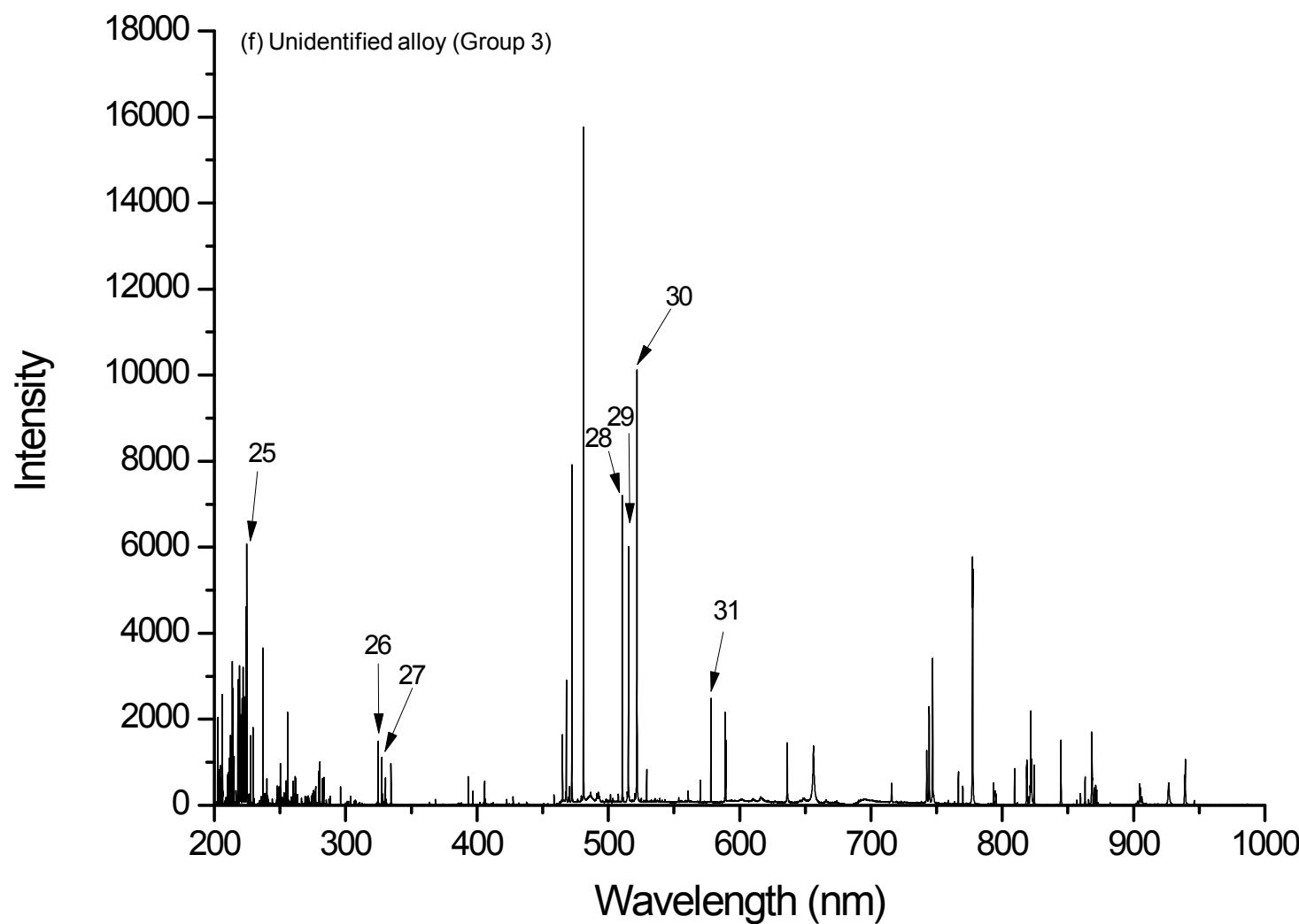


Fig. 2Sg

- 32. Co II 228.615
- 33. Co II 230.785
- 34. Co II 237.862
- 35. Cr II 267.716
- 36. Cr II 283.563
- 37. Cr II 284.325
- 38. Cr II 284.984
- 39. Cr I 357.869
- 40. Cr I 359.349
- 41. Cr I 425.435
- 42. Cr I 427.480

