

UNIVERSIDADE FEDERAL DE SÃO CARLOS
CENTRO DE CIÊNCIAS EXATAS E DE TECNOLOGIA
DEPARTAMENTO DE QUÍMICA
PROGRAMA DE PÓS-GRADUAÇÃO EM QUÍMICA

“USO DA TÉCNICA *LASER-INDUCED BREAKDOWN SPECTROSCOPY* (LIBS) PARA ANÁLISE DIRETA DE LIGAS METÁLICAS: ESTRATÉGIAS DE NORMALIZAÇÃO, CALIBRAÇÃO UNIVARIADA E MULTIVARIADA E MODELOS DE CLASSIFICAÇÃO”

Jeyne Pricylla de Castro*

Dissertação apresentada como parte dos requisitos para obtenção do título de MESTRA EM QUÍMICA, área de concentração: QUÍMICA ANALÍTICA.

Orientador: Edenir Rodrigues Pereira Filho

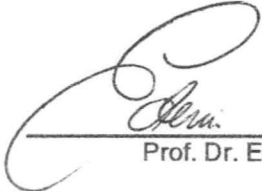
* bolsista FAPESP: 2014/22408-4

**São Carlos – SP
Fevereiro de 2017**



Folha de Aprovação

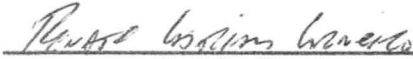
Assinaturas dos membros da comissão examinadora que avaliou e aprovou a Defesa de Dissertação de Mestrado da candidata Jeyne Pricylla de Castro, realizada em 10/02/2017:



Prof. Dr. Edenir Rodrigues Pereira Filho
UFSCar



Prof. Dr. Dário Santos Junior
UNIFESP



Prof. Dr. Renato Lajarim Carneiro
UFSCar

*"A mente que se abre a um novo conhecimento jamais
retorna ao seu tamanho original".*

Albert Einstein

*"Porque dele, e por meio dele, e para Ele, são todas as coisas; glória, pois a Ele
eternamente. Amém. "*

Romanos 11,36

**Dedico este trabalho aos pilares da minha vida:
Deus, minha família e meus amigos.**

Agradecimentos

- ✓ À Deus por tudo que tem feito em minha vida. Por ser meu sustento e minha força. A Ele toda honra, louvor e glória para todo o sempre.
- ✓ À toda a minha família, principalmente os meus pais Vânia e Vandeil, por me ensinarem tudo que sou hoje. Sou grata a eles pelo extraordinário exemplo de amor, dedicação e por sempre acreditarem em mim, apoiando todas as minhas escolhas.
- ✓ Ao meu namorado Leonardo, o qual trouxe mais alegria a minha vida. Obrigada por todo amor, paciência e companheirismo.
- ✓ Às minhas amigas Jéssica, Maíra e Taíse. Cada uma diferente da outra, mas juntas compartilhamos alegrias e tristezas, formando uma família. Obrigada por todo amor e amizade. Ao amigo Daniel, por toda a alegria, apoio e companheirismo que nos trouxe desde que chegou em São Carlos. As amigas de Piumhi, que mesmo com a distância, estão sempre presentes. E ao grupo Jovens Sarados São Paulo Apóstolo, por tudo que representa em minha vida.
- ✓ Ao professor Edenir, por toda orientação e por compartilhar ensinamentos valiosos. Obrigada pela amizade e por todos os conselhos.
- ✓ Aos professores Joaquim e Ana Rita, por todos os ensinamentos e colaborações nas reuniões de grupo.
- ✓ A todos os amigos do GAIA, por todo o apoio, ajuda, ensinamentos e por tornar a convivência no laboratório mais amena e feliz.
- ✓ Ao professor Dr. Ronei J. Poppi e ao Dr. José Augusto Dacol pelas determinações por XRF.
- ✓ Ao Dr. Marcos Monteiro por ter cedido gentilmente algumas amostras de ligas metálicas.
- ✓ À Thermo Scientific pelas medidas de algumas amostras no XRF Niton XL3T 985 ULTRA.
- ✓ Aos professores da banca, por todos os ensinamentos e contribuições.
- ✓ À UFSCar e aos funcionários do Departamento de Química por toda ajuda e colaboração.
- ✓ À FAPESP (Processo 2014/22408-4) pela bolsa de estudos durante esses dois anos de mestrado.
- ✓ A todos que de alguma forma, direta ou indiretamente, contribuíram e fizeram parte deste trabalho. Muito obrigada!

LISTA DE ACRÔNIMOS

ANOVA	Análise de Variância (<i>Analysis of variance</i>)
CCD	Dispositivo de Carga Acoplada (<i>Charge coupled device</i>)
CLS	Regressão por quadrados mínimos – clássico (<i>Classical Least Squares, CLS</i>)
ICP OES	Espectrometria de emissão óptica com plasma indutivamente acoplado (<i>Inductively coupled plasma optical emission spectrometry</i>)
KNN	Método dos K-ésimos Vizinhos mais próximos (<i>K-Nearest Neighbor</i>)
LDA	Análise Discriminante Linear (<i>Linear Discriminant Analysis</i>)
LIBS	Espectroscopia de emissão em plasma induzido por laser (<i>Laser induced breakdown spectroscopy</i>)
MLR	Regressão linear múltipla (<i>Multiple linear regression</i>)
PC	Componente Principal (<i>Principal Component</i>)
PCR	Regressão por componentes principais (<i>Principal Component Regression</i>)
PLS	Mínimos quadrados parciais (<i>Partial least squares</i>)
PLS-DA	Mínimos Quadrados Parciais para análise discriminante (<i>Partial Least Squares-Discriminant Analysis</i>)
PRESSval	Soma dos quadrados dos erros de previsão (<i>Predict Residual Error Sum of Squares</i>)
REP	Erro Relativo de Previsão (<i>Relative error of prediction</i>)
RSD	Desvio padrão relativo (<i>Relative Standard Deviation, RSD</i>)
SBR	Relação Sinal Radiação de Fundo (<i>Signal-to-background ratio</i>)
SEC	Erro Padrão de Calibração (<i>Standard Error of calibration</i>)
SECV	Erro Padrão de Validação Cruzada (<i>Standard Error of Cross-validation</i>)
SEP	Erro Padrão de Previsão (<i>Standard error of prediction</i>)
SEV	Erro Padrão de Validação (<i>Standard Error of validation</i>)
SIMCA	<i>Soft Independent Modeling of Class Analogy</i>
SNR	Relação Sinal Ruído (<i>Signal-to-noise ratio</i>)
XRF	Fluorescência de Raios-X (<i>X-Ray fluorescence</i>)

LISTA DE TABELAS

TABELA 4.1.1- Identificação (ID), descrição e imagem das amostras 1 a 20.	19
TABELA 4.1.2 - Identificação (ID), descrição e imagem das amostras 21 a 38.	20
TABELA 4.1.3- Identificação (ID), descrição e imagem das amostras 39 a 54.	21
TABELA 4.3.1. Planejamento fatorial do tipo Doehlert.	24
TABELA 4.3.2. Identificação e equações dos parâmetros da tabela ANOVA.	26
TABELA 4.4.1- Ilustração dos modos de normalização utilizando a média dos espectros normalizados.	28
TABELA 4.4.2- Ilustração dos modos de normalização utilizando a soma dos espectros normalizados.	29
TABELA 5.1. 1. Resultados (% m/m) do XRF para as amostras 1 a 27.	34
TABELA 5.1. 2. Resultados (% m/m) do XRF para as amostras 28 a 45_2.	35
TABELA 5.1. 3. Resultados (% m/m) do XRF para as amostras 45_3 a 54_2.	36
TABELA 5.2. 1. ANOVA para o Fe II 259,940 nm da amostra 54.	38
TABELA 5.3. 1. Comparação entre os modelos univariados (área e altura) e multivariados para os 10 analitos estudados.	57

LISTA DE FIGURAS

FIGURA 2.3.1- Esquema ilustrativo dos principais componentes da técnica LIBS. O sistema ilustrado é composto por fonte de laser, laser, espelhos, lentes, Espectrômetro, detector e computador.	14
FIGURA 4.2.1- Visão geral do equipamento LIBS J200 da Applied Spectra. A maçaneta frontal dá acesso ao porta-amostra.	22
FIGURA 4.2.2- Equipamento XRF Innov-X com o compartimento da amostra (a) aberto e (b) fechado.	23
FIGURA 4.2.3- Equipamento XRF Niton XL3T da Thermo Scientific.	23
FIGURA 5.2. 1. Gráficos de contorno (delay time versus energia) e seus respectivos modelos para a área do sinal. (a) Zn I 481,053 nm; (b) Cu I 324,754 nm; (c) Mo I 386,411 nm; (d) Ni I 361,939 nm; (e) Fe II 259,940 nm; (f) Cr II 283,563 nm.	40
FIGURA 5.2. 2. Gráficos de contorno e seus respectivos modelos para a área do sinal. Delay time versus energia (g) Fe II 259,940 nm; (h) V I 437,924 nm; (i) Mn II 294,920 nm; (j) Al I 394,400 nm; (k) delay time versus spot size (Al I 394,400 nm); (l) energia versus spot size (Al I 394,400 nm).	41
FIGURA 5.2. 3. Gráficos de contorno e seus respectivos modelos para a área do sinal. (m) delay time versus energia (Fe II 259,940 nm); (n) delay time versus spot size (Fe II 259,940 nm); (o) energia versus spot size (Fe II 259,940 nm).	42
FIGURA 5.2. 4. Comparação entre os valores de área obtidas e previstas com as condições otimizadas. (a) Amostra 1 (Zn I 481,053 nm); (b) 2 (Cu I 324,754 nm); (c) 10 (Mo I 386,411 nm); (d) 10 (Ni I 361,939 nm); (e) 35 (Fe II 259,940 nm) e (f) 44 (Cr II 283,563 nm).	45
FIGURA 5.2. 5. Comparação entre os valores de área obtidas e previstas com as condições otimizadas. (g) Amostra 45 (Fe II 259,940 nm); (h) 51 (V I 437,924 nm); (i) 54 (Mn II 294,920 nm); (j) 38 (Al I 394,400 nm) e (k) 54 (Fe II 259,940 nm).	46
FIGURA 5.2. 6. Gráficos das razões das áreas das linhas de emissão. (a) 357,869/283,563; (b) 357,869/359,349; (c) 357,869/425,435; (d) 357,869/427,48; (e) 283,563/359,349 e (f) 283,563/425,435.	48
FIGURA 5.2. 7. Gráficos das razões das áreas das linhas de emissão. (g) 283,563/427,48; (h) 359,349/425,435; (i) 359,349/427,48 e (j) 425,435/427,48.	49

FIGURA 5.3. 1. Resultados da calibração univariada utilizando a área e altura do sinal e calibração multivariada utilizando PLS em todos os modos de normalização para o Al. (a) Valores de SEV (%) e (b) Valores do teste F.	51
FIGURA 5.3. 2. Resultados da calibração univariada utilizando a área e altura do sinal e calibração multivariada utilizando PLS em todos os modos de normalização para o Cr. (a) Valores de SEV (%) e (b) Valores do teste F.	51
FIGURA 5.3. 3. Resultados da calibração univariada utilizando a área e altura do sinal e calibração multivariada utilizando PLS em todos os modos de normalização para o Cu. (a) Valores de SEV (%) e (b) Valores do teste F.	51
FIGURA 5.3. 4. Resultados da calibração univariada utilizando a área e altura do sinal e calibração multivariada utilizando PLS em todos os modos de normalização para o Fe. (a) Valores de SEV (%) e (b) Valores do teste F.	52
FIGURA 5.3. 5. Resultados da calibração univariada utilizando a área e altura do sinal e calibração multivariada utilizando PLS em todos os modos de normalização para o Mn. (a) Valores de SEV (%) e (b) Valores do teste F.	52
FIGURA 5.3. 6. Resultados da calibração univariada utilizando a área e altura do sinal e calibração multivariada utilizando PLS em todos os modos de normalização para o Mo. (a) Valores de SEV (%) e (b) Valores do teste F.	52
FIGURA 5.3. 7. Resultados da calibração univariada utilizando a área e altura do sinal e calibração multivariada utilizando PLS em todos os modos de normalização para o Ni. (a) Valores de SEV (%) e (b) Valores do teste F.	53
FIGURA 5.3. 8. Resultados da calibração univariada utilizando a área e altura do sinal e calibração multivariada utilizando PLS em todos os modos de normalização para o Ti. (a) Valores de SEV (%) e (b) Valores do teste F.	53
FIGURA 5.3. 9. Resultados da calibração univariada utilizando a área e altura do sinal e calibração multivariada utilizando PLS em todos os modos de normalização para o V. (a) Valores de SEV (%) e (b) Valores do teste F.	53
FIGURA 5.3. 10. Resultados da calibração univariada utilizando a área e altura do sinal e calibração multivariada utilizando PLS em todos os modos de normalização para o Zn. (a) Valores de SEV (%) e (b) Valores do teste F.	54
FIGURA 5.3. 11. Valores de referência (medido) e os valores previstos (multivariado e univariado) para o Al. (a) Conjunto de calibração e (b) Conjunto de validação.	58
FIGURA 5.3. 12. Valores de referência (medido) e os valores previstos (multivariado e univariado) para o Cr. (a) Conjunto de calibração e (b) Conjunto de validação.	59

FIGURA 5.3. 13. Valores de referência (medido) e os valores previstos (multivariado e univariado) para o Cu. (a) Conjunto de calibração e (b) Conjunto de validação. ...	59
FIGURA 5.3. 14. Valores de referência (medido) e os valores previstos (multivariado e univariado) para o Fe. (a) Conjunto de calibração e (b) Conjunto de validação.	59
FIGURA 5.3. 15. Valores de referência (medido) e os valores previstos (multivariado e univariado) para o Mn. (a) Conjunto de calibração e (b) Conjunto de validação. ...	60
FIGURA 5.3. 16. Valores de referência (medido) e os valores previstos (multivariado e univariado) para o Mo. (a) Conjunto de calibração e (b) Conjunto de validação. ...	60
FIGURA 5.3. 17. Valores de referência (medido) e os valores previstos (multivariado e univariado) para o Ni. (a) Conjunto de calibração e (b) Conjunto de validação.....	60
FIGURA 5.3. 18. Valores de referência (medido) e os valores previstos (multivariado e univariado) para o Ti. (a) Conjunto de calibração e (b) Conjunto de validação.	61
FIGURA 5.3. 19. Valores de referência (medido) e os valores previstos (multivariado e univariado) para o V. (a) Conjunto de calibração e (b) Conjunto de validação.....	61
FIGURA 5.3. 20. Valores de referência (medido) e os valores previstos (multivariado e univariado) para o Zn.	61
FIGURA 5.3. 21. Valores de SEV para o Cr em diferentes conjuntos de dados.	62
FIGURA 5.3. 22. Valores de RSD (%) para os diferentes analitos estudados.	63
FIGURA 5.4. 1. Espectro característico da amostra 3 pertencente a classe 1.....	63
FIGURA 5.4. 2. Espectro característico da amostra 2 pertencente a classe 2.....	64
FIGURA 5.4. 3. Espectro característico da amostra 6 pertencente a classe 3.....	64
FIGURA 5.4. 4. Exatidão, sensibilidade, taxa de falso alarme e especificidade para os modelos de classificação (KNN, SIMCA e PLS-DA) para as matrizes de (a) Concentrações do XRF e (b) Altura do sinal LIBS.	65
FIGURA 5.4. 5. Exatidão, sensibilidade, taxa de falso alarme e especificidade para os modelos de classificação (KNN, SIMCA e PLS-DA) para as matrizes de (c) Área do sinal LIBS.	66

RESUMO

USO DA TÉCNICA *LASER-INDUCED BREAKDOWN SPECTROSCOPY* (LIBS) PARA ANÁLISE DIRETA DE LIGAS METÁLICAS: ESTRATÉGIAS DE NORMALIZAÇÃO, CALIBRAÇÃO UNIVARIADA E MULTIVARIADA E MODELOS DE CLASSIFICAÇÃO

Esse trabalho foi dedicado ao desenvolvimento de métodos analíticos para a determinação de Al, Cr, Cu, Fe, Mn, Mo, Ni, Ti, V e Zn em ligas metálicas. A principal proposta do estudo foi apresentar a *Laser-induced breakdown spectroscopy* (LIBS) como uma alternativa viável para a análise direta de amostras de ligas e aços utilizando ferramentas quimiométricas para interpretar os dados obtidos. Inicialmente, realizou-se a otimização dos parâmetros do equipamento LIBS utilizando planejamento fatorial do tipo Doehlert, variando a energia do laser em 7 níveis (30 a 80 mJ), *delay time* em 5 níveis (0 a 2 μ s) e o *spot size* em 3 níveis (50 a 150 μ m). A condição de compromisso escolhida foi 60 mJ de energia, 0,9 μ s de *delay time* e 100 μ m de *spot size*, a qual foi aplicada em 80 amostras. Os valores de referência dos analitos foram obtidos utilizando a técnica de Fluorescência de Raios-X (*X-ray fluorescence*, XRF) para a construção de modelos de calibração. Para minimizar as variações do sinal e as diferenças das matrizes das amostras, foram testados doze modos de normalizações e duas estratégias de calibração. Foram estudadas: calibração multivariada utilizando *Partial Least Squares* (PLS) e calibração univariada empregando área e altura de várias linhas de emissão. Assim, buscou-se a identificação do melhor modo de normalização, linha de emissão e estratégia de calibração para cada analito. Para a maioria dos analitos, não houve diferença significativa entre os modos de normalização e também entre a calibração univariada e multivariada. Além dos modelos de calibração, foram aplicados modelos de classificação para identificar as amostras em 3 grupos diferentes. *K-nearest neighbor* (KNN), *Soft independent modeling of class analogy* (SIMCA) e *Partial-least squares-discriminant analysis* PLS-DA foram utilizados em 3 matrizes diferentes: concentrações obtidas por XRF (valores de referência), área e altura das linhas de emissão da LIBS (total de 57 linhas de emissão). Ao comparar os modelos, foram avaliadas algumas figuras de mérito como exatidão, sensibilidade, taxa de falso alarme e especificidade. O modelo de classificação que obteve melhores resultados foi o KNN. Como conclusão do trabalho, o planejamento fatorial foi útil para obter uma condição adequada de análise para todos os analitos e amostras simultaneamente, economizando tempo e recursos. Os modos de normalização foram eficazes para minimizar as variações dos sinais e as diferenças nas matrizes das amostras. Os modelos univariados foram mais satisfatórios do que os multivariados. No caso dos modelos de classificação, foi possível identificar as amostras, sendo o modelo KNN mais eficiente do que os demais.

ABSTRACT

USE OF *LASER-INDUCED BREAKDOWN SPECTROSCOPY* (LIBS) TECHNIQUE FOR DIRECT ANALYSES OF METTALIC ALLOYS: NORMALIZATION STRATEGIES, UNI AND MULTIVARIATE CALIBRATIONS AND CLASSIFICATION MODELS

This academic master's dissertation was devoted to the development of analytical methods for the determination of Al, Cr, Cu, Fe, Mn, Mo, Ni, Ti, V and Zn in alloys and steels. The main purpose of the study was to present the Laser-induced Breakdown Spectroscopy (LIBS) as a viable alternative for the direct analysis of alloys and steels using chemometric tools to interpret the obtained data. Initially, the optimization of the parameters of the LIBS equipment was done using Doehlert design, varying the laser energy in 7 levels (30 to 80 mJ), delay time in 5 levels (0 to 2 μ s) and spot size in 3 levels (50 to 150 μ m). The chosen compromise condition was 60 mJ of energy, 0.9 μ s of delay time and 100 μ m of spot size, which were applied to 80 samples. The reference values of the analytes were obtained using the X-ray Fluorescence (XRF) technique for the construction of calibration models. To minimize signal variations and sample matrix differences, twelve normalization modes were tested and two calibration strategies were studied: multivariate calibration using Partial Least Squares (PLS) and univariate calibration using area and height of several emission lines. Thus, we search to identify the best mode of normalization, emission line and calibration strategy for each analyte. For most analytes, there was no significant difference between the normalization modes and also between the univariate and multivariate calibration. Classification models were applied to identify the samples in 3 different groups. K-nearest neighbor (KNN), Soft independent modeling of class analogy (SIMCA) and Partial-least squares-discriminant analysis PLS-DA were used in 3 different matrices: concentrations obtained using XRF, height and area of the LIBS emission lines (total of 57 emission lines). When comparing the models, some merit figures were evaluated, such as accuracy, sensitivity, false alarm rate and specificity. The classification model that obtained the best results was KNN. As a conclusion of the work, factorial design was useful to obtain an adequate analysis condition for all analytes and samples simultaneously, saving time and resources. Normalization modes were effective to minimize signal variations and differences in sample matrices. Univariate models were more satisfactory than multivariate models. In the case of classification models, it was possible to identify the samples, being the KNN model more efficient than the others.

SUMÁRIO

1- Introdução	1
2- Revisão Bibliográfica.....	4
2.1- Planejamento fatorial.....	4
2.2- Métodos Quimiométricos.....	7
2.2.1- Pré-tratamento dos dados	7
2.2.2- Modelos de regressão.....	9
2.2.3- Modelos de classificação	12
2.3- Aplicação da Laser-induced breakdown spectroscopy na análise de ligas metálicas	13
3- Objetivo e justificativa	17
4- Materiais e métodos.....	18
4.1- Amostras	18
4.2- Instrumentação	22
4.3- Otimização das condições do equipamento LIBS utilizado	23
4.4- Modelos de calibração univariado e multivariado.....	26
4.5- Modelos de classificação	31
5- Resultados e discussão	33
5.1. Caracterização das amostras	33
5.2- Otimização dos parâmetros da LIBS.....	37
5.3- Modelos de calibração univariado e multivariado.....	50
5.4- Modelos de classificação	63
6- Conclusão	67
7- Referências	68
Anexos	71

1- Introdução

A importância dos metais e suas ligas para novas aplicações se configura como um grande fator de desenvolvimento científico e tecnológico nos campos da medicina, aeroespacial, automobilística, entre outros. Desta forma, é notório o aumento do interesse pela elaboração de diferentes e eficazes ligas metálicas e a busca por técnicas analíticas capazes de aferir a qualidade do que está sendo efetivamente produzido.

A união de dois ou mais metais formam as ligas metálicas, as quais são preparadas pelo aquecimento conjunto dos metais até a sua fusão completa e depois o esfriamento dos mesmos até solidificar completamente. As suas propriedades vão depender de vários fatores: dos elementos e as proporções em que se encontram nas ligas, estrutura cristalina da liga e até mesmo os tratamentos térmicos (aquecimento da liga, seguido do resfriamento mais rápido ou mais lento), termodinâmicos ou superficial a que se deve submeter a liga metálica para alcançar a propriedade desejada¹.

As ligas metálicas são classificadas como ferrosas e não ferrosas. As principais ligas ferrosas são os aços, os quais podem ter duas denominações: aços-carbono e aços-liga. Os aços-carbono são constituídos basicamente de carbono e ferro em sua composição, possuindo apenas alguns elementos residuais retidos durante o processo de fabricação, como o manganês, silício, fósforo e enxofre. Esses aços são usados quando os requisitos de resistências mecânica e à corrosão não são muito severos, possuindo vantagens como o custo relativamente baixo e pouca exigência de tratamentos elaborados para sua produção. Os aços-liga possuem outros elementos adicionados na sua composição química como: cromo, níquel, molibdênio, tungstênio, manganês, silício e cobre. Elementos como vanádio, nióbio, boro e titânio também podem ser adicionados em pequenos teores para melhorar ainda mais a qualidade do aço. Um aço de baixa liga é aquele que possui menos de 10% de elementos de liga além do Fe e um de alta liga possui 10% ou mais de elementos de liga. A principal finalidade da adição desses elementos é favorecer os efeitos benéficos dos tratamentos térmicos¹. Assim, a adição de cada elemento possui efeitos nos aços:

- Carbono: propriedades referentes a dureza e resistência;
- Manganês: resistência mecânica;

- Níquel: melhora as propriedades mecânicas e resistência a corrosão;
- Cobre: resistência a corrosão atmosférica;
- Cromo: resistência mecânica (teores baixos) e resistência ao desgaste (teores altos);
- Molibdênio: melhora as propriedades mecânicas a temperaturas mais elevadas;
- Chumbo: melhora a usinagem e acabamento;
- Boro: melhora a têmpera do aço;
- Zircônio: desoxidante e aumenta a tenacidade;
- Selênio: facilita a usinagem².

De acordo com o Instituto Aço Brasil, a produção acumulada de aço bruto no país de janeiro a agosto de 2016 totalizou cerca de 20,3 milhões de toneladas. As exportações de produtos siderúrgicos até agosto de 2016 representaram 8,7 milhões de toneladas com faturamento de 3,5 milhões de dólares, crescimento de 2,7% em volume e redução de 21,7% em valor, quando comparados ao mesmo período do ano anterior. No que se refere as importações, registrou-se em agosto de 2016 o volume de 110 mil toneladas (123 milhões de dólares), totalizando 963 mil toneladas de produtos siderúrgicos importados de janeiro a agosto de 2016 com um faturamento de 1 bilhão de dólares. O Brasil é o maior produtor de aço bruto da América Latina³.

As ligas não ferrosas são formadas por outros elementos químicos, os quais se apresentam em maior quantidade, que não seja o ferro. Aquelas que possuem elementos como cobre, alumínio, titânio, manganês e níquel representam uma grande importância comercial. Ao serem comparadas com ligas ferrosas, apresentam propriedades interessantes como:

- Menor densidade;
- Maior condutividade elétrica e;
- Maior resistência à corrosão.

Dependendo da quantidade e do analito presente na liga, as suas características serão diferentes. Assim, pequenas variações podem representar mudanças substanciais nas propriedades das mesmas, sendo necessárias análises químicas para garantir suas propriedades⁴. Na TABELA 1.1, tem-se as características e aplicações de algumas ligas metálicas².

TABELA 1. 1. Características e aplicações de diferentes ligas metálicas.

Ligas (elementos)	Características	Aplicações
Alumínio (Al, Si e Cu)	Resistência a corrosão e a altas temperaturas	Utensílios domésticos
Alumínio (Al, Mn, Cu e Mg)	Superfície brilhante, alta resistência mecânica e liga leve	Aviação, peças automotivas e para bicicletas
Alumínio (Mg e Al)	Superfície brilhante e resistência mecânica	Instrumentos científicos
Cobre - Latão (Zn e Cu)	Resistência a corrosão, flexibilidade e boa aparência	Tubos, torneiras e objetos de decoração
Cobre - Bronze (Sn e Cu)	Resistência a corrosão e flexibilidade	Sinos, lustres e objetos de decoração
Cobre (Zn, Sn e Cu)	Resistência a corrosão causada pela água do mar	Tubulações, indústria naval e equipamentos marítimos
Cobre (Ni e Cu)	Elevada dureza, inoxidabilidade e alto brilho	Confecção de moedas
Níquel (Ni e Cu)	Resistência a corrosão e superfície brilhante	Indústria alimentícia
Níquel (Ni e Cr)	Alto ponto de fusão e baixa condutividade elétrica	Resistências elétricas
Níquel (Ni e Fe)	Propriedades magnéticas	Cabos telefônicos
Ouro (Au, Cu e Ag)	Maior dureza	Jóias e circuitos elétricos
Chumbo (Sn e Pb)	Fusível a baixas temperaturas	Soldas elétricas

Desta forma, essa dissertação de mestrado acadêmico visou a proposição de métodos rápidos e confiáveis para a análise direta de ligas metálicas e aços. No desenvolvimento da parte experimental utilizou-se a técnica espectroscopia de emissão em plasma induzido por laser (*Laser-induced breakdown spectroscopy*, LIBS) que permite a análise sem preparo (ou mínimo preparo) de amostras sólidas (análise direta).

2- Revisão Bibliográfica

2.1- Planejamento fatorial

A quimiometria tem por objetivo extrair o máximo de informações relevantes de dados obtidos por experimentos químicos, onde engloba uma variedade de métodos matemáticos, estatísticos e lógicos⁵. Pode ser dividida em quatro itens:

- Análise exploratória dos dados;
- Calibração multivariada;
- Modelos de classificação;
- Planejamento fatorial⁶.

O planejamento fatorial é uma ferramenta versátil que estuda a influência de variáveis, extraíndo do sistema em estudo informações úteis com a execução de um número reduzido de experimentos. O planejamento fatorial possui diferentes vantagens quando comparado com experimentos efetuados de forma univariada, como por exemplo:

- São mais eficientes e econômicos;
- Abrangentes, pois há o estudo de duas ou mais variáveis simultaneamente;
- Desenvolvimento de modelos matemáticos avaliando a pertinência dos mesmos e os efeitos significativos dos fatores estudados;
- Como a influência das variáveis é analisada ao mesmo tempo, menos experimentos são realizados, possibilitando a economia de recursos financeiros (baixo consumo de reagentes) e esforços (menos trabalho laboratorial);
- Desde o início da parte experimental é possível conhecer o número de experimentos necessários para estudar o problema e atingir os objetivos desejados^{5,7}.

Em qualquer planejamento deve-se decidir quais são os fatores (variáveis) e as respostas de interesse. Os fatores são aquelas variáveis qualitativas (tipo de catalisador, por exemplo) ou quantitativas (volume de determinado reagente, por exemplo) controláveis, as quais serão estudadas em diferentes níveis. Já as respostas são variáveis de saída do sistema que podem ser influenciadas ou não por modificações nos fatores, sendo as variáveis dependentes, onde os valores vão depender dos níveis dos fatores. Quando identificados todos os fatores e respostas,

é importante definir qual será o objetivo que se deseja alcançar com os experimentos e, desta forma, selecionar o planejamento ideal para o problema em estudo⁸.

A primeira etapa é realizar uma triagem das variáveis estudadas utilizando planejamento fatorial completo (2 a 4 variáveis) ou planejamento fatorial fracionário (mais de 5 variáveis), onde é possível identificar quais as variáveis mais importantes sobre uma dada resposta. Os níveis dessas variáveis podem ser codificados, pois se as variáveis independentes são qualitativas é necessário codificar seus níveis para estimar um modelo de regressão. E se as variáveis independentes são quantitativas e os números da matriz variam grandemente em valor absoluto, a codificação pode diminuir os erros nas estimativas dos parâmetros⁶. Para k variáveis, um planejamento completo de dois níveis requer $2 \times 2 \times 2 \times \dots \times 2 = 2^k$ ensaios diferentes, sendo conhecido por planejamento fatorial 2^k . Por exemplo, quando se tem três variáveis para serem testadas em dois níveis, são realizados $2^3 = 8$ experimentos⁸.

Após a realização da triagem, é necessário em muitos casos uma segunda etapa que permite identificar com maior precisão a região mais adequada para a realização dos experimentos. Assim, um ajuste fino é realizado com o objetivo de identificar uma condição de compromisso, sendo calculados modelos de regressão obtidos a partir de planejamentos do tipo Doehlert, Box-Behnken e composto central. Nesses planejamentos, o número de experimentos não difere muito do planejamento completo e fracionário, pois as variáveis diminuem, porém, o número de níveis aumenta. Um outro aspecto, é que diferentemente da triagem, tem-se um conhecimento maior do sistema em estudo⁶.

O planejamento composto central combina um planejamento fatorial completo ou fracionário de dois níveis com pontos adicionais (pontos axiais ou estrela) e ponto central com cerca de 20% das réplicas⁶. Com o planejamento do tipo Doehlert é possível efetuar uma otimização com as variáveis estudadas em diferentes quantidades de níveis. Assim, a variável que apresentou maior efeito ou a de maior interesse pelo experimentador, pode ser estudada em mais níveis e com um maior refinamento⁹. Já o planejamento fatorial Box-Behnken é um planejamento fatorial incompleto onde os níveis são variados entre -1 e 1. Além disso, as condições extremas (1 e 1 ou -1 e -1) não são avaliadas, não sendo indicado para situações onde o objetivo é avaliar as respostas nos extremos⁷. Assim, a escolha de

um ou outro tipo de planejamento vai depender da necessidade de cada experimentador.

Com os resultados do planejamento escolhido, é possível efetuar uma análise da variância (*Analysis of Variance, ANOVA*). Todos os cálculos podem ser efetuados em programas computacionais como Matlab, Octave (programa livre) e Excel. Através dos modelos de regressão podem ser obtidos gráficos de superfície, onde é visualizada a região de máximo sinal analítico e gráficos de contorno, os quais mostram a visão superior da superfície de resposta⁶.

Um exemplo que utiliza a LIBS juntamente com planejamento fatorial foi apresentado por NUNES *et al.*¹⁰ onde os autores utilizaram planejamento do tipo Doehlert para otimizar os fatores que afetam o desempenho do equipamento LIBS. Os parâmetros investigados foram tempo de integração, número de pulsos de laser acumulados, *delay time* e ganho de amplificação. Os autores concluíram que a utilização do procedimento de otimização é uma ferramenta adequada e após esse processo foi possível determinar macro e micronutrientes, Al e Si em amostras de plantas.

Em um trabalho realizado por ANDRADE *et al.*¹¹, os autores determinaram contaminantes, micro e macronutrientes em amostras de fertilizantes sólidos utilizando a LIBS. Assim, planejamento fatorial foi utilizado para selecionar as melhores condições dos parâmetros da LIBS e a melhor pressão da pastilha na preparação dos fertilizantes sólidos. Primeiramente, foi realizado um planejamento fatorial fracionário 2^{4-1} variando a energia, *delay time*, *spot size* e a pressão da pastilha. E em uma segunda etapa realizou-se um planejamento composto central variando apenas a energia e o *delay time*, mantendo o *spot size* e a pressão da pastilha fixos. A função de desejabilidade foi utilizada para avaliar as variáveis e com os dois planejamentos foi possível identificar uma condição otimizada aplicável a todas as amostras.

Um outro trabalho de ANDRADE *et al.*¹², os autores determinaram metais essenciais e tóxicos em 18 amostras de ervas medicinais por LIBS e espectrometria de emissão óptica com plasma indutivamente acoplado (*Inductively coupled plasma optical emission spectrometry, ICP OES*). Os parâmetros da LIBS (energia, *delay time* e *spot size*) foram avaliados com o uso do planejamento do tipo Doehlert com uso da função de desejabilidade. Assim, foi possível obter uma

condição de compromisso para todos os tipos das amostras com um *delay time* de 1 μ s e energia de 70 mJ. O *spot size* foi fixado em 100 μ m.

2.2- Métodos Quimiométricos

A identificação de semelhanças e diferenças entre vários tipos de amostras pode ser realizada por métodos quimiométricos. Estes métodos podem ser divididos em:

- Métodos supervisionados: a análise de cada amostra é atribuída a uma classe preestabelecida, a qual é utilizada durante os cálculos dos modelos de classificação.
- Métodos não supervisionados de reconhecimento de padrões: não necessita de nenhum conhecimento prévio a respeito da classificação das amostras, assim as mesmas serão agrupadas naturalmente conforme a informação contida nos dados experimentais¹³.

Entretanto, antes de realizar qualquer tipo de análise nos dados é importante efetuar um pré-tratamento nos mesmos.

2.2.1- Pré-tratamento dos dados

O pré-tratamento dos dados é de grande importância, pois visa a redução de variações indesejáveis, as quais não foram eliminadas durante os experimentos e que afetarão análises futuras. Os tipos de variações indesejáveis mais comuns são relacionados com flutuações da linha de base, ruído experimental e desalinhamento de sinais. Assim, após o pré-tratamento a matriz de dados se torna mais adequada para uma análise quimiométrica. Antes de qualquer pré-tratamento, é necessário organizar os dados em uma matriz contendo linhas e colunas, constituindo uma matriz \mathbf{X} , onde cada linha corresponde a uma amostra i e cada coluna se refere a uma variável j . Assim, uma matriz de dados é dada por \mathbf{X} ($I \times J$), com um total de I linhas (amostras) e J colunas (variáveis)¹³.

A normalização dos dados é um tipo de pré-processamento. Há vários tipos de normalização, por exemplo: normalização pela área total, pela norma, por padrão interno. Assim, nessa transformação os valores de cada uma das variáveis de uma dada amostra i são divididos por um fator de normalização¹³⁻¹⁵.

A normalização pela área total tem por objetivo minimizar o efeito da intensidade total de perfis de resposta, pela variabilidade da amostra e do caminho óptico. Neste caso, o fator de normalização é a soma dos valores absolutos $\left(\sum_{j=1}^n |x_{i,j}|\right)$ de todas as variáveis para uma dada amostra, como é mostrado na Equação 1¹⁶.

$$X_{i,norm} = \frac{x_i}{\sum_{j=1}^n |x_{i,j}|} \quad (1)$$

A normalização pela norma é utilizada na remoção de variações sistemáticas relacionadas com o tamanho da amostra ou quando a concentração absoluta dos analitos nas amostras é incontrolável. Assim, iguala-se a magnitude de cada amostra. Neste caso, o fator de normalização pode ser a norma Euclidiana $\|x_i\|_2$, mostrada na Equação 2¹³.

$$\|x_i\|_2 = \sqrt{\sum_{j=1}^J x_{ij}^2} \quad (2)$$

A padronização interna é um procedimento também utilizado para minimizar a flutuação dos resultados analíticos. No caso da LIBS, o fator de normalização é a intensidade de alguma linha de emissão de um elemento cuja concentração é constante em todas as amostras. Além disso, o sinal do padrão interno deve flutuar da mesma forma que o analito de interesse. Assim, a utilização de padrão interno pode minimizar efeito de matriz e melhorar a qualidade dos modelos de calibração¹⁷.

Em ZAYTSEV *et al.*¹⁷, os autores compararam duas estratégias de calibração (uni e multivariada) para a determinação de Cr, Mn, Ni e Si em aços inoxidáveis de alta liga. Na calibração univariada utilizaram o sinal do Fe como padrão interno, onde obtiveram melhores resultados do que a calibração multivariada utilizando regressão por componentes principais (*Principal Component Regression, PCR*).

Em um trabalho de SARKAR *et al.*¹⁸ os autores estudaram 11 modos de normalização (intensidade máxima, intensidade mínima, norma, Fe como padrão interno, soma do espectro total, entre outros) e a partir do erro padrão de previsão

(*Standard error of prediction, SEP*) e o erro relativo de previsão (*Relative error of prediction, REP*), concluíram que o melhor modo de normalização é aquele que utiliza a norma dos espectros.

Há também outros tipos de pré-processamento, como centrar os dados na média e o autoescalamento, os quais são os mais usados quando se faz análise exploratória dos dados e modelos de calibração.

Ao centrar os dados na média, ocorre uma translação de eixos para o valor médio de cada um deles sem perder a estrutura dos dados. Assim, calcula-se o valor médio de cada coluna (\bar{x}_j), o qual é subtraído de cada um dos valores da respectiva coluna (x_{ij}), como é mostrado na Equação 3.

$$x_{ij(cm)} = x_{ij} - \bar{x}_j \quad (3)$$

Este tipo de pré-processamento é utilizado frequentemente em dados espectrais, para que o intercepto do modelo seja eliminado e para que a direção de maior variância seja coincidente com o de maior variância interna dos dados¹³. Já o autoescalamento, além de calcular o valor médio de cada coluna e subtrair-lo de cada um dos valores da respectiva coluna, divide-se esse resultado pelo desvio-padrão dessa coluna (s_j). Portanto, além de centrar os dados na média, os dados são autoescalados pelo desvio-padrão, mostrado na Equação 4.

$$x_{ij(a)} = \frac{x_{ij} - \bar{x}_j}{s_j} \quad (4)$$

Assim, a média e o desvio padrão dos dados autoescalados será 0 e 1, respectivamente. Este tipo de pré-processamento é empregado quando as variáveis são discretas (dados de concentração, por exemplo) ou quando as variáveis possuem unidades diferentes, assim iguala-se o impacto de cada uma delas¹³.

2.2.2- Modelos de regressão

É possível obter uma expressão matemática, $y=f(x)$, que determina uma propriedade de interesse, em função de seus espectros. Desta forma, o

esperado é que variações em uma determinada propriedade esteja relacionada com alterações nas informações espectrais. Esta é a principal hipótese desta dissertação de mestrado. Essa propriedade pode ser desde a concentração de um elemento químico até a propriedade física de uma liga metálica. Assim, a matriz de dados originais é configurada por duas informações importantes. A primeira delas é chamada de medidas instrumentais, denominadas de variáveis independentes (matriz **X**). A outra parte deste conjunto de dados é denominado de variáveis dependentes (matriz **Y**) e possui os valores de alguma propriedade de interesse ou concentrações de analitos¹³.

A construção de modelos de regressão pode economizar tempo e o custo de análise de rotina, principalmente em indústrias, pois em muitos casos é mais conveniente, rápido e econômico obter um espectro de uma amostra do que determinar diretamente a propriedade desejada¹³. Essas características são especialmente vantajosas no caso de medidas em linha efetuadas diretamente no setor de produção de uma indústria.

No cálculo de modelos é utilizado um conjunto de calibração que deve ser representativo, ou seja, conter todas as fontes de variação que as amostras de previsão (ou conjunto de validação) possam ter, desde aos efeitos espectrais quanto às variações causadas por interferências. Além disso, é necessário que as amostras do conjunto de calibração estejam distribuídas uniformemente ao longo de todo o intervalo, apresentando assim uma boa previsão¹³.

Alguns métodos para calcular os modelos de regressão multivariada são:

- Regressão por quadrados mínimos – clássico (*Classical Least Squares, CLS*);
- Regressão linear múltipla (*Multiple linear regression, MLR*);
- PCR;
- Mínimos quadrados parciais (*Partial least squares, PLS*).

Um dos métodos mais utilizados é o PLS, construído em uma única etapa, onde as informações da matriz **X** e da propriedade de interesse são consideradas durante a decomposição e a compressão dos dados. Ao construir um modelo é necessário escolher o número de fatores, o que pode ser feito através da validação cruzada (*cross validation*). A validação cruzada é realizada da seguinte maneira: é excluída uma (*leave-one-out*) ou mais amostras *i* do conjunto de calibração, e em seguida o modelo é calculado sem essas amostras. O modelo

obtido é utilizado para calcular a propriedade de interesse das amostras excluídas e em seguida realiza-se uma comparação entre os valores estimados (previstos, \hat{y}_i) da propriedade de interesse das amostras excluídas com os valores reais (valores de referência, y_i) calculando-se os resíduos (Equação 5). Estes resíduos são expressos pelo valor da soma dos quadrados dos erros de previsão efetuados pela validação cruzada (*Predicted Residual Error Sum of Squares*, $PRESS_{cv}$) ou pelo erro padrão de validação cruzada (*Standard Error of Cross-validation*, SECV). O número ótimo de fatores ou de variáveis latentes será aquele que produz um menor valor de $PRESS_{cv}$ ou SECV, mostrados pela Equação 6 e 7, respectivamente. Além disso, o SECV pode ser utilizado para avaliar a capacidade do modelo de fazer previsão¹³.

$$\text{Resíduo} = y_i - \hat{y}_i \quad (5)$$

$$PRESS_{cv} = \sum_{i=1}^l (y_i - \hat{y}_i)^2 \quad (6)$$

$$SECV = \sqrt{\frac{PRESS_{cv}}{n}} \quad (7)$$

Onde y_i é o valor real ou de referência, \hat{y}_i é o valor previsto e n é o número de amostras utilizadas na validação interna cruzada.

Para finalizar a etapa de validação do modelo é necessário calcular os parâmetros de qualidade (*figures of merit*) do modelo proposto como: sensibilidade, sensibilidade analítica, seletividade, limites de detecção e de quantificação e desvio padrão relativo. Após isso o modelo está pronto para fazer previsões da propriedade de interesse em novas amostras, as quais devem ser preparadas ou obtidas de modo semelhante as do conjunto de calibração. Os dados experimentais devem ser coletados usando os mesmos procedimentos e instrumentos e pré-processados da mesma maneira^{13,19}.

Em um trabalho de ANDRADE *et al.*²⁰, considerou-se a comparação da ferramenta quimiométrica PLS com o padrão tradicional de curvas de calibração e determinação de suas vantagens e limitações. Na parte experimental, fez-se então uma análise de LIBS em dez amostras de latão, onde nove eram de composição conhecida e uma desconhecida. Assim como era esperado, PLS é substancialmente mais rápido que a abordagem clássica da curva de calibração, não requer o cálculo

das intensidades de linhas espectrais, pois foi utilizado todo o perfil espectral e pode lidar com uma quantidade substancial de dados em poucos segundos. Além disso, foi confirmado então a possibilidade de utilizar PLS para prever com precisão, pelo menos, a concentração dos elementos majoritários da liga, mesmo quando a resolução espectral é pequena.

2.2.3- Modelos de classificação

Os métodos de classificação podem ser agrupados em duas categorias, paramétricos ou probabilísticos e não paramétricos ou não probabilísticos (determinísticos). A categoria dos probabilísticos considera que as variáveis satisfazem a uma distribuição normal, utilizando a homogeneidade da matriz de variância-covariância das classes para assumir e posicionar o hiperplano ou a região limitada de classificação. Assim assumem o conhecimento da função de densidade de probabilidade de cada classe. Os métodos SIMCA (*Soft Independent Modeling of Class Analogy, SIMCA*) e análise discriminante linear (*Linear Discriminant Analysis, LDA*) são exemplos de probabilísticos. Já a categoria dos determinísticos geram discriminantes baseados na matriz dos dados X e no conhecimento prévio da classe de cada amostra, sem considerar qualquer informação a respeito das medidas estatísticas e suas distribuições, por exemplo, o método dos K-ésimos Vizinhos mais próximos (*K-Nearest Neighbor, KNN*)¹³.

No KNN é necessário definir o conjunto de treinamento, onde cada amostra é excluída uma única vez e então classificada utilizando as amostras restantes. Assim, é calculada a distância (Euclideana ou de Mahalanobis) entre a amostra excluída e todas as outras amostras do conjunto de treinamento, sendo que a amostra excluída é classificada de acordo com a maioria dos “votos” de seus vizinhos mais próximos. Cada vizinho dá um voto para a sua própria classe e assim a amostra é atribuída a classe mais votada. Em caso de empate, é escolhida a classe com a menor distância acumulada. Se coincidir com a classe verdadeira, a classificação foi bem-sucedida¹³.

O método SIMCA assume que as amostras parecidas tenderão para uma distribuição uniforme e modelável, onde um modelo de componentes principais (*Principal Component, PC*) é ajustado a cada classe do conjunto de treinamento, dando origem a um classificador para cada uma delas. É definido os limites de cada

classe no subespaço e construído uma hipercaixa no subespaço complementar ao das PCs. Assim, a classificação de uma amostra é feita comparando a variação nos seus dados que não foi explicada pelo subespaço das PCs com a variância residual média da classe em questão, onde a quantificação da comparação é feita por meio de um teste F. Se as classes não estiverem bem separadas, a amostra pode ser classificada em duas ou mais classes. E se uma amostra for distinta de todas as classes, a mesma não é classificada em nenhuma delas¹³.

Um outro exemplo de métodos de classificação é a análise discriminante pelo método de mínimos quadrados parciais (*Partial Least Squares-Discriminant Analysis, PLS-DA*), o qual utiliza o método de regressão PLS. É necessário definir o bloco das variáveis dependentes, o qual será codificado de acordo com a classe da amostra. Assim os valores estimados pelo modelo final serão valores reais próximos dos números originais. Por exemplo, quando os valores reais se aproximam de 1 e 0, estabelece um valor de corte entre 0 e 1 para delimitar as classes. Se o valor estimado estiver acima do valor de corte, a amostra pertence a classe em questão. Caso o valor estimado estiver abaixo do valor de corte, a amostra não pertence a classe que está sendo modelada¹³.

Um trabalho realizado por VORS *et al.*²¹ os autores utilizaram LIBS e modelo SIMCA para identificar ligas metálicas. As medidas foram realizadas com as mesmas amostras e o mesmo equipamento em duas datas diferentes para avaliar a robustez do modelo. Para a otimização dos parâmetros do modelo, planejamento fatorial completo foi utilizado, onde mostrou-se ser uma abordagem que permite a realização de um modelo de identificação multivariada muito robusto ao longo de um tempo bastante longo.

2.3- Aplicação da *Laser-induced breakdown spectroscopy* na análise de ligas metálicas

A LIBS é uma técnica analítica moderna que emprega um pulso curto de radiação de alta energia e uma lente para focalizar o mesmo sobre a superfície da amostra, gerando um plasma, o qual vaporiza uma pequena quantidade de amostra. A luz do plasma é coletada por uma fibra óptica e direcionada ao espectrômetro, o qual dispersa a luz emitida por átomos, íons e moléculas simples

excitados. Por fim, um detector registra os sinais de emissão. A FIGURA 2.3.1 mostra um esquema dos processos descritos neste parágrafo.