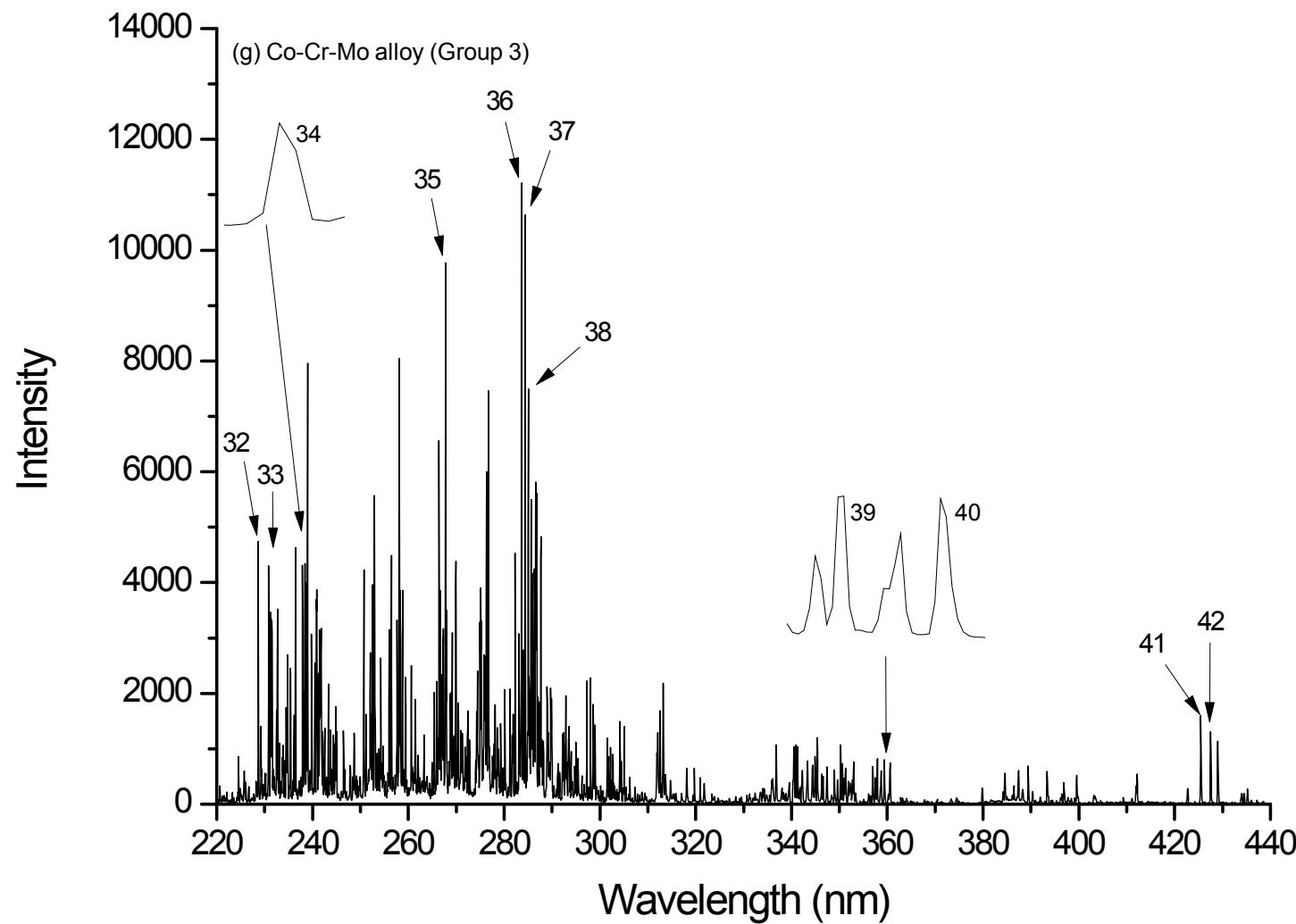


Fig. 2Sh

- 43. Ti II 307.864
- 44. Ti II 308.802
- 45. Ti I 334.188
- 46. Ti II 334.941
- 47. Ti II 368.520
- 48. V I 411.178
- 49. Ti I 430.109
- 50. V I 437.924
- 51. V I 440.820
- 52. Ti I 498.173