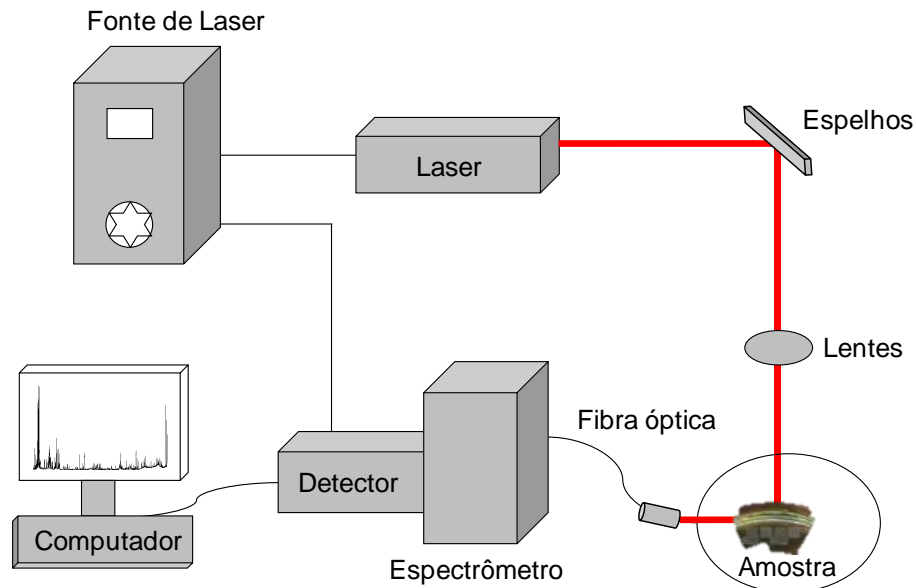


FIGURA 2.3.1- Esquema ilustrativo dos principais componentes da técnica LIBS. O sistema ilustrado é composto por fonte de laser, laser, espelhos, lentes, Espectrômetro, detector e computador.

As vantagens desta técnica em relação as outras, são: a análise direta e rápida de qualquer tipo de amostra (com pouco preparo ou sem nenhum preparo das amostras), o menor consumo da amostra (é uma técnica quase não destrutiva e que não gera resíduos) e a possibilidade de micro-análise (análise de micro-regiões da amostra, permitindo investigar superfícies sólidas com resolução espacial)²². Além disso, a LIBS pode ser empregada em aplicações industriais, onde vem se tornando uma técnica promissora, devido a determinação de vários elementos simultaneamente em tempo real, sem preparação da amostra²³. Entretanto, ainda apresenta alguns desafios como altos limites de detecção, reduzida precisão e carência de materiais calibrantes para massas da ordem de μg ²².

As medidas na LIBS podem ser influenciadas pelo comprimento de onda do laser, condições ambientais (gás e pressão), propriedades das amostras (mecânicas, físicas e químicas), taxa de repetição do laser, energia e *spot size* (fluência) e janela de detecção (*delay time* e tempo de integração)²⁴.

O aumento da energia do laser influencia na taxa de ablação e intensidade de emissão, se a duração do pulso e *spot size* são os mesmos. Assim, a

energia pode ser dada na forma de irradiância (W/cm^2), a qual depende da energia do pulso do laser, da largura do pulso e da área do pulso (*spot size*). E também na forma de fluência (J/cm^2), a qual é a energia por unidade de área²⁴.

Uma outra variável importante, é o *delay time*, atraso do início do laser até a abertura da janela (onde o sinal será medido). Portanto, é o tempo necessário para reduzir a emissão do contínuo, onde em seguida, irá iniciar a emissão iônica e depois atômica. A emissão do contínuo se dá pelo processo Bremsstrahlung (colisões de elétrons com íons e átomos) e pela recombinação de elétrons com íons²⁴.

No trabalho de STIPE *et al.*²³, os autores compararam três métodos de calibração para prever concentrações de Cr, Mn e Ni em 23 amostras de aços de alta e baixa liga. Os autores utilizaram a normalização por somatória do espectro total e normalização por um determinado sinal de Fe como padrão interno utilizando calibração univariada e por último PLS para gerar um modelo de regressão. Os sinais dos analitos foram as áreas das linhas de emissão selecionadas. A partir dos resultados, os autores concluíram que a normalização por somatória do espectro total depende fortemente da faixa espectral integrada, e assim os modelos gerados com esse tipo de normalização apresentaram altos erros de previsão. Entre o método de PLS e normalização por padrão interno, PLS mostrou ser uma boa alternativa, pois pode ser aplicado quando se estuda amostras completamente desconhecidas, além de identificar sinais de emissão importantes para a calibração.

Em trabalho publicado por GONZAGA e PASQUINI²⁵, foram determinados Cr e Ni em aço utilizando um equipamento LIBS compacto que possui baixo custo. Além disso, o equipamento apresentou tempos de integração longos e os autores empregaram calibração multivariada, para compensar a baixa energia de pulso da fonte de laser, a sensibilidade e a resolução mais pobre do espectrômetro usado. Os espectros obtidos foram derivados e aplicados para a construção de dois modelos preliminares de calibração (para Cr e Ni) usando PLS. Após esse processo de ajuste dos modelos de calibração, os mesmos foram validados utilizando validação cruzada. Este trabalho apresentou êxito para a determinação simultânea dos dois analitos, usando os novos modelos de analisador LIBS e calibração PLS com pré-seleção de variáveis. Os erros relativos médios foram da ordem de 4 a 7% para Cr e Ni, respectivamente.

Um exemplo da aplicação de LIBS em altos fornos, foi proposto por NOLL *et al.*²⁶, onde os autores aplicaram a LIBS na siderurgia e processamento industrial. As substâncias analisadas variaram de aço líquido em alto-forno até produtos finalizados. A utilização da LIBS melhorou a sensibilidade e limites de detecção. Os autores determinaram elementos leves e pesados no aço líquido e o controle de processo na produção baseou-se nos resultados da análise química da escória. Nestas determinações foi utilizada a LIBS que apresentou uma frequência analítica duas vezes maior quando comparada com métodos convencionais. O sistema de microanálise de varredura foi desenvolvido com frequências de até 1 kHz e uma resolução espacial < 15µm.

Outros trabalhos que utilizaram como técnica a LIBS e amostras de ligas metálicas podem ser encontrados em GUPTA *et al.*²⁷, ORTIZ *et al.*²⁸, HAO *et al.*²⁹ e ZHANG *et al.*³⁰.

3- Objetivo e justificativa

A análise de ligas e metais é efetuada em muitos casos com a técnica de Fluorescência de Raios-X (*X-Ray fluorescence, XRF*). Essa técnica é bastante consolidada no meio industrial e suas vantagens para análise rápida de amostras sólidas é inegável e inquestionável. Com a XRF é possível efetuar, por exemplo, determinações sem padrão ao utilizar as ferramentas de parâmetros fundamentais³¹. Entretanto, a XRF tem algumas limitações, como a determinação de elementos leves (aqueles com número atômico menores que o do Na) e dificuldades de determinação em amostras com concentrações inferiores a 1% em massa. Uma outra limitação em relação a ligas metálicas é a espessura do material. Se o material de interesse encontra-se deteriorado, haverá uma composição pouco significativa do restante do material. Ao remover a superfície deteriorada são inseridas algumas operações unitárias que comprometem vantagens relacionadas com o emprego da técnica XRF, a qual é a baixa manipulação das amostras.

Assim, a proposta dessa dissertação foi apresentar a LIBS como uma alternativa viável para a análise direta de ligas e metais, possibilitando assim a obtenção de uma ampla gama de informações importantes. Essas informações foram avaliadas com o uso de ferramentas quimiométricas que possibilitaram um melhor entendimento e aproveitamento dos dados gerados. Além disso, a combinação de LIBS e quimiometria permitiu a rápida obtenção de resultados.

4- Materiais e métodos

4.1- Amostras

Várias ligas metálicas como aços, latão, ligas de titânio, bronze, alumínio e hastelloy foram utilizadas na proposição de modelos de regressão e classificação. Algumas amostras foram gentilmente cedidas pelo Departamento de Engenharia de Materiais (DEMA) da UFSCar e outras foram adquiridas no comércio local da cidade de São Carlos - SP. Para as amostras 40, 42, 47 e 54, tem-se duas réplicas. Já para as amostras 41, 43, 44, 45, 46, 48, 49, 50, 51, 52 e 53, tem-se três réplicas. Assim, no total obteve-se 80 amostras, contando com as duplicatas e triplicatas. De forma geral as dimensões das amostras variaram de 0,5 (amostra 41) a 10,5 (amostra 42) cm para a largura e 0,2 (amostra 42) e 10 (amostra 32) cm para o comprimento. As TABELAS 4.1.1 a 4.1.3 mostram uma visão geral das amostras utilizadas, bem como uma breve descrição de cada uma delas. É importante destacar que as dimensões das amostras são bem distintas, sendo que algumas delas apresentaram formatos bastante irregulares. Apesar dessas observações, todas as amostras puderam ser acomodadas no porta-amostra do equipamento LIBS utilizado. Para as medidas por LIBS as amostras não passaram por nenhum tratamento prévio. O único cuidado tomado foi a efetuação de pulsos de laser preliminares para limpeza da superfície.

TABELA 4.1.1- Identificação (ID), descrição e imagem das amostras 1 a 20.






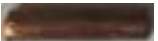














ID da Amostra	Descrição	Imagem	ID da Amostra	Descrição	Imagem
1	Aço bicromatizado		11	Amostra desconhecida	
2	Liga desconhecida		12	Amostra desconhecida	
3	Aço Carbono		13	Amostra desconhecida	
4	Aço Inox 304		14	Amostra desconhecida	
5	Aço inoxidável duplex 2205		15	Amostra desconhecida	
6	Alumínio AA7050		16	Amostra desconhecida	
7	Latão		17	Amostra desconhecida	
8	Liga de Ti 6Al 4V		18	Amostra desconhecida	
9	Liga de Ti puro		19	Amostra desconhecida	
10	Hastelloy c276		20	Aço Prata W	

TABELA 4.1.2 - Identificação (ID), descrição e imagem das amostras 21 a 38.





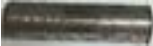






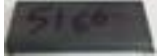






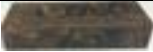
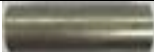














ID da Amostra	Descrição	Imagem	ID da Amostra	Descrição	Imagem
21	Latão		30	Cobre eletrolítico	
22	Bronze 620		31	Aço VC 131	
23	Aço prata carbono		32	Aço Inox 304	
24	Bronze TM 23		33	Aço Inox 420	
25	Aço P20		34	Aço 1045 retificado	
26	Aço VND		35	Aço 5160	
27	Bronze grafite		36	Aço VF 800	
28	Aço H13		37	Aço 52100	
29	Aço D2		38	Liga de Al 6351	

TABELA 4.1.3- Identificação (ID), descrição e imagem das amostras 39 a 54.

ID da Amostra	Descrição	Imagem	ID da Amostra	Descrição	Imagem
39	Aço VW3		47	Aço Inox F138	
40	Aço Inox F138		48	Aço Inox F138	
41	Aço Inox F138		49	Aço Inox F138	
42	Aço Inox F138		50	Aço Inox F138	
43	Liga de Ti puro		51	Liga de Ti 6Al 4V	
44	Liga Co_Cr_Mo		52	Aço Inox F138	
45	Aço Inox F138		53	Aço Inox F138	
46	Aço Inox F138		54	Aço Inox F138	

4.2- Instrumentação

Para a análise direta das ligas metálicas foi utilizada a técnica LIBS, cujo equipamento comercial é o modelo J200 da Applied Spectra (Fremont, Califórnia, EUA), o qual possui um laser de Nd:YAG que emite um pulso de laser em 1064 nm. O detector é um dispositivo de carga acoplada (*Charge Coupled Device*, CCD) e a duração do pulso é em torno de 10 ns. O espectrômetro tem 6 canais, com uma resolução de 0,08 – 0,1 nm dos canais 1 (186,940 a 311,195 nm) ao 4 (591,426 a 693,690 nm) e resolução de 0,11 – 0,14 nm no canal 5 (693,807 a 884,407 nm) e 6 (884,509 a 1042,026 nm). Em relação as variáveis instrumentais do equipamento, temos que a energia máxima é de 100 mJ, o tempo de aquisição dos sinais é fixo em 1,05 ms, o *delay time* pode ser variado de 0 a 2 μ s e o *spot size* de 50 a 250 μ m, estes parâmetros foram avaliados durante a execução desta dissertação. O equipamento LIBS utilizado está mostrado na FIGURA 4.2.1.



FIGURA 4.2.1- Visão geral do equipamento LIBS J200 da Applied Spectra. A maçaneta frontal dá acesso ao porta-amostra.

No estabelecimento dos modelos de calibração e de classificação foi necessário efetuar uma caracterização prévia das amostras (TABELAS 4.1.1 a 4.1.3). Nesta caracterização foram obtidos valores de referência (concentrações) para os seguintes elementos: Al, Cr, Cu, Fe, Mn, Mo, Ni, Ti, V e Zn. Nesta etapa foram empregados dois equipamentos baseados na técnica XRF: Innov-X, modelo alfa 6500 (FIGURA 4.2.2) e Niton XL3T 985 ULTRA da Thermo Scientific (FIGURA

4.2.3). Essa técnica foi escolhida por ser a mais semelhante à LIBS, devido a não necessidade do preparo ou digestão da amostra e também por refletir de forma mais fidedigna o processo de análise utilizado na área metalúrgica.

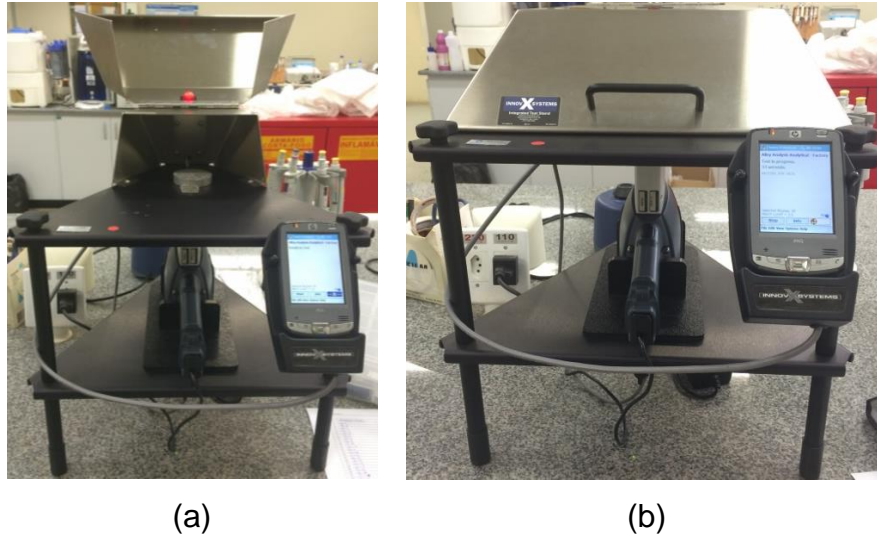


FIGURA 4.2.2- Equipamento XRF Innov-X com o compartimento da amostra (a) aberto e (b) fechado.



FIGURA 4.2.3- Equipamento XRF Niton XL3T da Thermo Scientific.

4.3- Otimização das condições do equipamento LIBS utilizado

Uma das principais etapas dos trabalhos dessa dissertação foi ajustar as condições operacionais do equipamento objetivando simultaneamente a obtenção de sinais com boa qualidade analítica para os 10 analitos em estudo. Sinais adequados seriam aqueles que apresentassem altas magnitudes (altura e área), alta

relação sinal radiação de fundo (*signal-to-background ratio*, SBR), alta relação sinal ruído (*signal-to-noise ratio*, SNR) e que não saturassem o detector. Neste sentido foi realizado um planejamento fatorial do tipo Doehlert com três variáveis: *delay time*, energia do laser e *spot size*, como é mostrado na TABELA 4.3.1. A primeira variável (*delay time*) foi testada em 5 níveis de 0 (codificado em -1) a 2 μs (codificado em 1). A energia do laser foi testada em 7 níveis de 30 (codificado em -0,866) a 80 mJ (codificado em 0,866), sendo testada em mais níveis, pois desempenha um papel importante na intensidade do sinal. Além disso, a energia apresenta uma influência nos valores de irradiância e fluência. Já o *spot size*, que também compromete a irradiância e fluência, foi testado em 3 níveis de 50 a 150 μm . Os três primeiros experimentos na TABELA 4.3.1 são o ponto central e foram utilizados para estimar o erro puro (PE) dos modelos de regressão propostos na otimização.

TABELA 4.3.1. Planejamento fatorial do tipo Doehlert.

Experimentos	<i>Delay time</i> (μs)		Energia (mJ)		<i>Spot size</i> (μm)	
	Codificado	Real	Codificado	Real	Codificado	Real
1	0	1	0	55	0	100
2	0	1	0	55	0	100
3	0	1	0	55	0	100
4	1	2	0	55	0	100
5	0,5	1,5	0,866	80	0	100
6	0,5	1,5	0,289	63	0,817	150
7	-1	0	0	55	0	100
8	-0,5	0,5	-0,866	30	0	100
9	-0,5	0,5	-0,289	47	-0,817	50
10	0,5	1,5	-0,866	30	0	100
11	0,5	1,5	-0,289	47	-0,817	50
12	-0,5	0,5	0,866	80	0	100
13	0	1	0,577	72	-0,817	50
14	-0,5	0,5	0,289	63	0,817	150
15	0	1	-0,577	38	0,817	150

Para este planejamento, utilizou-se 9 amostras com características diferentes: 1 (Zn), 2 (Cu), 10 (Mo e Ni), 35 (Fe), 38 (Al), 44 (Cr), 45 (Fe), 51 (V) e 54

(Fe e Mn). Detalhes sobre as mesmas, podem ser visualizadas nas TABELAS 4.1.1 a 4.1.3. Assim, aproximadamente 100 espectros foram adquiridos em uma varredura em linha reta sobre a superfície da amostra (*raster*) para cada um dos 15 experimentos da TABELA 4.3.1 totalizando 1500 espectros por amostra (100 espectros por experimento). Antes disso, foi realizado uma limpeza da superfície da amostra com as mesmas condições.

Calculou-se a média dos espectros e selecionou-se as linhas de emissão de Al I 394,400 nm; Cr II 283,563 nm; Cu I 324,754 nm; Fe II 259,940 nm; Mn II 294,920 nm; Mo I 386,411 nm; Ni I 361,939 nm; V I 437,924 nm e Zn 481,053 nm. Inicialmente estas linhas de emissão foram escolhidas, pois são as mais intensas para os 9 analitos em estudo. Para cada uma das linhas de emissão selecionadas calculou-se a área e a altura. Para o Ti, testou-se várias linhas de emissão e não foi possível obter nenhum modelo de regressão com coeficientes significativos ao nível de 95% de confiança para todas elas e no caso do Fe, foram calculados dois modelos adicionais, para abranger toda a diversidade de amostras. Assim, 22 modelos de regressão (11 para o sinal da área e 11 para altura) foram obtidos e 10 coeficientes considerados: b_0 (constante), b_1 , b_2 e b_3 (lineares), b_{11} , b_{22} e b_{33} (quadráticos) e b_{12} , b_{13} e b_{23} (interação). A tabela ANOVA de cada modelo gerado foi avaliada para verificar a significância dos coeficientes calculados. O nível de confiança adotado em todos os cálculos foi o de 95% de confiança. Os dados foram tratados no Matlab 2009 (Matworks, Natick, EUA) e no Excel (organização das matrizes). Com os modelos estabelecidos foram gerados gráficos de superfície e de contorno. No cálculo dos valores de área e altura dos sinais foi empregada a rotina `libs_par` que é apresentada no Anexo 1. Já para os cálculos dos modelos de regressão foi utilizada a rotina `regression2` que faz parte do livro de autoria de PEREIRA-FILHO⁶ e está livremente disponível em [facebook.com/planEdenir](https://www.facebook.com/planEdenir). Na TABELA 4.3.2 tem-se as equações e os parâmetros utilizados para os cálculos da ANOVA. Ao dividir a soma quadrática pelo grau de liberdade, tem-se a média quadrática (MQ) dos parâmetros do modelo.

TABELA 4.3.2. Identificação e equações dos parâmetros da tabela ANOVA.

Parâmetros	Soma Quadrática (SQ)	Graus de liberdade (G.L.)	Observações
Regressão, SQR	$\sum_{i=1}^n (\hat{y}_i - \bar{y})^2$	p-1	p = número de coeficientes
Resíduo, SQres	$\sum_{i=1}^n (y_i - \hat{y}_i)^2$	n-p	n = número de experimentos
Total, SQT	$\sum_{i=1}^n (y_i - \bar{y})^2$	n-1	m = número de experimentos independentes
Erro puro, SQep	$\sum_{i=1}^m \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2$	n-m	\hat{y}_i : resposta prevista (área ou altura)
Falta de ajuste, SQfaj	$\sum_{i=1}^m \sum_{j=1}^{n_i} (\hat{y}_i - \bar{y}_i)^2$	m-p	\bar{y} : média das respostas experimentais (área ou altura)
Varição explicada (R ²)	$\frac{SQR}{SQT}$		
Máxima variação explicável (R ² máximo)	$\frac{SQT - SQep}{SQT}$		

4.4- Modelos de calibração univariado e multivariado

Após estabelecer uma condição operacional de compromisso para os 10 analitos simultaneamente, aplicou-se a mesma nas 80 amostras. Para cada amostra foi efetuada uma varredura com aproximadamente 100 espectros após limpeza da superfície. Com os resultados, 12 diferentes modos de tratamentos foram testados a fim de minimizar as variações do sinal e as diferenças da matriz da amostra:

- 1- Média dos 100 espectros;
- 2- A norma individual (comprimento do vetor) é calculada para cada espectro e os dados são divididos por esse valor. A norma do espectro normalizado é igual a 1. Em seguida é calculada a média dos espectros normalizados (100 espectros normalizados);
- 3- A área individual (soma de todos os sinais) é calculada para cada espectro e os dados são divididos por esse valor. A área do espectro normalizado é 1. Em seguida é calculada a média dos espectros normalizados (100 espectros normalizados);
- 4- O máximo individual (sinal mais alto) é calculado para cada espectro e os dados são divididos por esse valor. A intensidade de sinal para a linha de emissão mais alta é 1. Em seguida é calculada a média dos espectros normalizados (100 espectros normalizados);

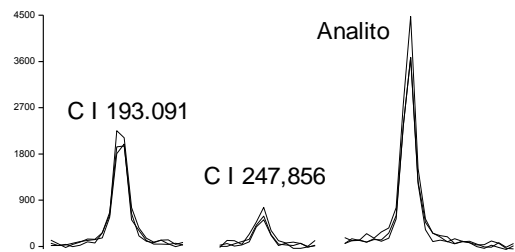
- 5- O espectro individual é dividido pela intensidade de sinal do carbono I 193,091 nm. Em seguida é calculada a média dos espectros normalizados (100 espectros normalizados). A intensidade do sinal resultante para o C I 193,091 nm é 1;
- 6- O espectro individual é dividido pela intensidade de sinal do carbono I 247,856 nm. Em seguida é calculada a média dos espectros normalizados (100 espectros normalizados). A intensidade do sinal resultante para o C I 247,856 nm é 1;
- 7- Soma dos 100 espectros;
- 8- A norma individual (comprimento do vetor) é calculada para cada espectro e os dados são divididos por esse valor. Em seguida é calculada a soma dos espectros normalizados (100 espectros normalizados);
- 9- A área individual (soma de todos os sinais) é calculada para cada espectro e os dados são divididos por esse valor. Em seguida é calculada a soma dos espectros normalizados (100 espectros normalizados);
- 10- O máximo individual (sinal mais alto) é calculado para cada espectro e os dados são divididos por esse valor. Em seguida é calculada a soma dos espectros normalizados (100 espectros normalizados);
- 11- O espectro individual é dividido pela intensidade de sinal do carbono I 193,091 nm. Em seguida é calculada a soma dos espectros normalizados (100 espectros normalizados);
- 12- O espectro individual é dividido pela intensidade de sinal do carbono I 247,856 nm. Em seguida é calculada a soma dos espectros normalizados (100 espectros normalizados).

As imagens ilustrativas dos procedimentos de normalização de 1 a 6 são mostrados na TABELA 4.4.1. Já os procedimentos de 7 a 12 são mostrados na TABELA 4.4.2. O Anexo 2 mostra a rotina computacional `libs-treat` que pode ser utilizada para o cálculo das 12 normalizações.