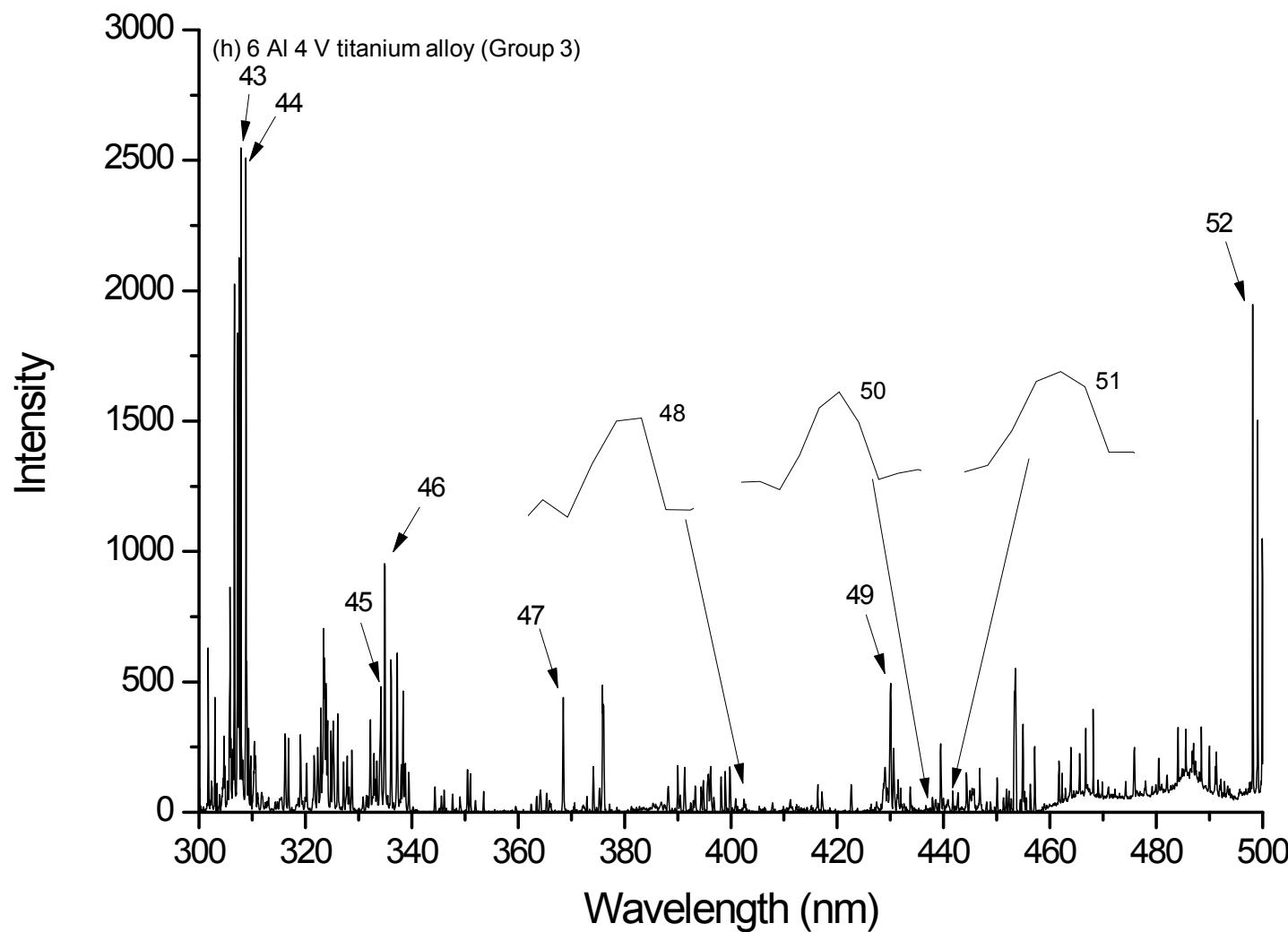


Fig. 2Si

53. Ni I 356.637
54. Ni I 361.939
55. Mo I 386.411
56. Mo I 379.825
57. Mo I 550.649

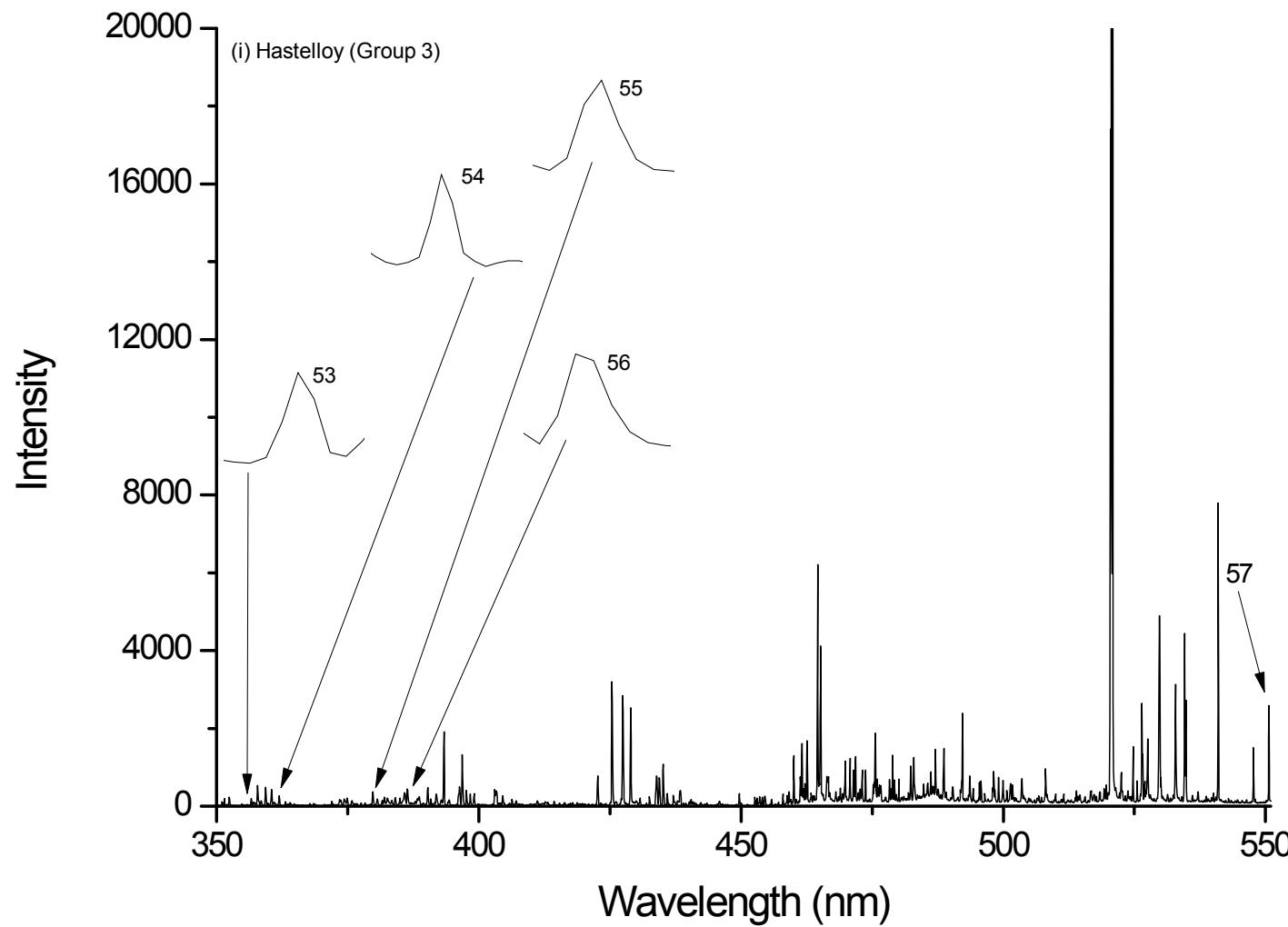
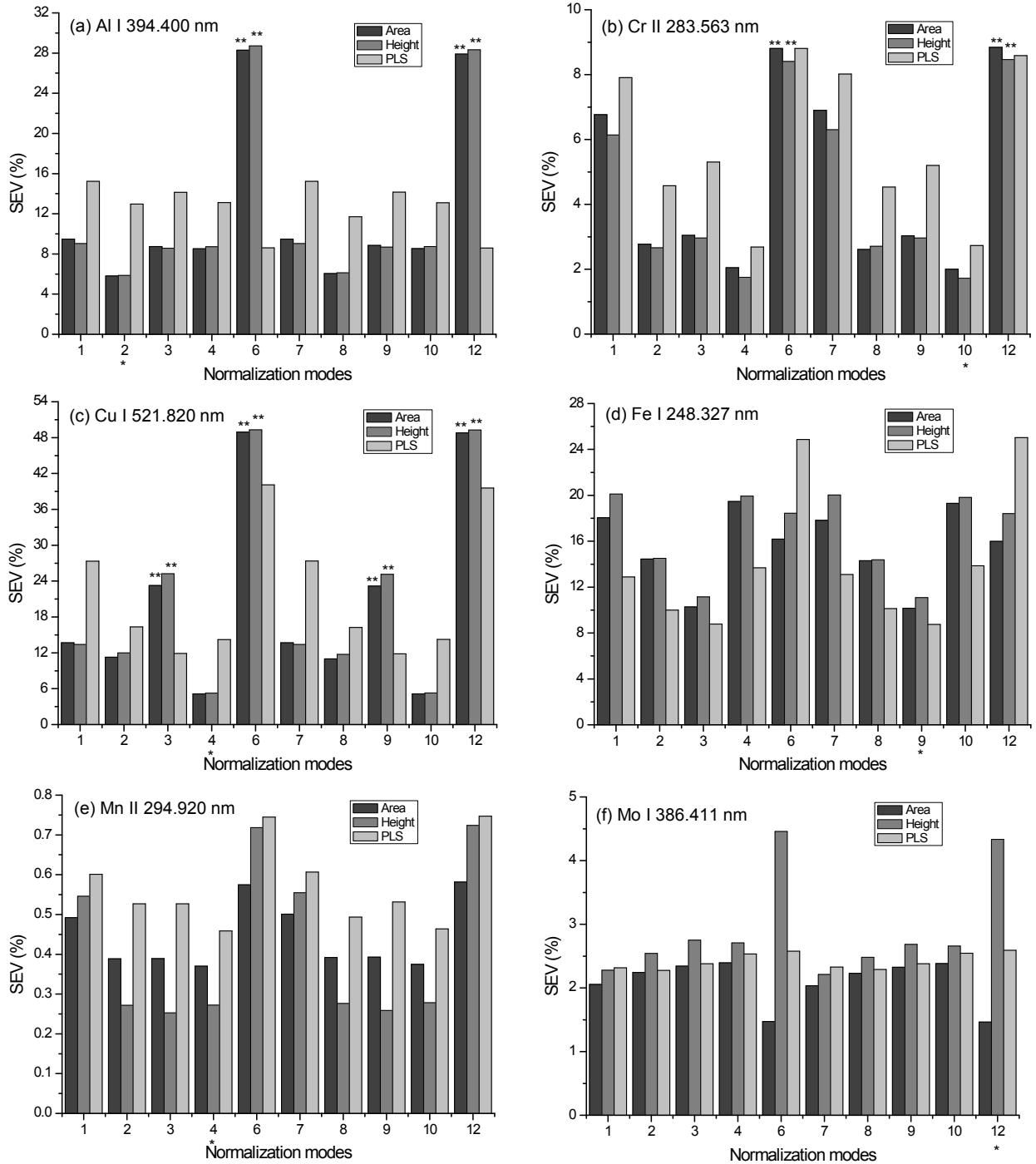