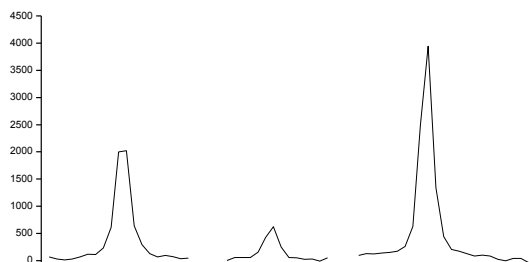
TABELA 4.4.1- Ilustração dos modos de normalização utilizando a média dos espectros normalizados.

Normalização e descrição gráfica

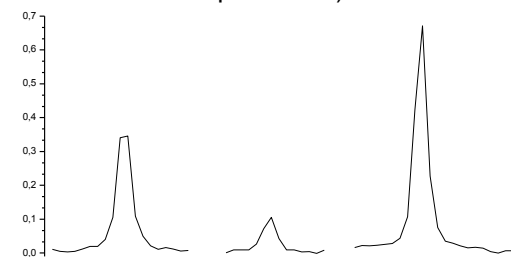
Dados originais com linhas de carbono e um analito hipotético



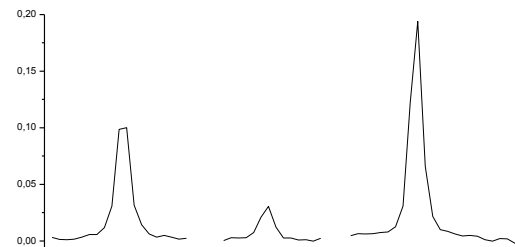
(1) Média dos espectros



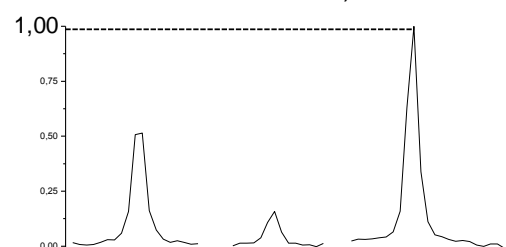
(2) Normalizado pela norma individual e média dos n pulsos (a norma final de cada espectro é 1)



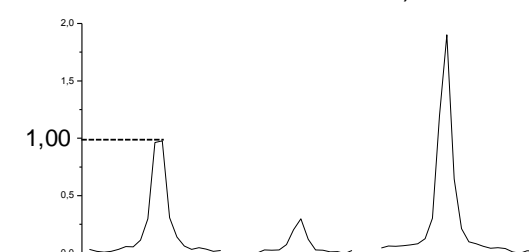
(3) Normalizado pela área individual e média dos n pulsos



(4) Normalizado pelo máximo individual e média dos n pulsos (o sinal da maior linha de emissão é 1)



(5) Normalizado pela intensidade individual do C I 193,091 nm e média dos n pulsos (o sinal de C I 193,091 é 1)



(6) Normalizado pela intensidade individual do C I 247,856 nm e média dos n pulsos (o sinal de C I 247,856 é 1)

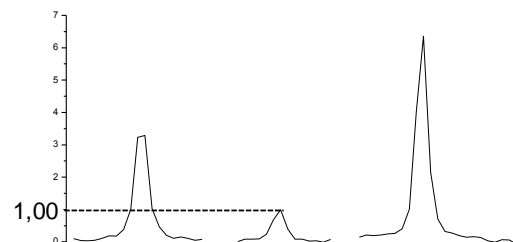
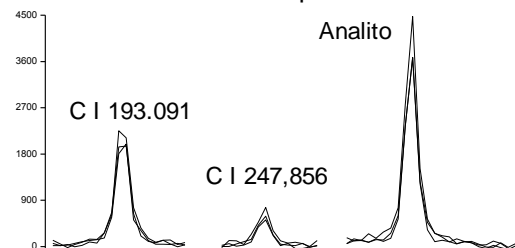


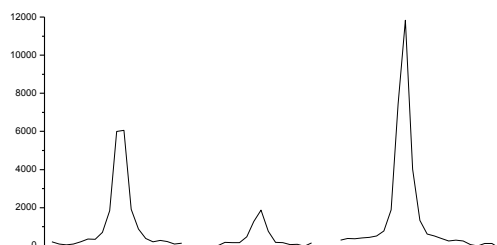
TABELA 4.4.2- Ilustração dos modos de normalização utilizando a soma dos espectros normalizados.

Normalização e descrição gráfica

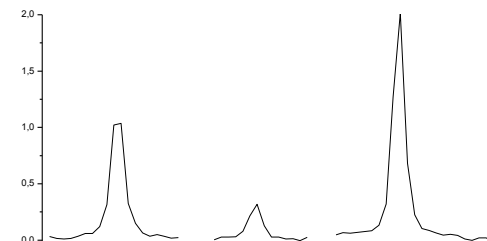
Dados originais com linhas de carbono e um analito hipotético



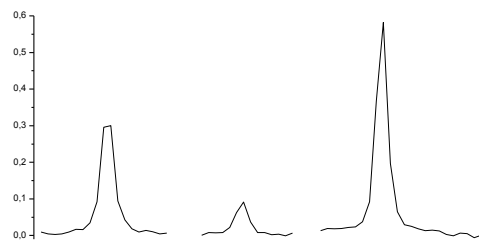
(7) Soma dos espectros



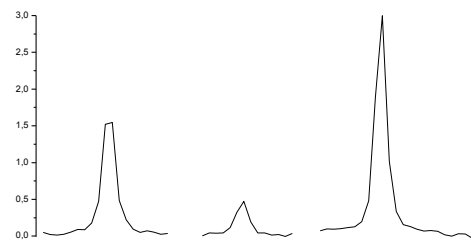
(8) Normalizado pela norma individual e soma dos n pulsos



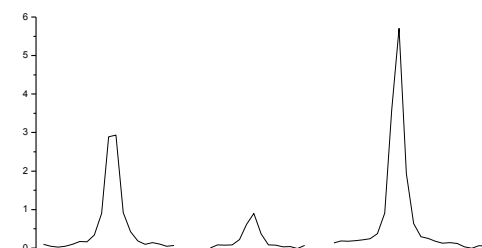
(9) Normalizado pela área individual e soma dos n pulsos



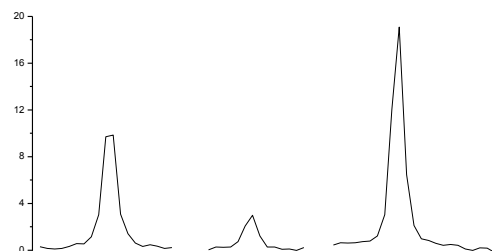
(10) Normalizado pelo máximo individual e soma dos n



(11) Normalizado pela intensidade individual do C I 193,091 nm e soma dos n pulsos



(12) Normalizado pela intensidade individual do C I 247,856 nm e soma dos n pulsos



Assim, calculou-se modelos preliminares de PLS utilizando o perfil total dos sinais e modelos univariados com as linhas mais intensas de cada analito (Al, Cr, Cu, Fe, Mn, Mo, Ni, Ti, V e Zn). Com os vetores de regressão dos modelos PLS gerados, selecionou-se as linhas de emissão dos analitos que mais influenciavam nas previsões das concentrações.

Com o objetivo de avaliar os modelos de calibração gerados, as 80 amostras foram divididas em 64 amostras de calibração e 16 amostras de validação através do algoritmo Kennard-Stone no Matlab³². Assim, foram calculados modelos PLS com validação cruzada, onde os dados foram centrados na média. Os parâmetros observados foram: variáveis latentes, erro padrão de calibração (*Standard Error of calibration, SEC*), SECV e erro padrão de validação (*Standard Error of validation, SEV*), limite de detecção mínimo e máximo. Para os modelos univariados, utilizou-se várias linhas de emissão para cada analito, onde a área e a altura do sinal foram selecionados. Os parâmetros observados foram: SEC, SEV e limite de detecção.

Com os resultados gerados, tais como parâmetros relacionados com os erros e limites de detecção, selecionou-se: o melhor modo de normalização, o modelo mais eficiente (univariado ou multivariado) e a melhor linha de emissão para cada analito. Assim, testes F foram realizados para verificar se havia diferença significativa entre os resultados.

No cálculo do limite de detecção, para a calibração univariada foi empregada a Equação 8:

$$LOD = \frac{3 \times SD_{branco}}{s} \quad (8)$$

Onde o SD_{branco} é calculado a partir das variâncias do *background* da linha de emissão de interesse e s é a sensibilidade da curva analítica obtida. Vale lembrar que foram realizadas 10 medidas na LIBS para o cálculo das variâncias.

Para a calibração multivariada, utilizou-se a nova abordagem da IUPAC, descrita por ALLEGRINI e OLIVIERI³³. Esse cálculo leva em consideração a variância do y (concentrações dos analitos obtidas pela técnica de XRF). Se a concentração do analito está abaixo do LOD_{min} , o analito não está presente e se a concentração está acima do LOD_{max} , o analito está presente. Porém, se a mesma estiver entre o LOD_{min} e LOD_{max} , é necessário calcular o seu LOD específico e

comparar com a concentração. Com esses cálculos, há a possibilidade de obter o erro do tipo I, onde afirma-se que o analito está presente quando ele pode estar ausente e o erro do tipo II, onde afirma-se que o analito está ausente quando ele pode estar presente.

4.5- Modelos de classificação

Para a proposição dos modelos de classificação as amostras foram divididas em 3 classes de acordo com a concentração de Fe:

- (1) Concentração de Fe de 71 a 99% (16 amostras)
- (2) Concentração de Fe de 16 a 67% (33 amostras)
- (3) Concentração de Fe de 0 a 6% (31 amostras).

Essas 3 classes não refletem nenhuma norma ou regulamentação específica e foram estabelecidas apenas para avaliar a capacidade dos espectros de emissão em discriminar amostras de metais. Na avaliação dos modelos de classificação 3 matrizes diferentes foram utilizadas: concentrações dos analitos obtidas pela técnica XRF, sinal da área e altura com todas as linhas de emissão estudadas (57 linhas no total). Assim, calculou-se três modelos: KNN, SIMCA e PLS-DA³⁴. Os seguintes parâmetros dos modelos foram avaliados: exatidão, sensibilidade, taxa de falso alarme e especificidade, os quais estão exemplificados nas equações de 9 a 12.

$$\text{Exatidão} = \frac{NA}{NT} \quad (9)$$

$$\text{Sensibilidade de } i = \frac{\sum NA_y}{\sum NT_y} \quad (10)$$

$$\text{Taxa de falso alarme de } i = \frac{NE_i}{NT_i} \quad (11)$$

$$\text{Especificidade de } i = \frac{NA_i}{NT_i} \quad (12)$$

Onde:

NA = número total de acertos;

NT = número total de amostras;

i = amostras da classe que se deseja calcular o parâmetro;

y = são as amostras das outras classes com exceção de i ;

NAy = todos os acertos de y ;

NTy = número total de amostras de y ;

NEi = número de erros de i ;

NTi = número total de amostras de i ;

NAi = número de acertos em i ;

NTi = número total de amostras de i .

5- Resultados e discussão

5.1. Caracterização das amostras

Os dados de referência dos analitos nas amostras foram adquiridos através da técnica XRF. A XRF é uma técnica não destrutiva que permite identificar os elementos presentes em uma amostra (qualitativa) e também a proporção (concentração) em que se encontram. Uma fonte de radiação de alta energia (radiação X) provoca a excitação dos átomos, promovendo os elétrons a níveis mais energéticos (estado excitado). Os elétrons retornam naturalmente ao seu estado fundamental, ocorrendo uma emissão de energia. Essa energia é característica de cada elemento químico, a qual permite a identificação e correspondente quantificação. Nas TABELAS 5.1.1 a 5.1.3, tem-se as concentrações (% m/m) de todas as amostras, adquiridas por meio da técnica XRF.

Utilizou-se dois equipamentos de XRF, pois o XRF Innov-X não detectou as concentrações de Al nas amostras. Assim, tivemos a oportunidade de testar o equipamento da Thermo Scientific em algumas amostras. Portanto, as amostras que apresentavam Al em suas concentrações, foram analisadas por esse equipamento. Testou-se também em outras amostras que não possuía concentração de Al para verificar a concordância entre os dois equipamentos. Os resultados foram muito próximos de um equipamento para outro. Portanto, as concentrações apresentadas nas TABELAS de 5.1.1 a 5.1.3, possuem resultados de ambos equipamentos. Vale ressaltar, que os dois equipamentos portáteis possuem uma biblioteca interna de dados de ligas metálicas.

TABELA 5.1. 1. Resultados (% m/m) do XRF para as amostras 1 a 27.

Amostras	Co	Cr	Cu	Fe	Ni	Pb	Zn	Mn	Mo	Ti	V	Al
Am 1	0,41	0,21	0,10	16,51	0,00	0,00	82,79	0,00	0,00	0,00	0,00	0,00
Am 2	0,00	0,00	59,63	0,22	0,00	1,44	38,31	0,00	0,00	0,00	0,00	0,00
Am 3	0,00	0,02	0,00	99,69	0,00	0,00	0,00	0,30	0,00	0,00	0,00	0,00
Am 4	0,00	18,19	0,00	71,47	8,02	0,00	0,00	1,48	0,02	0,00	0,00	0,00
Am 5	0,00	22,48	0,00	67,46	5,57	0,00	0,00	1,59	2,40	0,00	0,00	0,00
Am 6	0,00	0,13	1,56	0,12	0,00	0,00	4,23	0,00	0,00	0,10	0,00	88,50
Am 7	0,00	0,00	99,97	0,03	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
Am 8	0,00	0,00	0,00	0,69	0,00	0,00	0,00	0,12	0,00	87,18	4,09	7,56
Am 9	0,00	0,00	0,00	0,33	0,00	0,00	0,00	0,00	0,00	99,68	0,00	0,00
Am 10	0,07	19,47	0,12	5,85	56,39	0,00	0,00	0,27	14,01	0,29	0,42	0,00
Am 11	0,00	17,61	0,20	62,70	14,61	0,00	0,07	1,82	2,85	0,00	0,13	0,00
Am 12	0,00	0,00	52,87	4,00	43,14	0,00	0,00	0,00	0,00	0,00	0,00	0,00
Am 13	0,00	0,00	100,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
Am 14	0,00	17,82	0,08	62,93	14,27	0,00	0,00	1,91	2,89	0,00	0,06	0,00
Am 15	0,00	0,00	0,00	0,21	0,00	0,00	0,00	0,00	0,00	87,44	4,18	7,87
Am 16	0,00	0,00	50,13	5,79	41,68	0,21	0,80	0,00	0,00	0,00	0,00	0,00
Am 17	0,00	0,00	0,00	0,39	0,00	0,00	0,00	0,00	0,00	94,96	4,66	0,00
Am 18	0,00	0,00	0,00	0,18	0,03	0,01	0,00	0,00	0,00	95,14	4,67	0,00
Am 19	0,00	0,00	0,00	0,22	0,00	0,00	0,00	0,00	0,00	99,79	0,00	0,00
Am 20	0,00	1,44	0,00	98,19	0,00	0,00	0,00	0,37	0,02	0,00	0,00	0,00
Am 21	0,00	0,00	61,89	0,35	0,10	3,02	34,39	0,00	0,00	0,00	0,00	0,00
Am 22	0,00	0,00	93,99	0,07	0,20	0,74	1,03	0,06	0,00	0,00	0,00	0,00
Am 23	0,00	0,80	0,04	98,34	0,00	0,00	0,00	0,83	0,00	0,00	0,00	0,00
Am 24	0,00	0,03	80,77	0,51	0,48	6,77	6,19	0,00	0,00	0,05	0,00	0,00
Am 25	0,00	1,97	0,00	95,87	0,00	0,00	0,00	1,42	0,17	0,10	0,00	0,00
Am 26	0,00	0,53	0,16	97,47	0,00	0,00	0,00	1,20	0,08	0,00	0,09	0,00
Am 27	0,00	0,00	87,56	0,06	0,00	0,00	0,00	0,00	0,00	0,06	0,00	0,00

TABELA 5.1. 2. Resultados (% m/m) do XRF para as amostras 28 a 45_2.

Amostras	Co	Cr	Cu	Fe	Ni	Pb	Zn	Mn	Mo	Ti	V	Al
Am 28	0,00	5,24	0,14	89,42	0,26	0,00	0,00	0,39	1,17	0,00	0,86	0,00
Am 29	0,00	11,02	0,10	84,31	0,23	0,00	0,00	0,26	0,60	0,00	0,93	0,00
Am 30	0,00	0,02	99,58	0,34	0,00	0,06	0,00	0,01	0,00	0,00	0,00	0,00
Am 31	0,00	11,05	0,04	87,08	0,27	0,00	0,00	0,35	0,27	0,00	0,17	0,00
Am 32	0,00	18,27	0,28	71,93	8,03	0,00	0,00	1,27	0,12	0,00	0,11	0,00
Am 33	0,00	13,28	0,22	84,53	0,00	0,00	0,00	0,00	0,04	0,00	0,00	0,00
Am 34	0,00	0,47	0,15	95,96	0,57	0,00	0,00	0,94	0,16	0,00	0,00	0,00
Am 35	0,00	0,90	0,06	96,53	0,00	0,00	0,00	0,82	0,00	0,00	0,00	0,00
Am 36	0,00	8,40	0,23	87,61	0,24	0,00	0,00	0,53	2,29	0,00	0,51	0,00
Am 37	0,00	1,35	0,27	96,25	0,13	0,00	0,08	0,37	0,02	0,00	0,00	0,00
Am 38	0,00	0,14	0,03	0,22	0,00	0,00	0,00	0,44	0,00	0,02	0,00	97,96
Am 39	0,00	0,99	0,58	91,60	0,23	0,00	0,10	0,28	0,06	0,00	0,15	0,00
Am 40_1	0,00	17,26	0,19	60,56	13,63	0,00	3,75	1,75	2,86	0,00	0,00	0,00
Am 40_2	0,00	17,35	0,00	60,67	13,81	0,00	3,73	1,61	2,83	0,00	0,00	0,00
Am 41_1	0,00	0,00	0,00	0,41	0,00	0,00	2,64	0,00	0,00	96,96	0,00	0,00
Am 41_2	0,00	0,00	0,00	0,27	0,00	0,00	2,17	0,00	0,00	97,57	0,00	0,00
Am 41_3	0,00	0,00	0,00	0,28	0,00	0,00	1,65	0,00	0,00	98,08	0,00	0,00
Am 42_1	0,00	17,65	0,00	60,53	13,63	0,00	3,70	1,65	2,85	0,00	0,00	0,00
Am 42_2	0,00	17,46	0,00	60,89	13,58	0,00	3,55	1,67	2,86	0,00	0,00	0,00
Am 43_1	0,00	0,00	0,00	0,17	0,00	0,00	0,81	0,00	0,00	99,03	0,00	0,00
Am 43_2	0,00	0,00	0,00	0,17	0,00	0,00	0,77	0,00	0,00	99,07	0,00	0,00
Am 43_3	0,00	0,00	0,00	0,22	0,00	0,00	0,43	0,00	0,00	99,36	0,00	0,00
Am 44_1	65,34	27,77	0,00	0,00	0,00	0,00	0,37	0,83	5,71	0,00	0,00	0,00
Am 44_2	64,66	28,02	0,00	0,08	0,00	0,00	0,51	0,82	5,91	0,00	0,00	0,00
Am 44_3	64,61	28,33	0,00	0,00	0,00	0,00	0,37	0,77	5,93	0,00	0,00	0,00
Am 45_1	0,00	17,75	0,09	62,64	14,62	0,00	0,00	1,87	2,87	0,00	0,07	0,00
Am 45_2	0,00	17,88	0,00	62,88	14,25	0,00	0,00	1,91	2,92	0,00	0,07	0,00

TABELA 5.1. 3. Resultados (% m/m) do XRF para as amostras 45_3 a 54_2.

Amostras	Co	Cr	Cu	Fe	Ni	Pb	Zn	Mn	Mo	Ti	V	Al
Am 45_3	0,00	17,88	0,00	62,94	14,27	0,00	0,00	1,80	2,95	0,00	0,07	0,00
Am 46_1	0,00	17,84	0,00	62,98	14,37	0,00	0,00	1,82	2,93	0,00	0,00	0,00
Am 46_2	0,00	17,82	0,00	62,76	16,05	0,00	0,00	1,80	2,94	0,00	0,00	0,00
Am 46_3	0,00	17,76	0,19	62,71	14,59	0,00	0,00	1,76	2,92	0,00	0,00	0,00
Am 47_1	0,00	17,53	0,30	63,10	14,25	0,00	0,00	1,89	2,95	0,00	0,00	0,00
Am 47_2	0,00	17,68	0,20	63,06	14,17	0,00	0,00	1,94	2,93	0,00	0,00	0,00
Am 48_1	0,00	17,69	0,00	62,25	14,54	0,00	0,68	1,89	2,86	0,00	0,00	0,00
Am 48_2	0,00	17,62	0,00	62,86	14,72	0,00	0,00	1,87	2,87	0,00	0,00	0,00
Am 48_3	0,00	17,54	0,43	62,18	14,16	0,00	0,13	1,83	2,86	0,00	0,00	0,00
Am 49_1	0,00	17,71	0,00	62,85	14,66	0,00	0,00	1,75	2,86	0,00	0,07	0,00
Am 49_2	0,00	17,90	0,00	62,97	14,24	0,00	0,00	1,83	2,96	0,00	0,00	0,00
Am 49_3	0,00	17,99	0,00	62,79	14,18	0,00	0,00	1,90	2,93	0,00	0,12	0,00
Am 50_1	0,00	17,81	0,13	62,62	14,62	0,00	0,14	1,74	2,87	0,00	0,08	0,00
Am 50_2	0,00	17,76	0,16	62,27	14,53	0,00	0,20	1,81	2,88	0,41	0,00	0,00
Am 50_3	0,00	17,68	0,00	62,52	14,83	0,00	0,28	1,75	2,86	0,00	0,09	0,00
Am 51_1	0,00	0,00	0,00	0,21	0,00	0,00	0,17	0,00	0,00	95,14	4,49	0,00
Am 51_2	0,00	0,00	0,00	0,21	0,00	0,00	0,17	0,00	0,00	95,34	4,29	0,00
Am 51_3	0,00	0,00	0,00	0,96	0,00	0,00	0,12	0,00	0,00	94,53	4,41	0,00
Am 52_1	0,00	17,84	0,09	61,66	15,62	0,00	0,00	1,72	3,09	0,00	0,00	0,00
Am 52_2	0,00	17,67	0,00	61,40	14,03	0,00	2,03	1,82	3,05	0,00	0,00	0,00
Am 52_3	0,00	17,82	0,10	62,81	14,46	0,00	0,00	1,82	3,01	0,00	0,00	0,00
Am 53_1	0,00	17,71	0,00	61,39	13,85	0,00	2,17	1,83	3,06	0,00	0,00	0,00
Am 53_2	0,00	17,80	0,00	62,73	14,62	0,00	0,00	1,80	3,01	0,00	0,00	0,00
Am 53_3	0,00	20,71	0,10	55,08	19,22	0,00	0,00	1,60	3,30	0,00	0,00	0,00
Am 54_1	0,00	17,23	0,00	60,84	13,85	0,00	3,48	1,71	2,89	0,00	0,00	0,00
Am 54_2	0,00	17,23	0,00	60,41	13,55	0,00	4,11	1,83	2,88	0,00	0,00	0,00

5.2- Otimização dos parâmetros da LIBS

Para a otimização dos parâmetros relacionados com energia, *delay time* e *spot size*, fez-se um planejamento fatorial do tipo Doehlert. Esse planejamento foi utilizado por ter a possibilidade de avaliar as variáveis em diferentes níveis. Em estudos anteriores do grupo de pesquisa, já foi observado que a energia é a variável que mais influencia nas medidas, seguida do *delay time*. Assim, a energia foi variada em 7 níveis diferentes e o *delay time* em 5 níveis, enquanto que o *spot size*, por ser a variável que menos influencia, foi variado em 3 níveis.