Fig. 3S

Standard error of validation for (a) Al, (b) Cr, (c) Cu, (d) Fe, (e) Mn, (f) Mo, (g) Ni, (h) Ti, (i) V and (j) Zn (SEC was used) using univariate and multivariate models applying the normalization modes proposed. An asterisk represents the mode selected and type of the signal. Normalization modes with two asterisks presented significant difference with the selected one with 95% of confidence level.



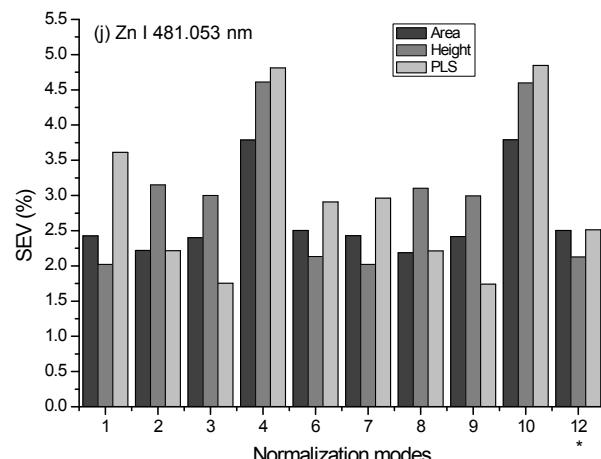
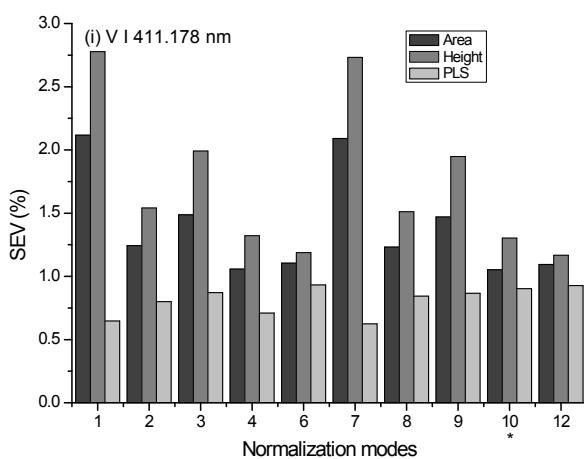
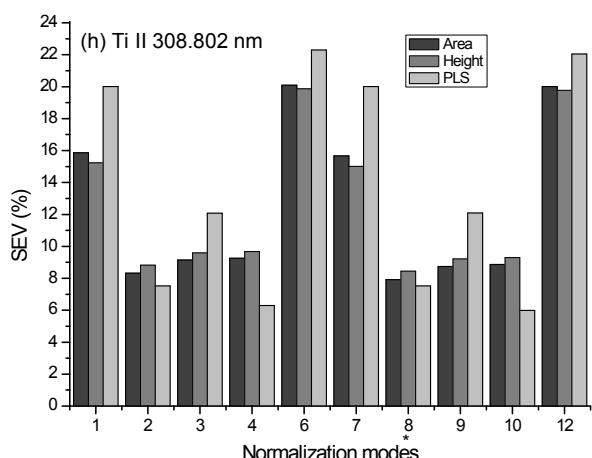
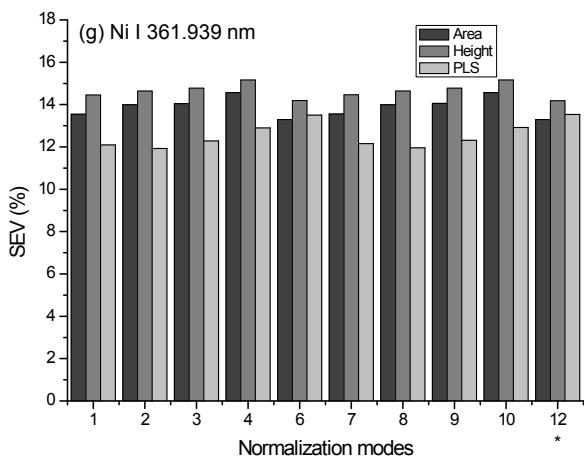


Fig. 4S

LOD's values for all normalization modes of area and height signal and LOD_{\min} , LOD_{\max} and LOD_{μ} for PLS (multivariate): (a) Al, (b) Cr, (c) Cu, (d) Fe, (e) Mn, (f) Mo, (g) Ni, (h) Ti, (i) V and (j) Zn.