Com o planejamento Doehlert, obteve-se 22 modelos de regressão (11 para a área e 11 para altura) para 9 amostras diferentes e 9 analitos. No caso do Fe, como já foi citado, dois modelos adicionais foram calculados para abranger toda a diversidade de amostras. Assim, o modelo obtido para a área do sinal do Fe II 259,940 nm na amostra 54 foi:

$$\begin{aligned} \text{Área} = & 10992 - 18750DT + 12665E - 5108SS + 13324DT^2 - 12914DT \times E & (13) \\ & \pm 2630 \quad \pm 3857 \quad \pm 3857 \quad \pm 3854 \quad \pm 6443 \quad \pm 8449 \end{aligned}$$

Onde: DT= *delay time*; E = energia; SS = *spot size*. Os coeficientes mostrados foram superiores ao intervalo de confiança.

Neste caso, os coeficientes b_0 , b_1 , b_2 , b_3 , b_{11} e b_{12} foram significativos ao nível de 95% de confiança. Para obter uma área alta, o *delay time* e o *spot size* devem estar em um nível baixo, pois possuem efeito negativo, como observado na equação 13. Já a energia, por possuir um efeito positivo, a mesma deve estar em um nível mais alto. Apenas o termo quadrático do *delay time* foi significativo, onde há um ponto de inflexão mínimo com a concavidade voltada para cima, por ser um termo positivo. Há também a interação entre o *delay time* e energia, onde as duas variáveis devem estar em níveis opostos, pois o efeito é negativo. Como já foi citado, a energia deve estar em um nível alto e o *delay time* em um nível baixo. Se somente o b_0 fosse válido, a área do Fe para essa amostra seria 10992 ± 2630 (intervalo de confiança). Estes valores são obtidos por meio de mínimos quadrados e a significância dos mesmos é avaliada por meio da ANOVA. Para o caso do Fe da amostra 54, a tabela ANOVA está descrita na TABELA 5.2.1.

TABELA 5.2. 1. ANOVA para o Fe II 259,940 nm da amostra 54.

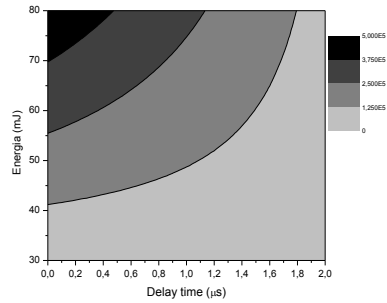
Amostra (linha de emissão)	Parâmetros	Soma Quadrática (SQ)	Graus de liberdade (GL)	Média Quadrática (MQ)	Teste F
54 (Fe II 259,940 nm)	Regressão (R)	2,545x10 ⁹	5	5,091x10 ⁸	Teste F1: MQR/MQres = 45,51
	Resíduo (res)	1,007x10 ⁸	9	1,118x10 ⁷	F tab (95%, 5, 9) = 3,481
	Total (T)	2,646x10 ⁹	14	1,890x10 ⁸	
	Erro puro (ep)	8,290x10 ⁶	2	4,145x10 ⁶	Teste F2: MQfaj/MQep = 3,184
	Falta de ajuste (faj)	9,240x10 ⁷	7	1,320x10 ⁷	F tab (95%, 7, 2) = 19,35
	Variação explicada (R ²)	0,9619	R	0,9808	
Máxima variação explicável (R ² máximo)	0,9969	R máximo	0,9984		

A SQR deve ser a mais alta possível, onde os valores previstos devem ser concordantes com aqueles obtidos experimentalmente. Já com a SQres é possível observar a quantidade de informação que não foi modelada pelo coeficiente de regressão, sendo que o ideal é ser a menor possível. A SQep tem por objetivo mostrar a magnitude do erro puro (3 réplicas do ponto central) e a SQfaj mostra a incapacidade do modelo em se ajustar aos pontos experimentais obtidos. Com a variação explicada observa-se a quantidade de informação que a SQR explica em relação a SQT, quanto mais próximo de 1, mais consistente é o modelo. No caso da TABELA 5.2.1, o R² foi 0,9619, ou seja, 96% da informação foi explicada pelos 6 coeficientes válidos e menos de 4% não foi modelada.

Neste caso, no teste F1, o F calculado foi maior que o F tabelado, indicando que a MQR e a MQres são estatisticamente diferentes, sendo o modelo confiável para efetuar previsões. No teste F2, o F calculado foi menor que o F tabelado, onde a MQep e MQfaj se confundem, sendo estatisticamente iguais. Essa situação é ideal para o modelo, pois a incapacidade dos valores experimentais em se igualarem àqueles previstos (falta de ajuste) se confunde com o erro aleatório (erro puro) cometido em laboratório. Portanto, o modelo não sofre falta de ajuste e pode-se utilizar a MQres como estimativa da variância dos coeficientes.

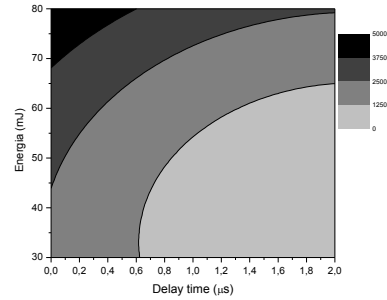
Os modelos para a área e altura foram concordantes, assim são mostrados nas FIGURAS 5.2.1 a 5.2.3 apenas os gráficos de contorno e as equações com os coeficientes significativos (*backward calculation*) para a área do sinal. O termo *backward calculation* significa que inicialmente foram calculados modelos com os 10 coeficientes descritos anteriormente. Com os coeficientes significativos (mantidos), calculou-se novamente os modelos que são mostrados nas figuras a seguir.

(a) Amostra 1 (Zn I 481,053 nm):



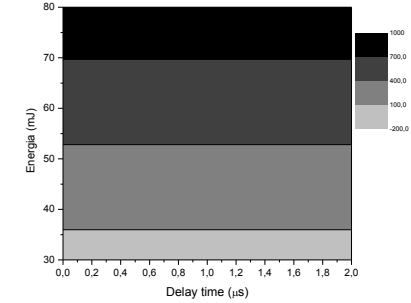
$$\begin{aligned} \text{Área} = & 155142 - 90883\text{DT} + 138788\text{E} - 114063\text{DT} \times \text{E} \\ & \pm 13445 \quad \pm 26037 \quad \pm 26038 \quad \pm 57039 \\ & R^2 = 0,95 \end{aligned}$$

(b) Amostra 2 (Cu I 324,754 nm):



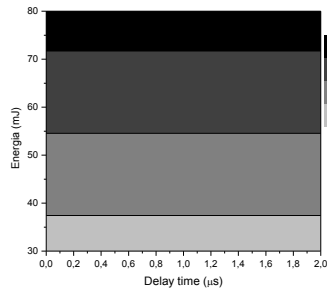
$$\begin{aligned} \text{Área} = & 1286 - 1132\text{DT} + 1425\text{E} + 492\text{DT}^2 + 938\text{E}^2 \\ & \pm 211 \quad \pm 246 \quad \pm 246 \quad \pm 416 \quad \pm 416 \\ & R^2 = 0,97 \end{aligned}$$

(c) Amostra 10 (Mo I 386,411 nm):



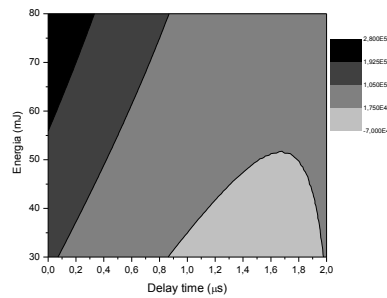
$$\begin{aligned} \text{Área} = & 439 + 514\text{E} \\ & \pm 162 \quad \pm 314 \\ & R^2 = 0,66 \end{aligned}$$

(d) Amostra 10 (Ni I 361,939 nm):



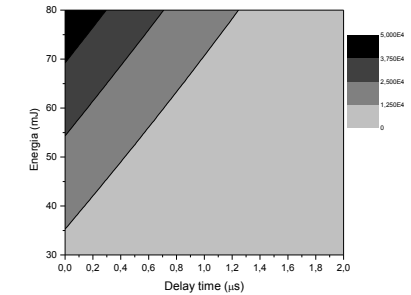
$$\begin{aligned} \text{Área} = & 611 + 758\text{E} \\ & \pm 265 \quad \pm 513 \\ & R^2 = 0,61 \end{aligned}$$

(e) Amostra 35 (Fe II 259,940 nm):



$$\begin{aligned} \text{Área} = & 49113 - 82764\text{DT} + 45617\text{E} + 58053\text{DT}^2 - 39363\text{DT} \times \text{E} \\ & \pm 5301 \quad \pm 7772 \quad \pm 7772 \quad \pm 12984 \quad \pm 17027 \\ & R^2 = 0,99 \end{aligned}$$

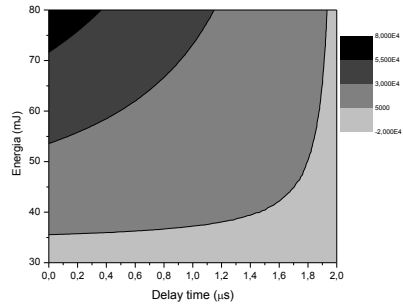
(f) Amostra 44 (Cr II 283,563 nm):



$$\begin{aligned} \text{Área} = & 5855 - 12263\text{DT} + 9794\text{E} + 7407\text{DT}^2 + 4459\text{E}^2 - 12307\text{DT} \times \text{E} \\ & \pm 1801 \quad \pm 2093 \quad \pm 2093 \quad \pm 3543 \quad \pm 3544 \quad \pm 4584 \\ & R^2 = 0,98 \end{aligned}$$

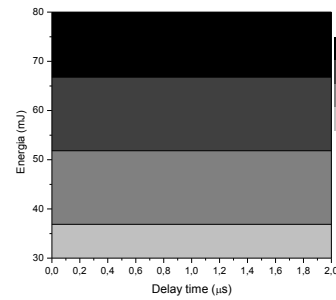
FIGURA 5.2. 1. Gráficos de contorno (*delay time versus energia*) e seus respectivos modelos para a área do sinal. (a) Zn I 481,053 nm; (b) Cu I 324,754 nm; (c) Mo I 386,411 nm; (d) Ni I 361,939 nm; (e) Fe II 259,940 nm; (f) Cr II 283,563 nm.

(g) Amostra 45 (Fe II 259,940 nm):



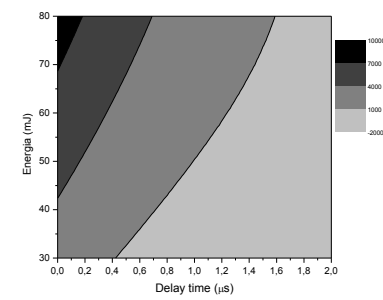
$$\begin{aligned} \text{Área} = & 17352 - 14597\text{DT} + 20056\text{E} - 19932\text{DT} \times \text{E} \\ & \pm 4007 \quad \pm 7760 \quad \pm 7761 \quad \pm 17000 \\ & R^2 = 0,84 \end{aligned}$$

(h) Amostra 51 (V I 437,924 nm):



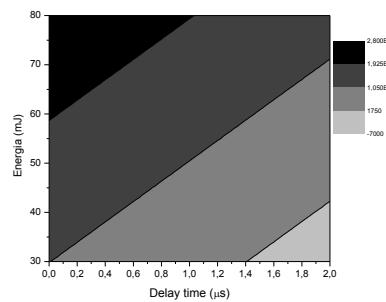
$$\begin{aligned} \text{Área} = & 43 + 29\text{E} \\ & \pm 14 \quad \pm 27 \\ & R^2 = 0,30 \end{aligned}$$

(i) Amostra 54 (Mn II 294,920 nm):



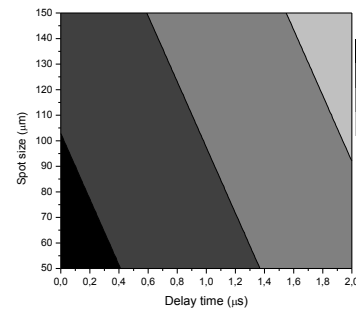
$$\begin{aligned} \text{Área} = & 1256 - 2378\text{DT} + 1595\text{E} + 1810\text{DT}^2 - 1699\text{DT} \times \text{E} \\ & \pm 237 \quad \pm 348 \quad \pm 348 \quad \pm 581 \quad \pm 761 \\ & R^2 = 0,98 \end{aligned}$$

(j) Amostra 38 (Al I 394,400 nm):



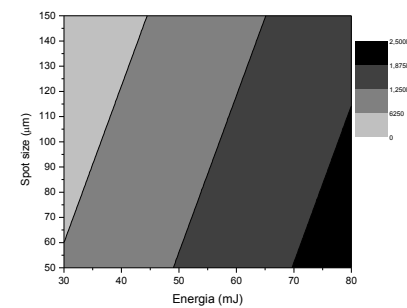
$$\begin{aligned} \text{Área} = & 11873 - 6265\text{DT} + 8744\text{E} \\ & \pm 1427 \quad \pm 2764 \quad \pm 2764 \\ & R^2 = 0,90 \end{aligned}$$

(k) Amostra 38 (Al I 394,400 nm):



$$\begin{aligned} \text{Área} = & 11873 - 6265\text{DT} - 2980\text{SS} \\ & \pm 1427 \quad \pm 2764 \quad \pm 2762 \\ & R^2 = 0,90 \end{aligned}$$

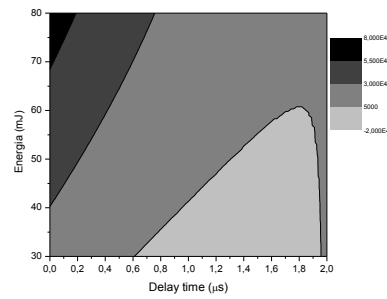
(l) Amostra 38 (Al I 394,400 nm):



$$\begin{aligned} \text{Área} = & 11873 + 8744\text{E} - 2980\text{SS} \\ & \pm 1427 \quad \pm 2764 \quad \pm 2762 \\ & R^2 = 0,90 \end{aligned}$$

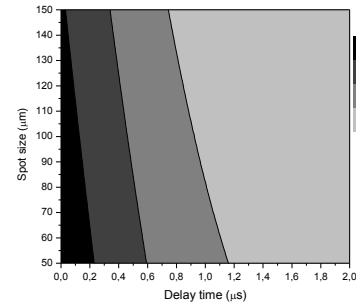
FIGURA 5.2. 2. Gráficos de contorno e seus respectivos modelos para a área do sinal. *Delay time versus energia* (g) Fe II 259,940 nm; (h) V I 437,924 nm; (i) Mn II 294,920 nm; (j) Al I 394,400 nm; (k) *delay time versus spot size* (Al I 394,400 nm); (l) *energia versus spot size* (Al I 394,400 nm).

(m) Amostra 54 (Fe II 259,940 nm):



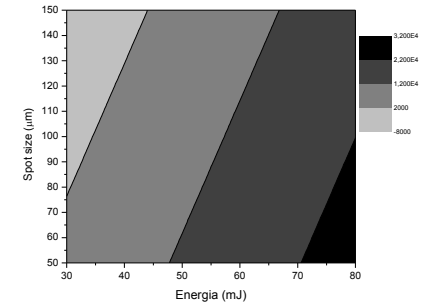
$$\begin{aligned} \text{Área} = & 10992 - 18750DT + 12665E + 13324DT^2 - 12914DT \times E \\ & \pm 2630 \quad \pm 3857 \quad \pm 3857 \quad \pm 6443 \quad \pm 8449 \\ & R^2 = 0,96 \end{aligned}$$

(n) Amostra 54 (Fe II 259,940 nm):



$$\begin{aligned} \text{Área} = & 10992 - 18750DT - 5108SS + 13324DT^2 \\ & \pm 2630 \quad \pm 3857 \quad \pm 3854 \quad \pm 6443 \\ & R^2 = 0,96 \end{aligned}$$

(o) Amostra 54 (Fe II 259,940 nm):



$$\begin{aligned} \text{Área} = & 10992 + 12665E - 5108SS \\ & \pm 2630 \quad \pm 3857 \quad \pm 3854 \\ & R^2 = 0,96 \end{aligned}$$

FIGURA 5.2. 3. Gráficos de contorno e seus respectivos modelos para a área do sinal. (m) *delay time versus* energia (Fe II 259,940 nm); (n) *delay time versus spot size* (Fe II 259,940 nm); (o) *energia versus spot size* (Fe II 259,940 nm).

Com exceção do Al (amostra 38) e do Fe (amostra 54), apenas a energia e o *delay time* foram significativos para os analitos, sendo que alguns apresentaram apenas dois coeficientes significativos (caso do Mo, Ni e V) e para o Cr, 6 coeficientes foram válidos. Já para o Al e para o Fe (amostra 54), o *spot size* também foi significativo, obtendo 3 gráficos de contorno para cada um deles.

Com os modelos calculados e propostos, observou-se que utilizando uma alta energia, baixo *delay time* e baixo *spot size*, altos sinais de área e altura são obtidos. Entretanto, para o Zn, ao utilizar essa condição extrema, é promovida a saturação do sinal (intensidades acima de 70000). Assim, foi necessário obter uma condição intermediária que satisfizesse todos os analitos e amostras. Objetivando a identificação de uma condição de compromisso, foi montado no Excel uma planilha com todos os modelos obtidos e testadas condições instrumentais que proporcionasse altos sinais para todos os analitos sem saturar o detector (caso do Zn). Finalmente, a melhor condição selecionada para todas as 80 amostras foi uma energia de 60 mJ, *delay time* de 0,9 μ s e *spot size* de 100 μ m. Com essa condição, não foi observado a saturação do sinal e nem problemas com auto absorção. A fluência do laser foi calculada a partir da equação 14:

$$I(\text{Fluência}) = \frac{E(\text{mJ}) \times \frac{1\text{J}}{10^3 \text{mJ}}}{1\text{pulso}} \times \frac{1}{A(\text{mm}^2) \times \frac{1^2 \text{cm}^2}{10^2 \text{mm}^2}} = \frac{\text{J}}{\text{cm}^2} \quad (14)$$

Para uma E = 60 mJ e SS = 100 μ m. Tem-se uma A = 7,85x10⁻³ mm².

$$I(\text{Fluência}) = \frac{60\text{mJ} \times \frac{1\text{J}}{10^3 \text{mJ}}}{1\text{pulso}} \times \frac{1}{7,85 \times 10^{-3} \text{mm}^2 \times \frac{1^2 \text{cm}^2}{10^2 \text{mm}^2}} = 764 \frac{\text{J}}{\text{cm}^2}$$

Assim, a fluência do laser na condição estabelecida foi de 764 J/cm². A fluência é definida como a quantidade de energia por unidade de área irradiada pelo laser, sendo determinada pelo diâmetro de focalização do laser (*spot size*) sobre a superfície da amostra e da energia aplicada por pulso. Com o aumento da fluência, observa-se também um aumento da intensidade das linhas de emissão. Assim, há uma maior eficiência de ablação e um maior número de átomos vaporizados no

interior do plasma. Entretanto, deve-se atentar aos efeitos de blindagem do plasma (devido a elevada densidade de elétrons) e efeitos de auto-absorção, os quais podem ser observados com uma fluência muito elevada.

Após obtermos a condição otimizada, a mesma foi aplicada a todas as amostras. Com as FIGURAS 5.2.4 e 5.2.5 visualiza-se a faixa de valores de área obtida experimentalmente com a condição otimizada (primeira caixa) e a área prevista pelo modelo (segunda caixa) para as 9 amostras e os analitos utilizados na parte de otimização. O quadrado pequeno no interior das caixas representa a área média e os extremos das caixas juntamente com os asteriscos representam o maior e o menor valor obtido experimentalmente (primeira caixa) e previsto pelo modelo (segunda caixa).

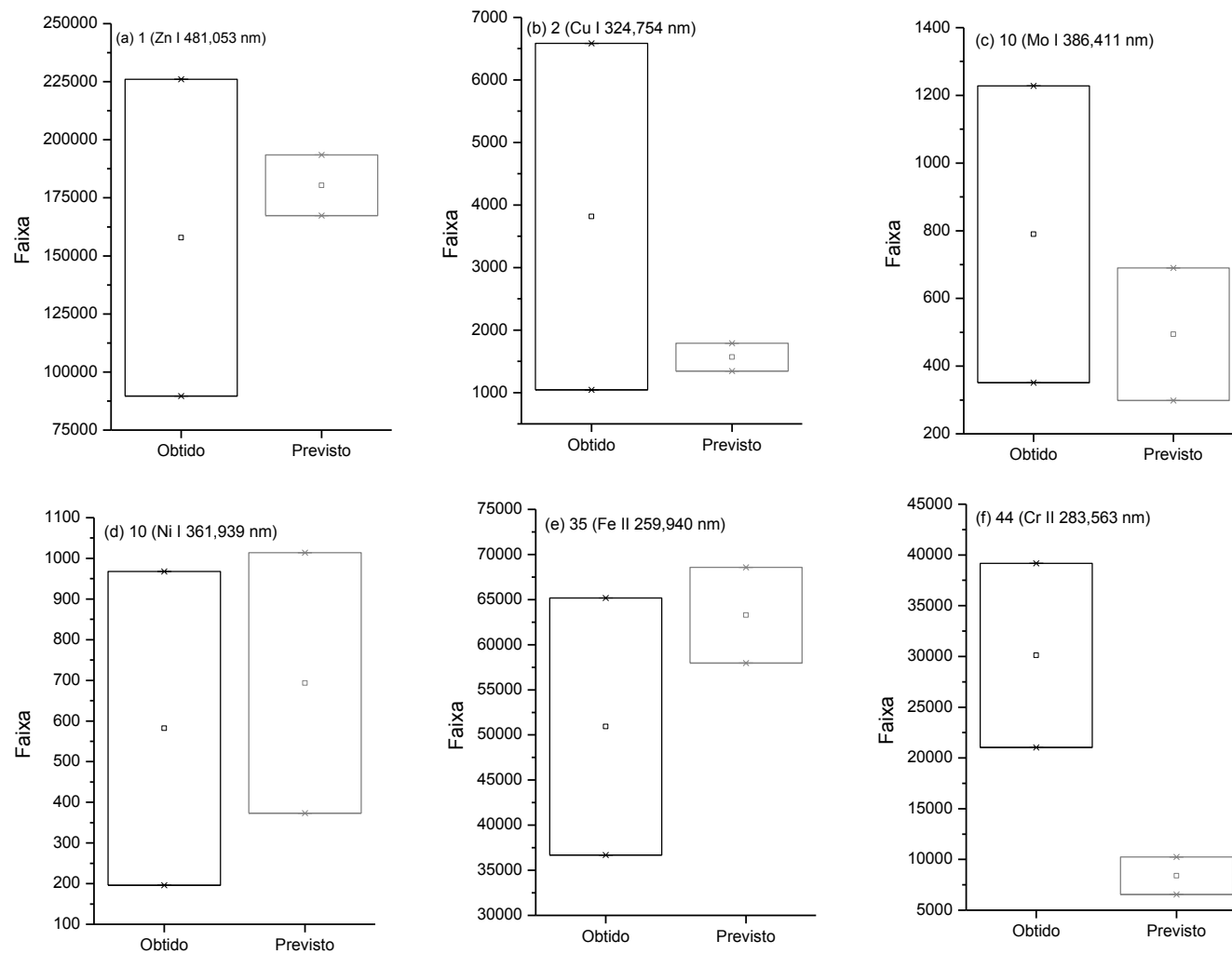


FIGURA 5.2. 4. Comparação entre os valores de área obtidas e previstas com as condições otimizadas. (a) Amostra 1 (Zn I 481,053 nm); (b) 2 (Cu I 324,754 nm); (c) 10 (Mo I 386,411 nm); (d) 10 (Ni I 361,939 nm); (e) 35 (Fe II 259,940 nm) e (f) 44 (Cr II 283,563 nm).

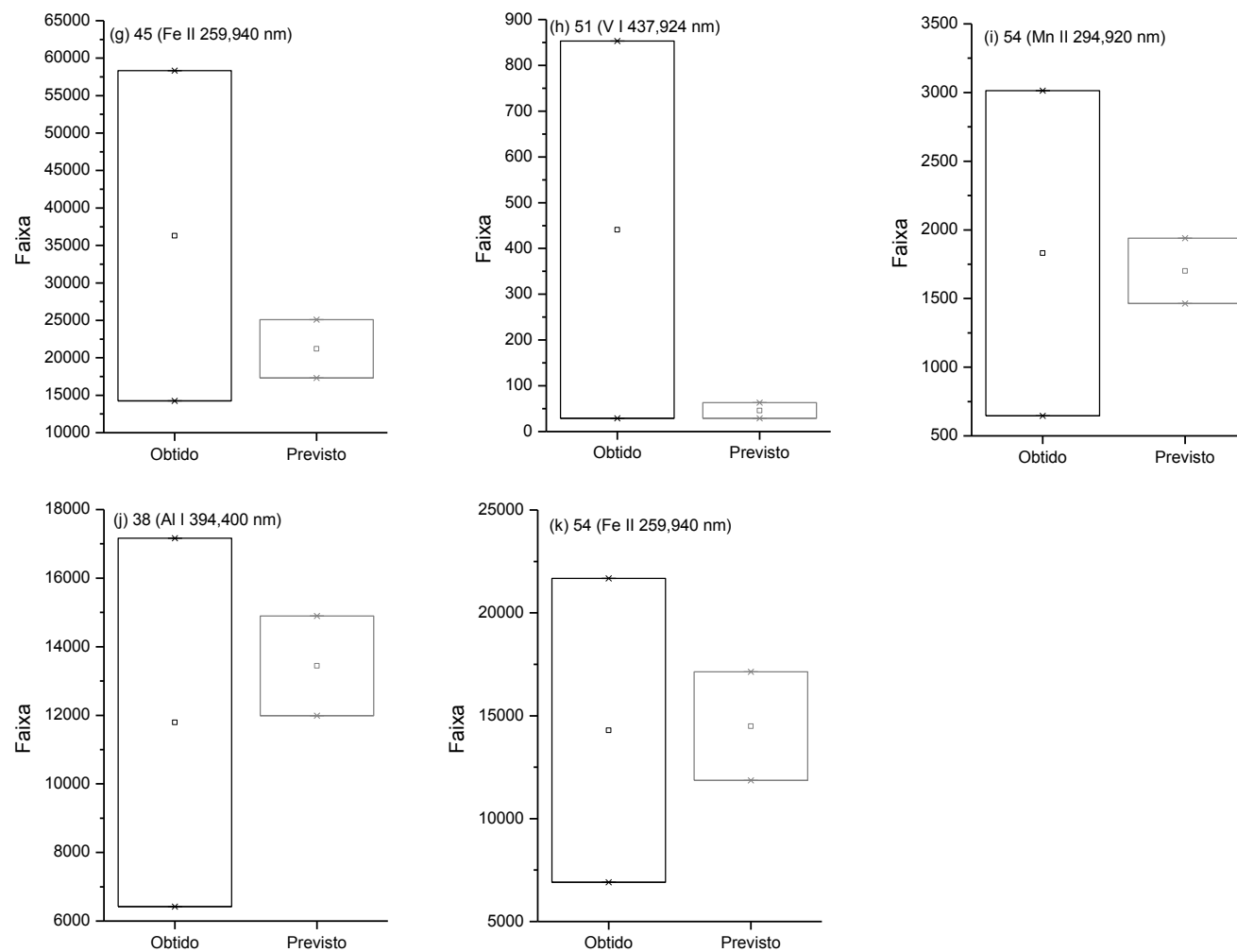


FIGURA 5.2. 5. Comparação entre os valores de área obtidas e previstas com as condições otimizadas. (g) Amostra 45 (Fe II 259,940 nm); (h) 51 (V I 437,924 nm); (i) 54 (Mn II 294,920 nm); (j) 38 (Al I 394,400 nm) e (k) 54 (Fe II 259,940 nm).

Na FIGURA 5.2.4a, tem-se os resultados para o Zn na amostra 1, onde a faixa da área obtida experimentalmente foi de aproximadamente 90000 a 225000. E a faixa da área prevista pelo modelo proposto (FIGURA 5.2.4a) foi de 170000 a 190000. Assim, o valor obtido experimentalmente foi concordante com o valor previsto pelo modelo, isso foi observado para todos os analitos e amostras, exceto para o Cr (FIGURA 5.2.4f), o qual não apresentou falta de ajuste no modelo, porém devido à complexidade das amostras, esse tipo de discrepância pode ocorrer.

Portanto, o uso do Doehlert foi útil para a otimização dos parâmetros instrumentais do equipamento LIBS, possibilitando o uso eficiente do tempo e recursos, produzindo modelos mais precisos e estudar as variáveis ao mesmo tempo e as interações entre elas.

Para avaliar a estabilidade das linhas de emissão, fez-se um teste com 5 linhas do Cr: II 283,563; I 357,869; I 359,349; I 425,435 e I 427,48 nm. As áreas das linhas foram selecionadas e fez-se a razão de uma pela outra, resultando em 10 razões, como é mostrado nas FIGURAS 5.2.6 e 5.2.7, a qual possui gráficos de caixa a direita com o valor mínimo (losango), média (quadrado) e o máximo (losango) dos espectros e os pontos (esquerda dos gráficos) representam a razão das linhas de emissão selecionadas para os 100 espectros. As barras de erro representam uma faixa de 95%, o tamanho da caixa corresponde ao desvio padrão e a linha horizontal, a mediana.

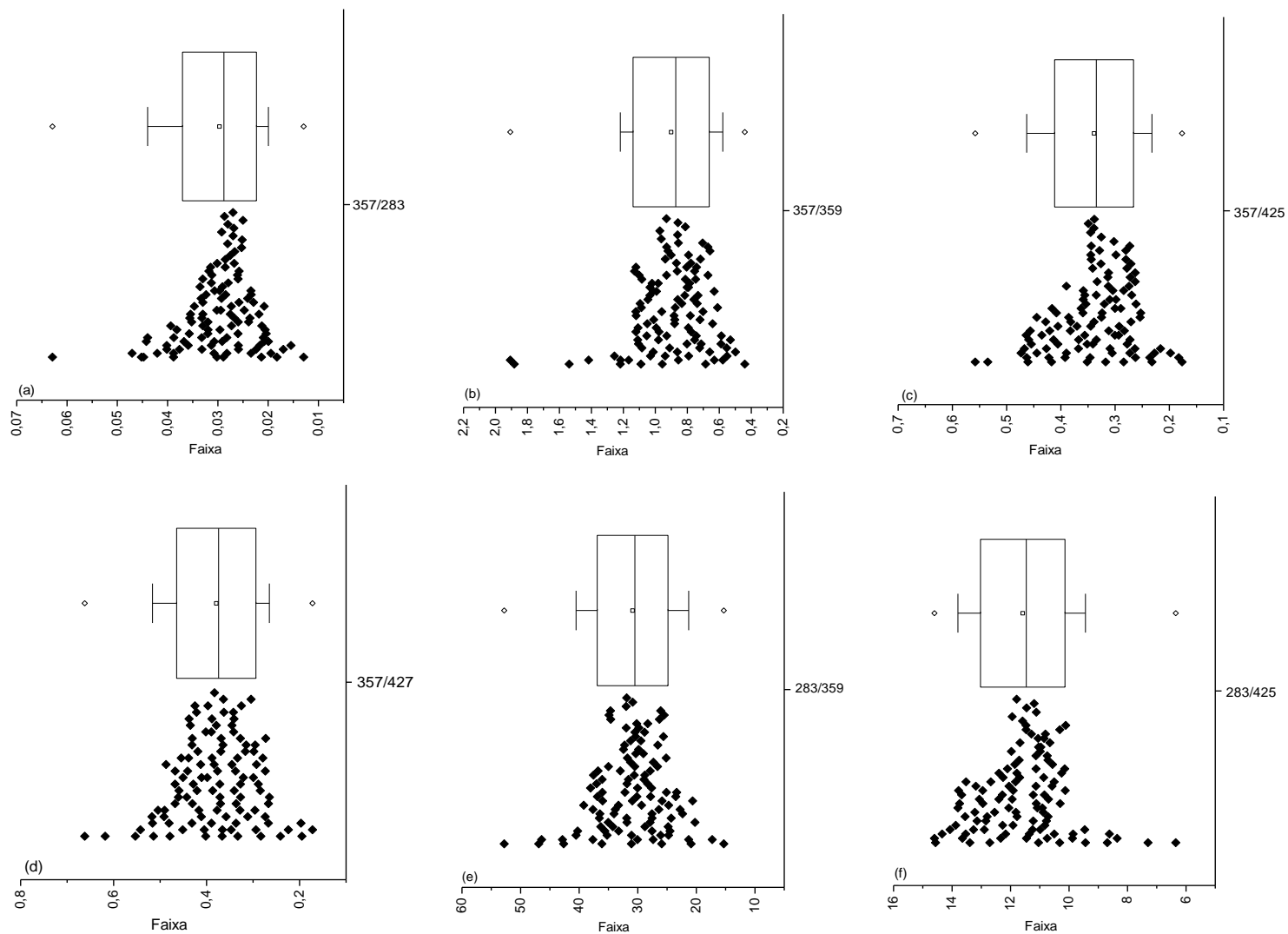


FIGURA 5.2. 6. Gráficos das razões das áreas das linhas de emissão. (a) 357,869/283,563; (b) 357,869/359,349; (c) 357,869/425,435; (d) 357,869/427,48; (e) 283,563/359,349 e (f) 283,563/425,435.

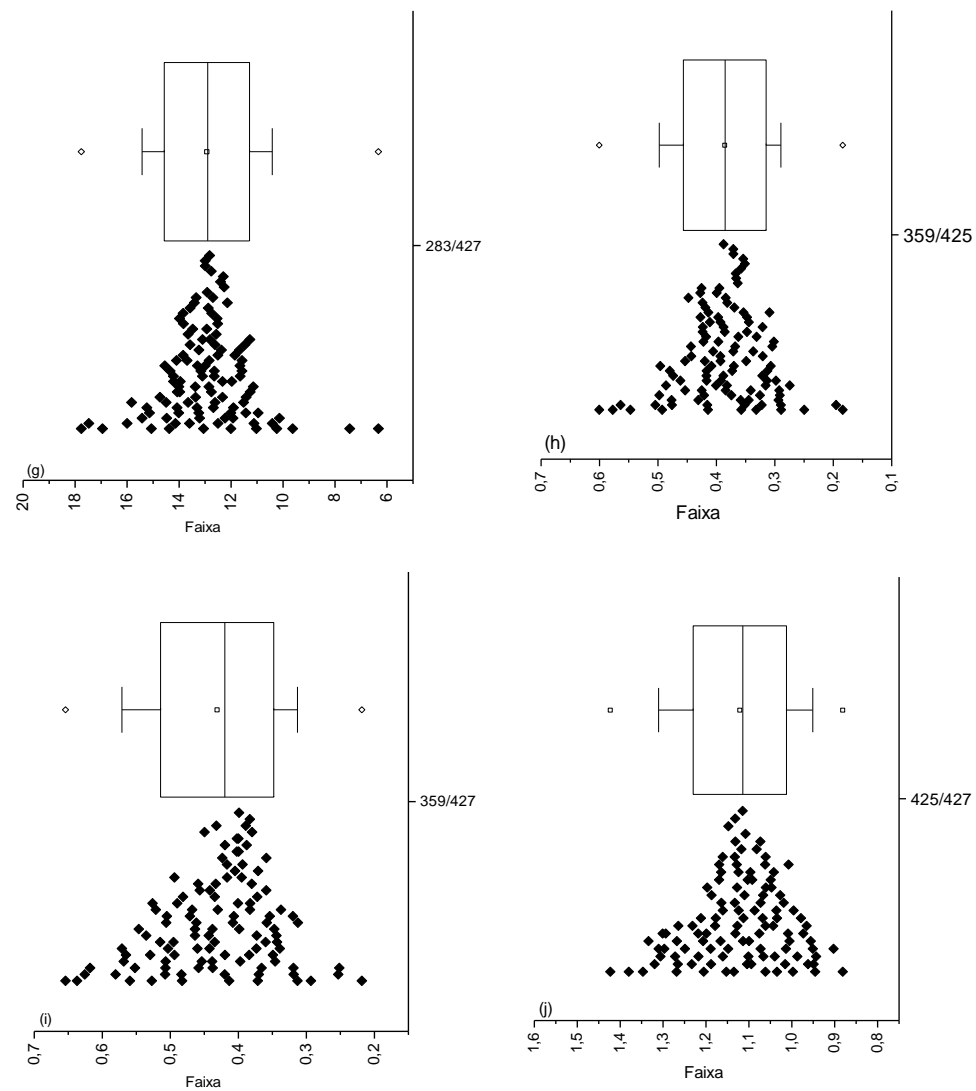


FIGURA 5.2. 7. Gráficos das razões das áreas das linhas de emissão. (g) 283,563/427,48; (h) 359,349/425,435; (i) 359,349/427,48 e (j) 425,435/427,48.

Com as FIGURAS 5.2.6 e 5.2.7 pode-se concluir que as linhas de emissão seguem uma distribuição normal, sendo assim os 100 espectros adquiridos em um raster são proporcionais, mostrando a estabilidade dos sinais.

5.3- Modelos de calibração univariado e multivariado

Após aplicar a condição otimizada a todas as amostras, os espectros foram submetidos a vários modos de normalização, como descritos nas TABELAS 4.4.1 E 4.4.2. O conjunto de dados foi dividido em 64 amostras de calibração e 16 amostras de validação através do algoritmo Kennard-Stone no Matlab. Modelos de calibração multivariado utilizando o perfil total do sinal (PLS) e modelos univariados com várias linhas de emissão foram avaliados para cada analito. Assim, o objetivo foi verificar o melhor modo de normalização, melhor estratégia de calibração (univariado ou multivariado) e a linha de emissão mais adequada para cada analito através dos menores valores de SEV. Com os resultados, testes F foram realizados a fim de verificar se havia diferença significativa entre as estratégias adotadas. Para observar essas diferenças estatísticas entre os modos de normalização, foi aplicado um teste F comparando a melhor normalização (menor SEV) com as demais para cada analito.

Nas FIGURAS 5.3.1 a 5.3.10 tem-se os valores de SEV (a) e os valores do teste F calculado (b) para os modelos univariados utilizando a área e altura do sinal e o modelo PLS para todos os analitos com os diferentes modos de normalização. No caso dos modelos univariados, os resultados mostrados são para as melhores linhas de emissão (aquelas que apresentaram menores valores de SEV) para cada analito. Nos gráficos do teste F, considera-se o valor 1 como o referencial (melhor normalização, representado por uma linha tracejada nos gráficos). Um asterisco significa a normalização escolhida e dois asteriscos indicam que tal normalização apresentou diferença significativa em comparação com a aquela que obteve melhores resultados. As normalizações pela linha do carbono 193,091 nm (normalizações 5 e 11) não foram aplicadas em nenhuma amostra, pois não houve sensibilidade suficiente.

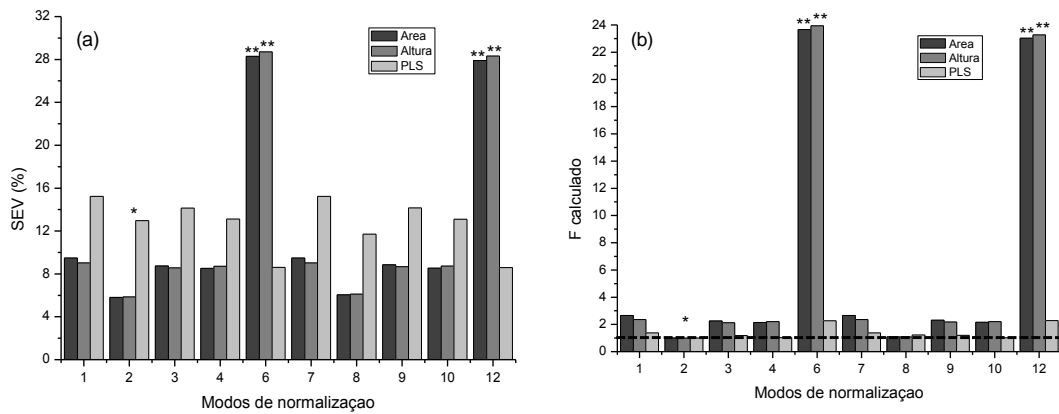


FIGURA 5.3. 1. Resultados da calibração univariada utilizando a área e altura do sinal e calibração multivariada utilizando PLS em todos os modos de normalização para o Al. (a) Valores de SEV (%) e (b) Valores do teste F.

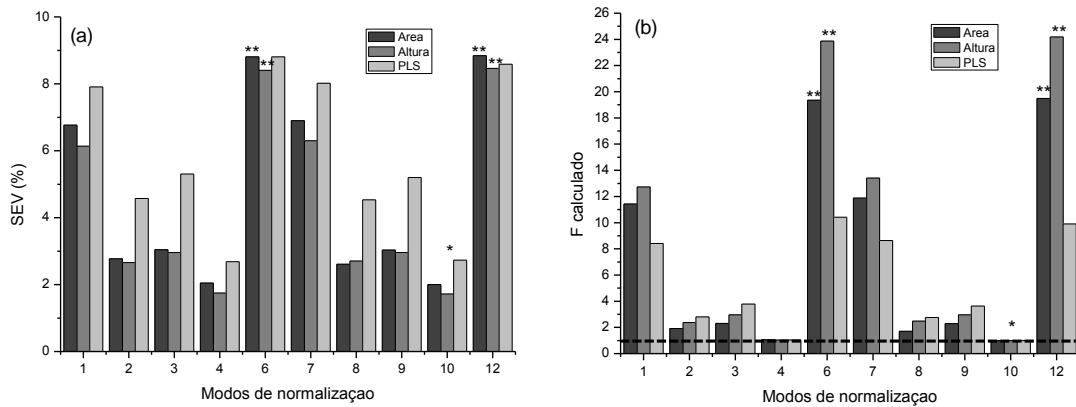


FIGURA 5.3. 2. Resultados da calibração univariada utilizando a área e altura do sinal e calibração multivariada utilizando PLS em todos os modos de normalização para o Cr. (a) Valores de SEV (%) e (b) Valores do teste F.

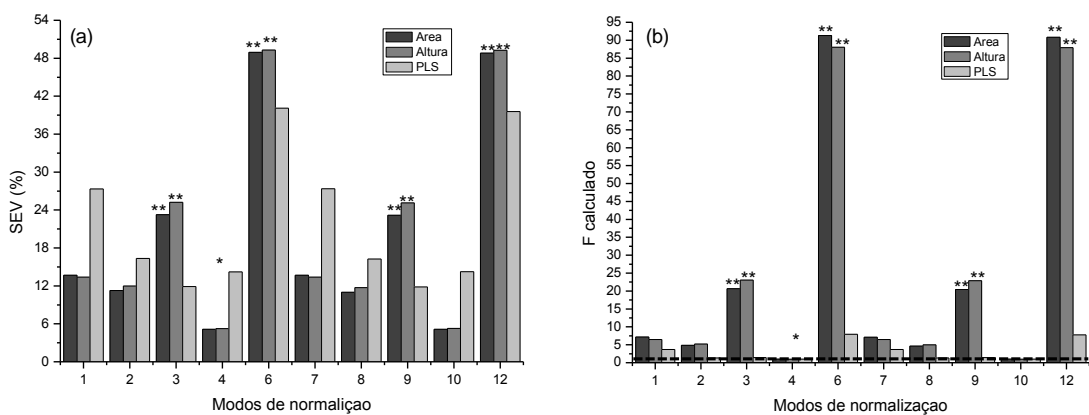


FIGURA 5.3. 3. Resultados da calibração univariada utilizando a área e altura do sinal e calibração multivariada utilizando PLS em todos os modos de normalização para o Cu. (a) Valores de SEV (%) e (b) Valores do teste F.

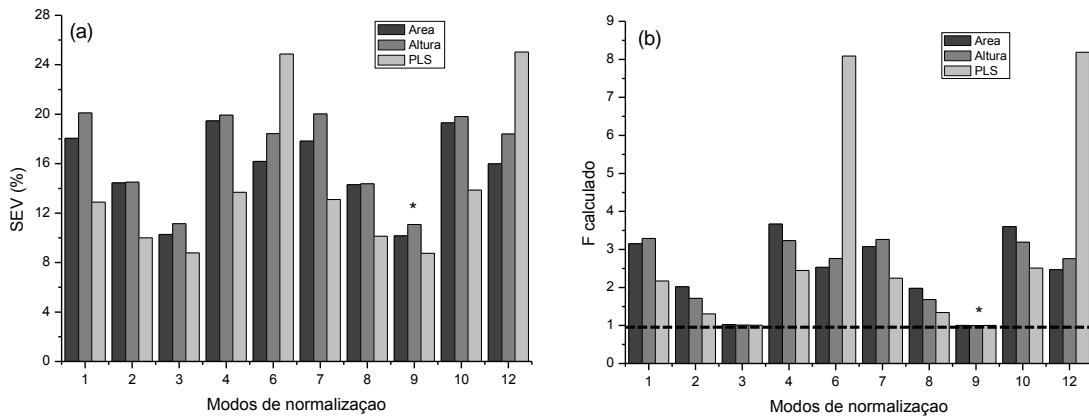


FIGURA 5.3. 4. Resultados da calibração univariada utilizando a área e altura do sinal e calibração multivariada utilizando PLS em todos os modos de normalização para o Fe. (a) Valores de SEV (%) e (b) Valores do teste F.

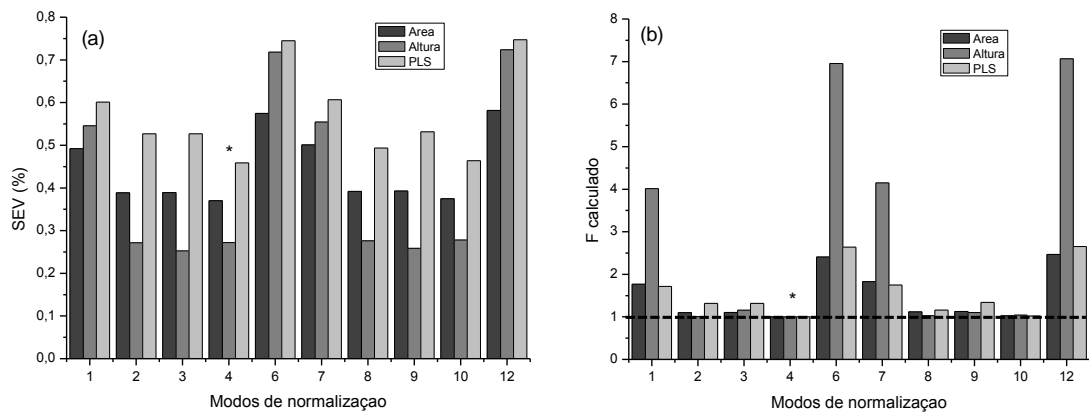


FIGURA 5.3. 5. Resultados da calibração univariada utilizando a área e altura do sinal e calibração multivariada utilizando PLS em todos os modos de normalização para o Mn. (a) Valores de SEV (%) e (b) Valores do teste F.

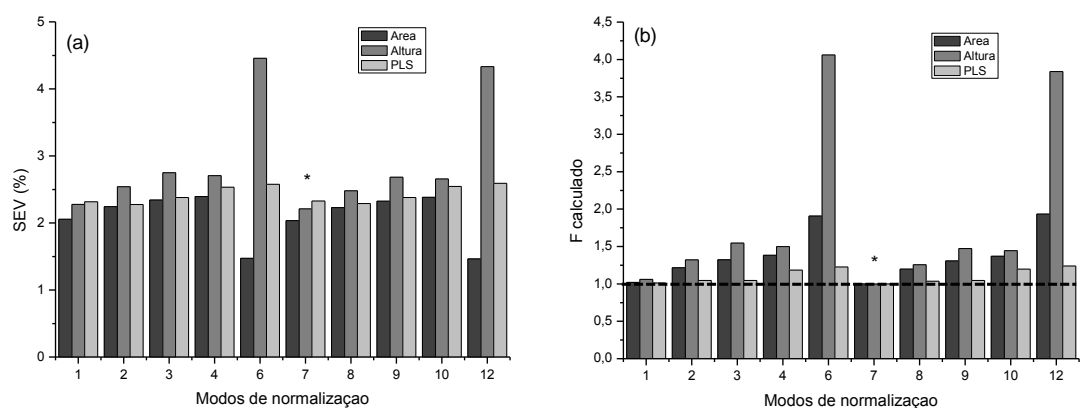


FIGURA 5.3. 6. Resultados da calibração univariada utilizando a área e altura do sinal e calibração multivariada utilizando PLS em todos os modos de normalização para o Mo. (a) Valores de SEV (%) e (b) Valores do teste F.

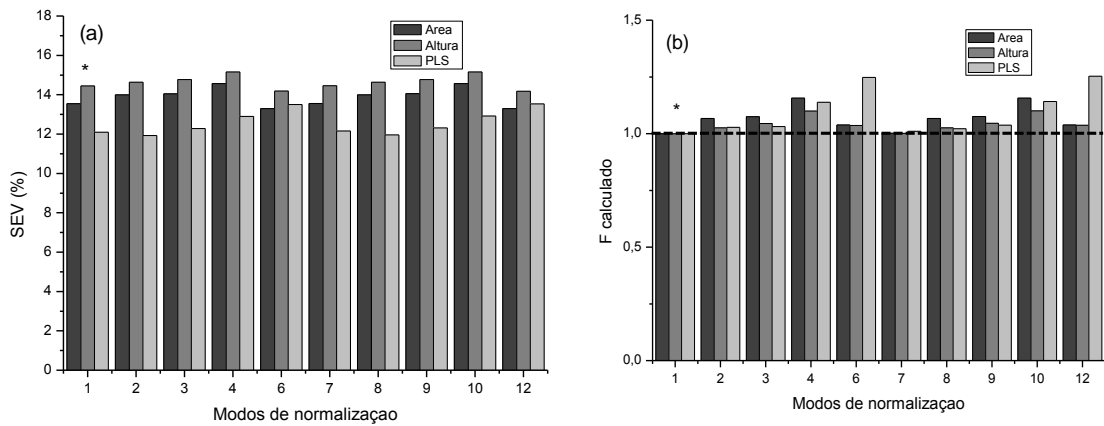


FIGURA 5.3. 7. Resultados da calibração univariada utilizando a área e altura do sinal e calibração multivariada utilizando PLS em todos os modos de normalização para o Ni. (a) Valores de SEV (%) e (b) Valores do teste F.

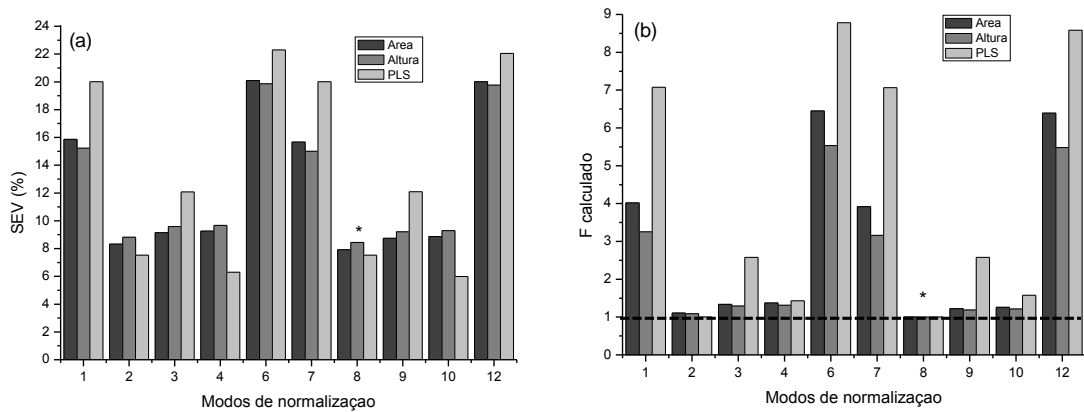


FIGURA 5.3. 8. Resultados da calibração univariada utilizando a área e altura do sinal e calibração multivariada utilizando PLS em todos os modos de normalização para o Ti. (a) Valores de SEV (%) e (b) Valores do teste F.

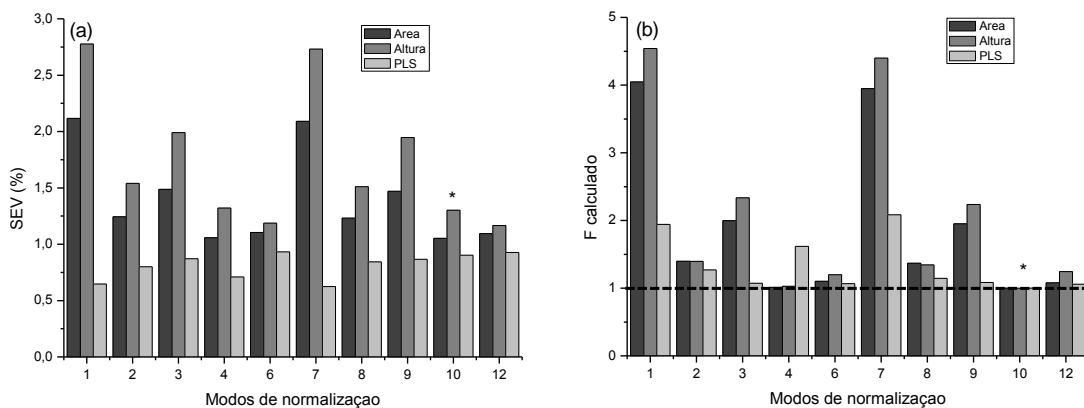


FIGURA 5.3. 9. Resultados da calibração univariada utilizando a área e altura do sinal e calibração multivariada utilizando PLS em todos os modos de normalização para o V. (a) Valores de SEV (%) e (b) Valores do teste F.

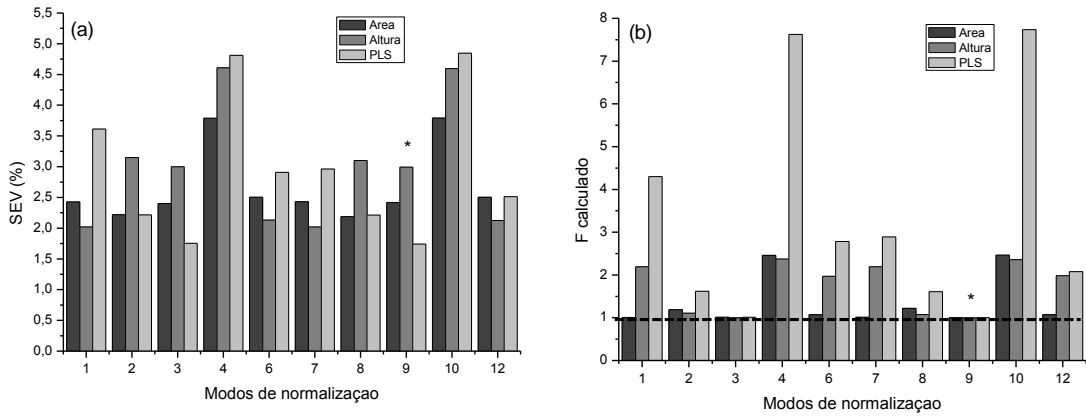


FIGURA 5.3. 10. Resultados da calibração univariada utilizando a área e altura do sinal e calibração multivariada utilizando PLS em todos os modos de normalização para o Zn. (a) Valores de SEV (%) e (b) Valores do teste F.

De acordo com os menores valores de SEV, as melhores normalizações foram escolhidas para todos os analitos:

- Al (FIGURA 5.3.1): norma individual e média dos espectros normalizados (marcado com um asterisco). Houve diferença significativa entre a melhor normalização (2) com aquelas que utilizaram a linha de emissão do carbono 247,856 nm (6 e 12, marcado com dois asteriscos) para a calibração univariada utilizando a área e altura do sinal.
- Cr (FIGURA 5.3.2): máximo individual e soma dos espectros normalizados (marcado com um asterisco). Houve diferença significativa entre a melhor normalização (10) com aquelas que utilizaram a linha de emissão do carbono 247,856 nm (6 e 12, marcado com dois asteriscos) para a calibração univariada utilizando a área e altura do sinal.
- Cu (FIGURA 5.3.3): máximo individual e média dos espectros normalizados (marcado com um asterisco). Houve diferença significativa entre a melhor normalização (4) com aquelas que utilizaram a área individual (3 e 9, marcado com dois asteriscos) e a linha de emissão do carbono 247,856 nm (6 e 12, marcado com dois asteriscos) para a calibração univariada utilizando a área e altura do sinal.
- Fe (FIGURA 5.3.4): área individual e soma dos espectros normalizados (marcado com um asterisco). Não houve diferença significativa entre a melhor normalização (9) com as demais.
- Mn (FIGURA 5.3.5): máximo individual e média dos espectros normalizados (marcado com um asterisco). Não houve diferença significativa entre a melhor normalização (4) com as demais.

- Mo (FIGURA 5.3.6): soma dos espectros (marcado com um asterisco). Não houve diferença significativa entre a melhor normalização (7) com as demais.
- Ni (FIGURA 5.3.7): média dos espectros (marcado com um asterisco). Não houve diferença significativa entre a melhor normalização (1) com as demais.
- Ti (FIGURA 5.3.8): norma individual e soma dos espectros normalizados (marcado com um asterisco). Não houve diferença significativa entre a melhor normalização (8) com as demais.
- V (FIGURA 5.3.9): máximo individual e soma dos espectros normalizados (marcado com um asterisco). Não houve diferença significativa entre a melhor normalização (10) com as demais.
- Zn (FIGURA 5.3.10): área individual e soma dos espectros normalizados (marcado com um asterisco). Não houve diferença significativa entre a melhor normalização (9) com as demais.

Assim, ao utilizar a normalização, o valor do erro padrão diminui, porém não é uma diminuição tão drástica a ponto de observar uma diferença estatística significativa. Assim, podemos observar que para a maioria dos analitos, a melhor normalização foi utilizando o máximo individual, seja fazendo a média dos espectros normalizados ou a soma dos mesmos. Com os resultados, conclui-se também que os valores obtidos para a média e soma dos espectros normalizados são semelhantes, não havendo diferença em se utilizar um ou outro.

A função `libs_treat` (ver anexo 2) desenvolvida para a realização dos cálculos de normalização foi concebida para ser aplicada a diversos tipos de amostras (orgânicas e inorgânicas). No GAIA a maioria das aplicações da técnica LIBS são direcionadas para amostras com altos teores de carbono (caso de alimentos e polímeros). Desta forma, resolveu-se aplicar a mesma função para amostras metálicas, mesmo não conhecendo a concentração de carbono nas mesmas. Entretanto, sabe-se que o carbono está presente mesmo em uma concentração variável. É importante mencionar que há uma linha de Fe I em 248,327 nm, sendo essa muito próxima a do C I 247,856. Como as amostras possuem uma concentração alta de Fe, deve ter ocorrido uma influência da mesma.

Através dos modelos preliminares, os vetores de regressão do PLS foram úteis para selecionar as linhas de emissão que influenciavam em cada modelo. Assim, várias linhas foram estudadas, variando desde 2 linhas para o Ni até 8 linhas para o Fe e Cr.

Na TABELA 5.3.1 tem-se a comparação dos resultados da calibração univariada e multivariada com a linha de emissão e a normalização selecionada para cada analito.

TABELA 5.3. 1. Comparação entre os modelos univariados (área e altura) e multivariados para os 10 analitos estudados.

Linhas de emissão (nm)	Faixa de concentração (%)	Normalização	Área		Altura		Multivariado (PLS)				
			SEV	LOD	SEV	LOD	SEV	VL	LOD _{min}	LOD _{max}	
Al	I 394,400	0 – 98	2	5,82	0,28	5,87	0,81	13,0	3	0,61	2,0
Cr	II 283,563	0 – 28	10	2,00	0,65	1,72	1,4	2,73	3	1,5	3,5
Cu	I 521,820	0 – 100	4	5,12	0,25	5,25	0,65	14,2	4	0,28	0,65
Fe	I 248,327	0 – 99	9	10,2	6,7	11,1	16	8,75	3	3,1	8,2
Mn	II 294,920	0 – 2	4	0,37	0,12	0,27	0,14	0,46	1	0,15	0,19
Mo	I 386,411	0 – 13	7	2,03	0,9	2,21	3,5	2,33	3	0,24	0,34
Ni	I 361,939	0 – 60	1	13,5	1,1	14,5	2,3	12,1	3	1,8	6,1
Ti	II 308,802	0 – 99	8	7,91	1,6	8,44	4,0	7,53	3	1,4	4,7
V	I 411,178	0 – 5	10	1,05	14	1,30	16	0,90	1	0,16	0,34
Zn	I 481,053	0 – 83	9	2,42	3,5	2,99	7,5	1,74	4	4,8	6,2

As linhas de emissão e as normalizações foram selecionadas de acordo com o menor valor de SEV. A única exceção foi para o analito Zn, o qual não foi separado em calibração e validação por possuir poucas amostras com esse analito. Assim o parâmetro observado para o Zn foi o SEC apenas. Os valores de SEV variaram de 0,27% para o Mn até 14% para o Cu e Ni. Em geral, os modelos univariados apresentaram menores valores de SEV para a maioria dos analitos, com exceção apenas do Fe, Ti e V. Um comportamento semelhante foi observado para os limites de detecção, onde para a maioria dos analitos na calibração univariada foram menores do que aqueles obtidos para os modelos multivariados. Esta observação é esperada, pois para o cálculo multivariado, houve a contribuição dos valores de y (variância das concentrações dos analitos). Ao aplicar um teste F entre os modelos multivariados e os univariados (tanto para a área quanto para a altura do sinal), não houve diferença significativa entre eles.

Nas FIGURAS 5.3.11 até 5.3.20, tem-se os modelos de calibração e validação para todos os analitos, multivariado com o perfil total do sinal e univariado utilizando a altura do sinal com a linha de emissão escolhida.

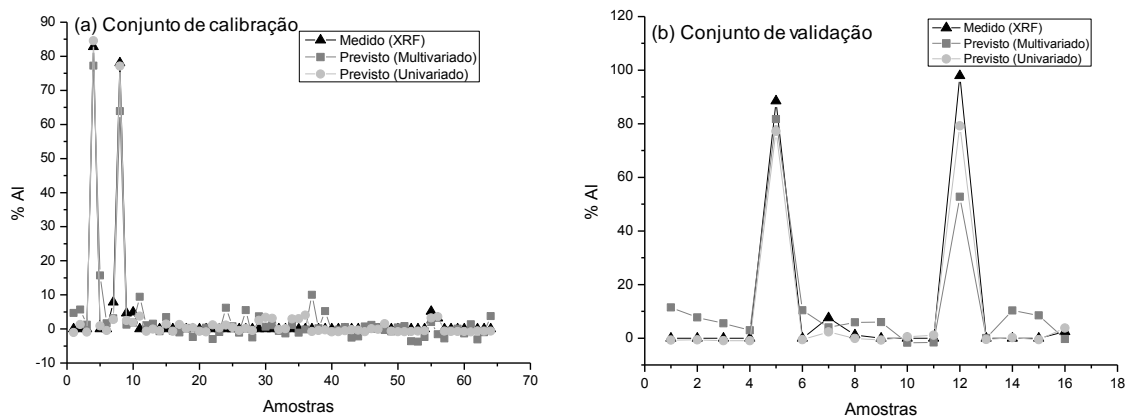


FIGURA 5.3. 11. Valores de referência (medido) e os valores previstos (multivariado e univariado) para o Al. (a) Conjunto de calibração e (b) Conjunto de validação.

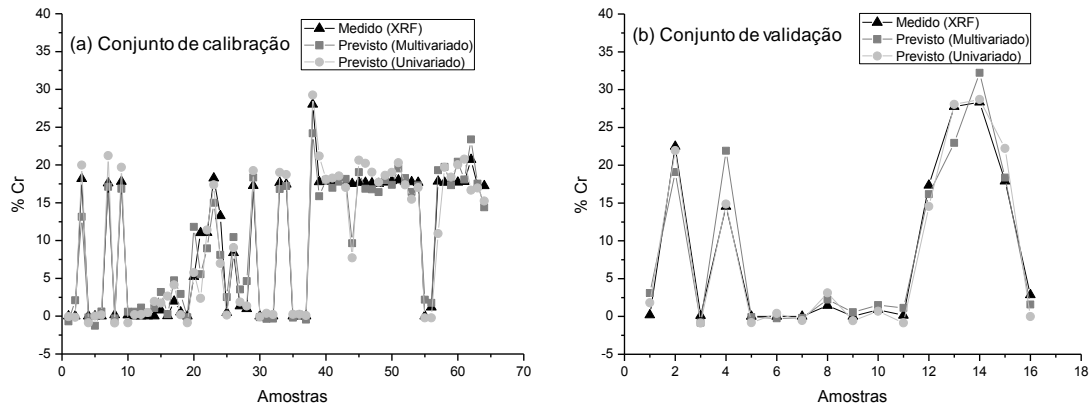


FIGURA 5.3. 12. Valores de referência (medido) e os valores previstos (multivariado e univariado) para o Cr. (a) Conjunto de calibração e (b) Conjunto de validação.

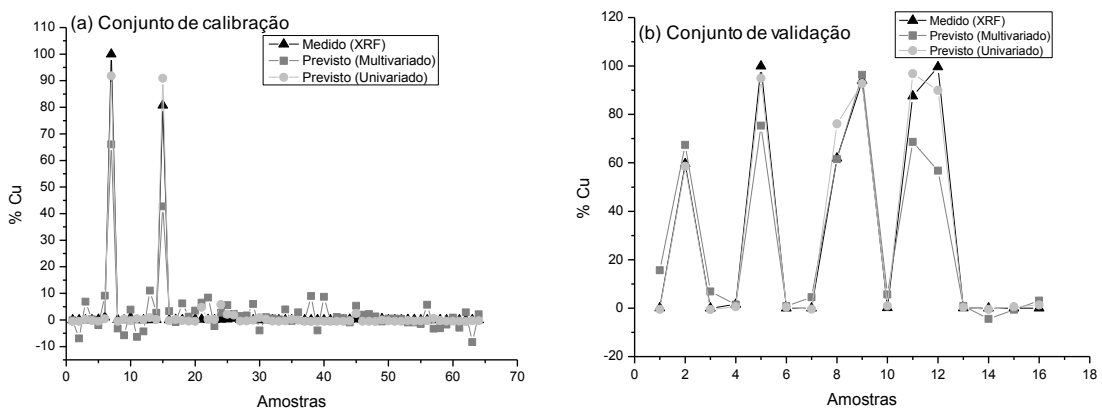


FIGURA 5.3. 13. Valores de referência (medido) e os valores previstos (multivariado e univariado) para o Cu. (a) Conjunto de calibração e (b) Conjunto de validação.

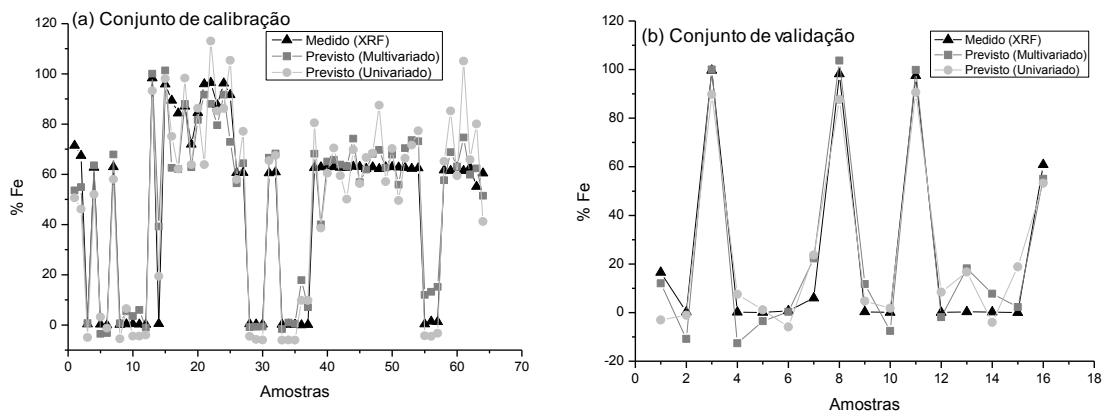


FIGURA 5.3. 14. Valores de referência (medido) e os valores previstos (multivariado e univariado) para o Fe. (a) Conjunto de calibração e (b) Conjunto de validação.

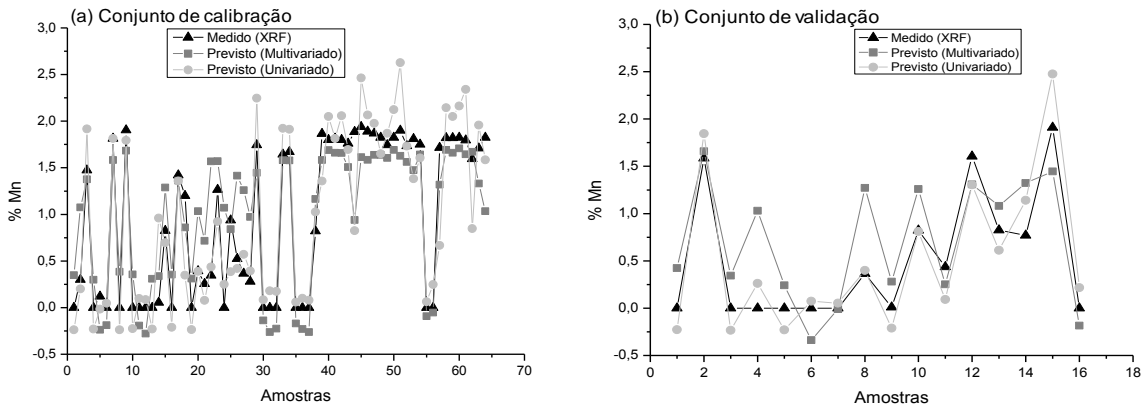


FIGURA 5.3. 15. Valores de referência (medido) e os valores previstos (multivariado e univariado) para o Mn. (a) Conjunto de calibração e (b) Conjunto de validação.

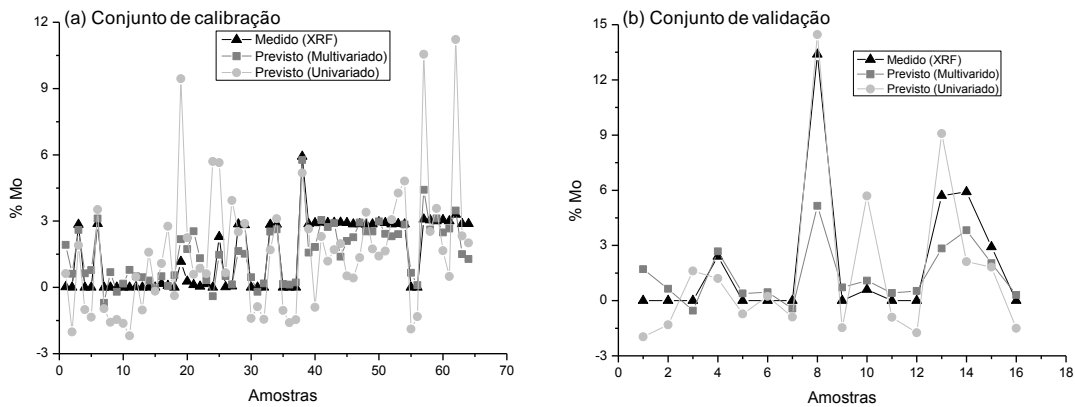


FIGURA 5.3. 16. Valores de referência (medido) e os valores previstos (multivariado e univariado) para o Mo. (a) Conjunto de calibração e (b) Conjunto de validação.

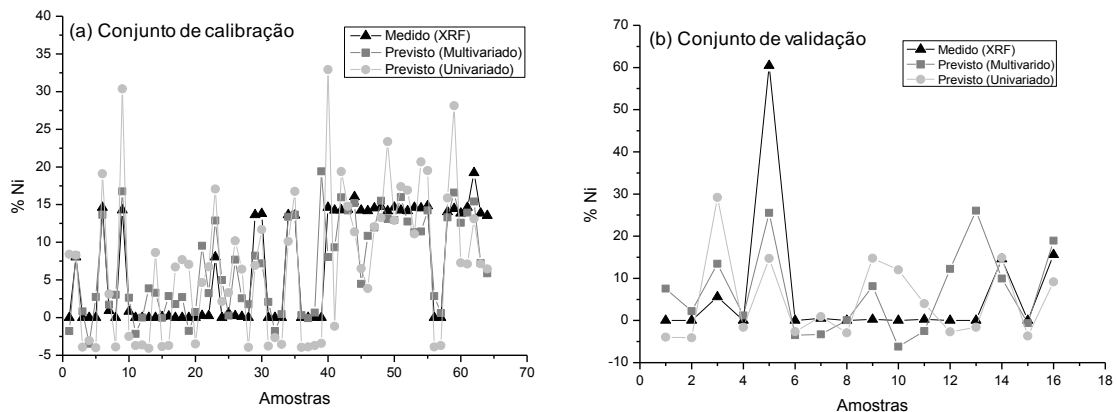


FIGURA 5.3. 17. Valores de referência (medido) e os valores previstos (multivariado e univariado) para o Ni. (a) Conjunto de calibração e (b) Conjunto de validação.

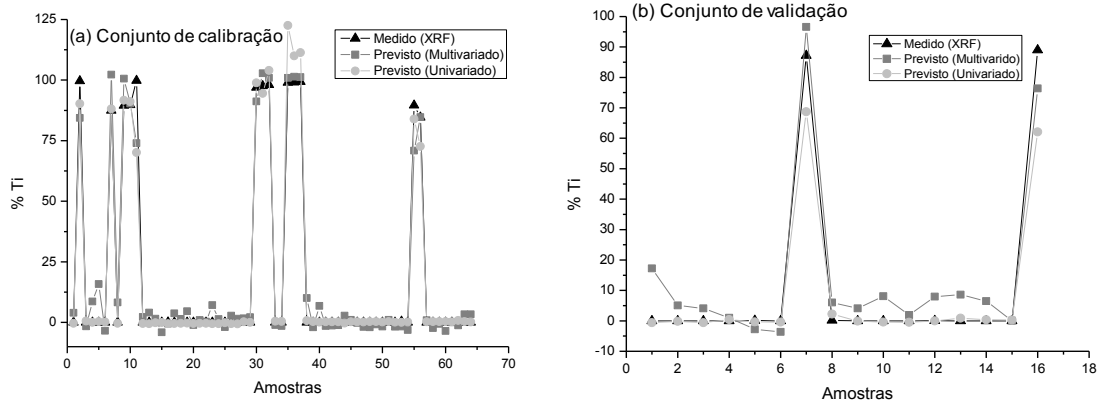


FIGURA 5.3. 18. Valores de referência (medido) e os valores previstos (multivariado e univariado) para o Ti. (a) Conjunto de calibração e (b) Conjunto de validação.

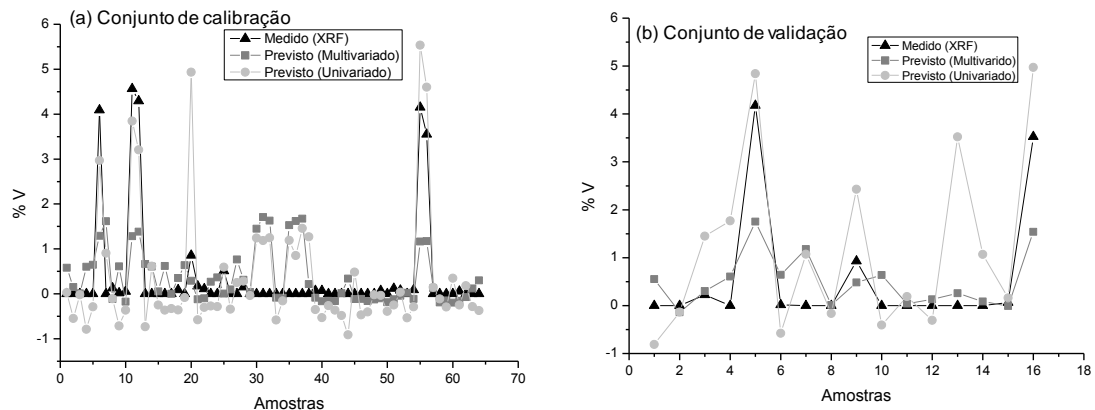


FIGURA 5.3. 19. Valores de referência (medido) e os valores previstos (multivariado e univariado) para o V. (a) Conjunto de calibração e (b) Conjunto de validação.

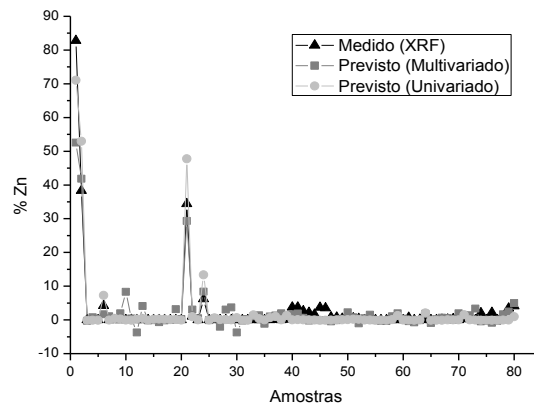


FIGURA 5.3. 20. Valores de referência (medidos) e os valores previstos (multivariado e univariado) para o Zn.

Como pode ser observado nas FIGURAS 5.3.11 a 5.3.20, os valores previstos pelos modelos univariados e multivariados foram concordantes com aqueles medidos por XRF, obtendo uma boa correlação (R^2) entre os mesmos variando de 0,42 (V) a 0,99 (Ti) para calibração multivariada e 0,23 (Ni) a 0,99 (Al) para a calibração univariada. Assim, foi útil utilizar a técnica de XRF portátil para obter os valores de referência dos analitos nas amostras, pois a análise das ligas metálicas foi direta como na

LIBS (sem a necessidade de tratamento das mesmas). Para o Zn, não foi possível dividir em conjunto de calibração e conjunto de validação devido a poucas amostras com concentração de Zn.

Portanto, ao considerar todos os parâmetros dos modelos, a calibração univariada utilizando a área ou altura do sinal geralmente apresentou melhores resultados do que a calibração multivariada, principalmente quando os valores de SEV são considerados.

Para avaliar se o primeiro conjunto de dados selecionado (rotina Kennard-Stone) foi obtido por acaso, selecionou-se 20 conjuntos distintos de dados (calibração e validação) aleatórios. Assim, calculou-se 20 modelos PLS para o Cr a fim de verificar se haveria diferença entre os valores de SEV, como é mostrado na FIGURA 5.3.21.

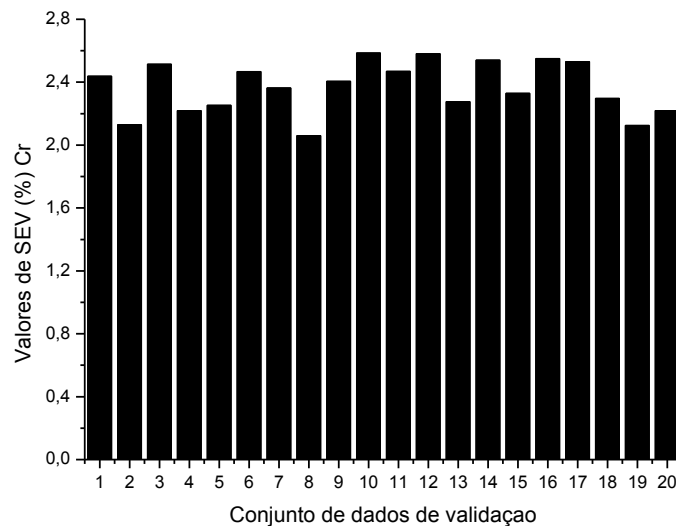


FIGURA 5.3. 21. Valores de SEV para o Cr em diferentes conjuntos de dados.

Observou-se que não houve diferença de um conjunto de dados para o outro, sendo muito semelhantes.

Um outro parâmetro avaliado foi o desvio padrão relativo (*Relative Standard Deviation, RSD*) para avaliar a repetibilidade dos dados. Escolheu-se diferentes amostras e selecionou-se as linhas de emissão dos analitos estudados. O cálculo foi feito a partir da média e desvio padrão dos espectros. A FIGURA 5.3.22 apresenta os valores para a altura do sinal.

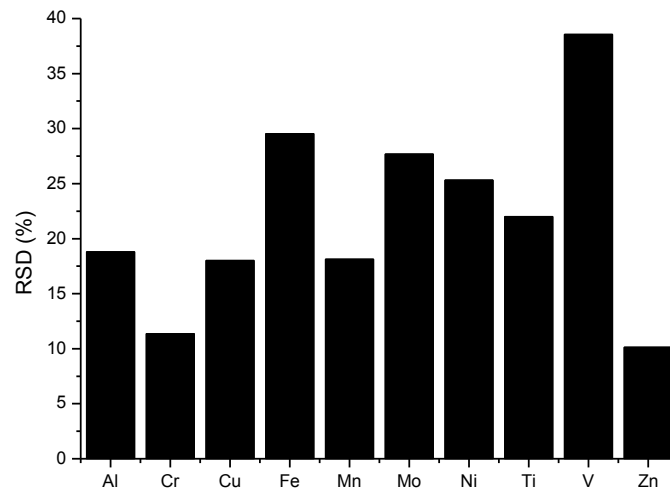


FIGURA 5.3. 22. Valores de RSD (%) para os diferentes analitos estudados.

Os valores de RSD variaram desde 10% para o Zn (concentração do analito de 83%) até 40% para o V (concentração do analito de 4%). Assim, quanto menor o valor de RSD, maior é a repetibilidade dos dados. Espera-se que os maiores valores de RSD são para os analitos que possuem baixas concentrações, porém isso não foi observado em todos os casos. Por exemplo, o Fe tem uma alta concentração nas amostras, porém apresentou valor de RSD da ordem de 28%.

5.4- Modelos de classificação

Nas FIGURAS 5.4.1 a 5.4.3 tem-se os espectros de 3 amostras pertencentes as 3 diferentes classes (detalhes na sessão 4.5).

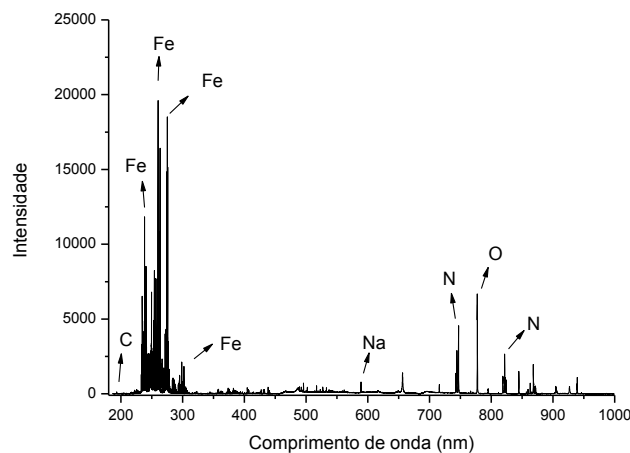


FIGURA 5.4. 1. Espectro característico da amostra 3 pertencente a classe 1.

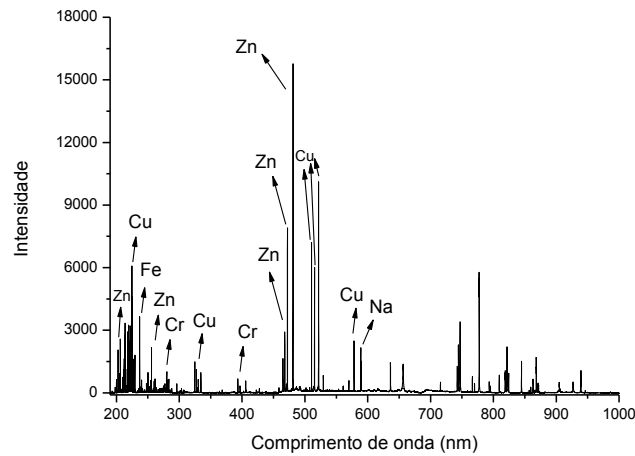


FIGURA 5.4. 2. Espectro característico da amostra 2 pertencente a classe 2.

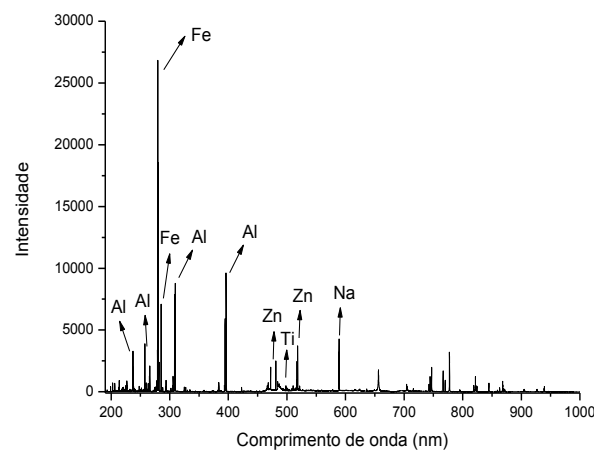


FIGURA 5.4. 3. Espectro característico da amostra 6 pertencente a classe 3.

O espectro da amostra 3, possui basicamente Fe e C em sua composição, sendo pertencente a classe 1 com concentração de Fe entre 71 a 99%. O espectro da amostra 2, além do Fe e C, possui Cr, Cu e Zn, sendo pertencente a classe 2 com concentração de Fe entre 16 a 67%. Já no espectro da amostra 6, por ser uma liga de Al, possui uma grande quantidade de Al e Fe entre 0 a 6%, sendo pertencente a classe 3.

Três conjuntos de dados foram selecionados para testar os modelos de classificação: concentrações do XRF, área e altura do sinal com todas as linhas de emissão estudadas (57 linhas de emissão) nos modos de normalização selecionados. Três modelos foram estudados: KNN, SIMCA e PLS-DA, em todos os casos, as matrizes foram auto escaladas.

Para o KNN, um vizinho foi selecionado para as 3 matrizes. Para o SIMCA, o número de componentes principais variou de 2 a 3 para as 3 classes e para o PLS-DA, uma variável latente foi selecionada para todos os casos.

Nas FIGURAS 5.4.4 e 5.4.5, tem-se as figuras de mérito relativas a exatidão, sensibilidade, taxa de falso alarme e especificidade (Equações 9 a 12), do conjunto de validação (16 amostras) para as 3 matrizes estudadas.

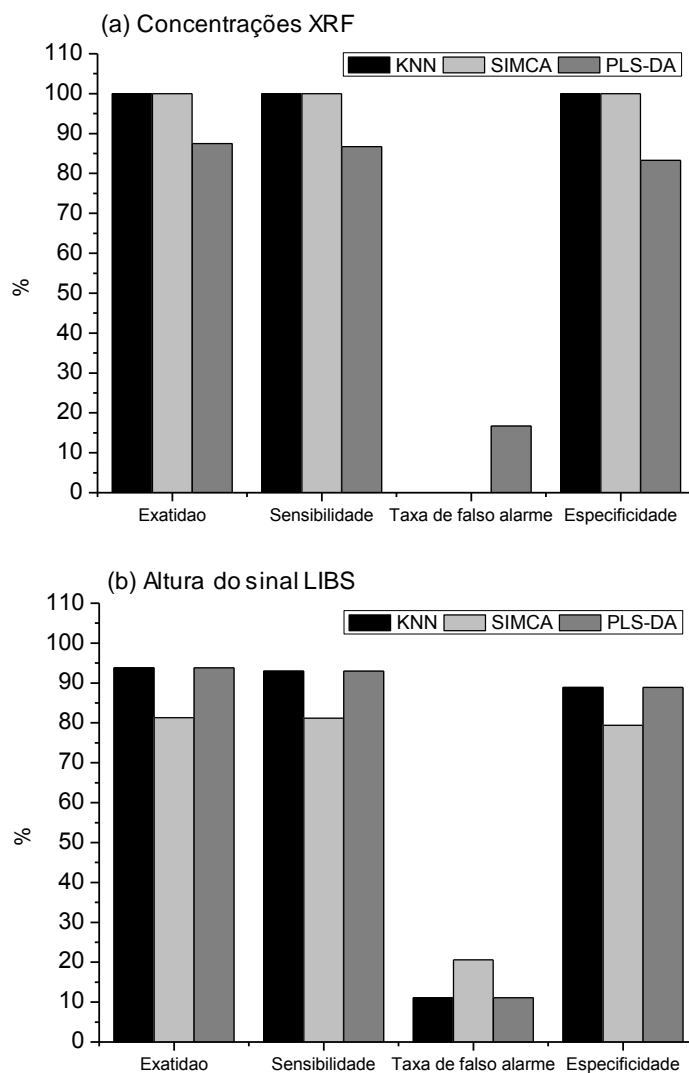


FIGURA 5.4. 4. Exatidão, sensibilidade, taxa de falso alarme e especificidade para os modelos de classificação (KNN, SIMCA e PLS-DA) para as matrizes de (a) Concentrações do XRF e (b) Altura do sinal LIBS.

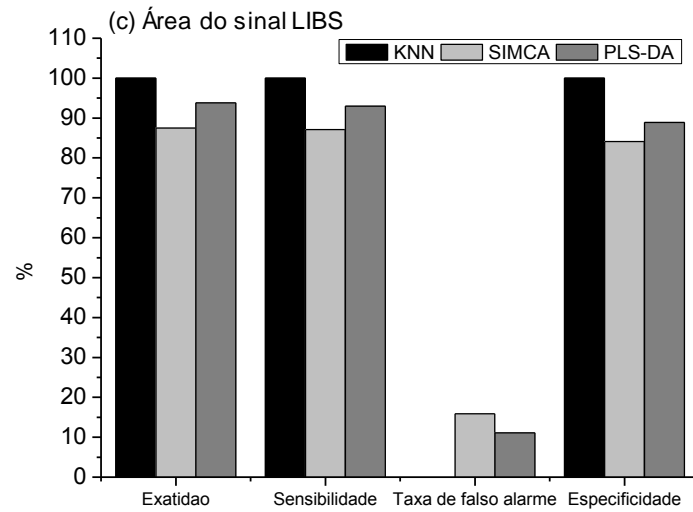


FIGURA 5.4. 5. Exatidão, sensibilidade, taxa de falso alarme e especificidade para os modelos de classificação (KNN, SIMCA e PLS-DA) para as matrizes de (c) Área do sinal LIBS.

Os modelos de KNN, SIMCA e PLS-DA foram eficientes para classificar as 80 amostras nas 3 diferentes classes. Entre eles, o KNN apresentou melhores resultados com exatidão de 100% para a matriz de concentração do XRF e da área do sinal.

6- Conclusão

Com essa dissertação de mestrado, concluiu-se que a utilização da técnica LIBS foi muito útil para análise direta de ligas metálicas. As etapas relacionadas com o preparo de amostra foram totalmente eliminadas sendo possível acomodar amostras de diferentes tamanhos e formas na porta amostra do equipamento.

Em comparação a técnicas de espectrometria (ICP OES, ICP MS, entre outras), a LIBS apresentou vantagens: não houve necessidade de preparo da amostra, não utilizou nenhum tipo de gás (Ar ou He) e não há necessidade de transportar a amostra ao plasma ou chama, pois o plasma é formado sobre a amostra pela utilização de um feixe de laser focalizado. Na indústria é muito utilizada a técnica de espectrometria de emissão com centelha elétrica ou arco. Essa técnica apresenta alta frequência de amostragem, sensibilidade e facilidade de operação. Porém, as amostras têm que ser sólidas e condutoras de eletricidade. Ao comparar LIBS e XRF, com a LIBS não houve problemas com a determinação de elementos leves e analitos com concentrações inferiores a 1%.

A combinação da LIBS com estratégias quimiométricas foi uma boa alternativa para propor modelos de regressão confiáveis. Além disso, com o uso de planejamento fatorial foi possível obter uma condição de compromisso para todos os analitos e amostras simultaneamente.

Essa dissertação também mostrou a aplicação de diferentes normalizações de sinal, as quais foram úteis para minimizar as diferenças entre as matrizes das amostras. Os modelos de calibração univariado foram mais eficientes quando comparados com os multivariados, porém não houve diferença significativa ao nível de 95% de confiança. E com os modelos de classificação, foi possível identificar as amostras nas 3 diferentes classes, sendo que o KNN mostrou um melhor desempenho. Essa dissertação de mestrado gerou um artigo científico que é apresentado no Anexo 3.

7- Referências

1. SOUZA, S. A.; "Composição química dos aços". 1ª ed., São Paulo, Editora Edgard Blücher LTDA, 1989.
2. CANTO, E.; "Minerais, Minérios e Metais - de onde vêm ? Para onde vão ?". 1ª ed., São Paulo, Editora Moderna, 1996.
3. Site do Instituto Aço Brasil. Disponível em: <<http://www.acobrasil.org.br/site2015/>>. Acessado em 17/10/2016.
4. DA-COL, J. A. "Uma breve visão sobre a determinação de metais em ligas metálicas e matrizes semelhantes: alguns métodos clássicos e espectroscópicos modernos". *Rev. Virtual Quim.*, 6 (4): 989, 2014.
5. PEREIRA-FILHO, E. R.; PEREIRA, F. M. V. & LEMOS, S. G. "Introducing factorial design concepts using experiments with multi-colored chocolates". *Chem. Educator*, **14**: 200, 2009.
6. PEREIRA-FILHO, E. R. "Planejamento Fatorial em Química". 1ª ed., São Carlos, Editora EdUSCar, 2015.
7. FERREIRA, S. L. C.; BRUNS, R. E.; FERREIRA, H. S.; MATOS, G. D.; DAVID, J. M.; BRANDÃO, G. C.; DA SILVA, E. G. P.; PORTUGAL, L. A.; DOS REIS, P. S.; SOUZA, A. S. & DOS SANTOS, W. N. L. "Box Behnken design: an alternative for the optimization of analytical methods". *Anal. Chim. Acta*, **597**: 179, 2007.
8. BARROS NETO, B.; SCARMINIO, I. S. & BRUNS, R. E. "Como fazer experimentos". 4ª ed., Porto Alegre, Artmed Editora S.A., 2010.
9. FERREIRA, S. L. C.; DOS SANTOS, W. N. L.; QUINTELLA, C. M.; NETO, B. B. & BOSQUE-SENDRA, J. M. "Doehlert matrix: a chemometric tool for analytical chemistry - review". *Talanta*, **63**: 1061, 2004.
10. NUNES, L. C.; DA SILVA, G. A.; TREVIZAN, L. C.; SANTOS JÚNIOR, D.; POPPI, R. J. & KRUG, J. F. " Simultaneous optimization by neuro-genetic approach for analysis of plant materials by Laser induced breakdown spectroscopy". *Spectrochim. Acta - Part B*, **64**: 565, 2009.
11. ANDRADE, D. F. & PEREIRA-FILHO, E. R. "Direct determination of contaminants and major and minor nutrients in solid fertilizers using Laser-induced breakdown spectroscopy (LIBS)". *J. Agric. Food Chem.*, **64**: 7890, 2016.
12. ANDRADE, D. F.; PEREIRA-FILHO, E. R. & KONIECZYNSKI, P. " Comparison of ICP OES and LIBS analysis of medicinal herbs rich in flavonoids from Eastern Europe". *J. Braz. Chem. Soc.*, **0**: 1, 2016.
13. FERREIRA, M. M. C. "Quimiometria. Conceitos, Métodos e Aplicações". 1ª ed., Campinas, Editora UNICAMP, 2015.
14. POŘÍZKA, P.; KLUS, J.; HRDLIČKA, A.; VRÁBEL, J.; ŠKARKOVÁ, P.; PROCHAZKA, D.; NOVOTNÝ, J.; NOVOTNÝ K. & KAISER, J. " Impact of Laser-induced breakdown spectroscopy data normalization on multivariate classification accuracy". *J. Anal. At. Spectrom.*, 2016. DOI: 10.1039/c6ja00322b.
15. KLUS, J.; Pořízka, P.; PROCHAZKA, D.; NOVOTNÝ, J.; NOVOTNÝ, K. & KAISER, J. " Effect of experimental parameters and resulting analytical signal statistics in Laser-induced breakdown spectroscopy". *Spectrochim. Acta - Part B*, **126**: 6, 2016.
16. DE SOUZA, A. M. & POPPI, R. J. "Experimento didático de quimiometria para análise exploratória de óleos vegetais comestíveis por espectroscopia no infravermelho médio e análise de componentes principais: um tutorial, parte I". *Quim. Nova*, 35 (1): 223, 2012.
17. ZAYTSEV, S. M.; POPOV, A. M.; CHERNYKH, E. V.; VORONINA, R. D.; ZOROV, N. B. & LABUTIN, T. A. "Comparison of single- and multivariate calibration for determination of Si, Mn, Cr and Ni in high-alloyed stainless steels by Laser-induced

- breakdown spectrometry". *J. Anal. At. Spectrom.*, **29**: 1417, 2014.
18. SARKAR, A.; KARKI, V.; AGGARWAL, S. K.; MAURYA, G. S.; KUMAR, R.; RAI, A. K.; MAO, X. & RUSSO, R. E. "Evaluation of the prediction precision capability of partial least squares regression approach for analysis of high alloy steel by Laser induced breakdown spectroscopy". *Spectrochim. Acta - Part B*, **108**: 8, 2015.
 19. VALDERRAMA, P.; BRAGA, J. W. B. & POPPI, R. J. "Estado da arte de figuras de mérito em calibração multivariada". *Quim. Nova*, **32** (5): 1278, 2009.
 20. ANDRADE, J. M.; CRISTOFORETTI, G.; LEGNAIOLI, S.; LORENZETTI, G.; PALLESCHI, V. & SHALTOUT, A. A. "Classical univariate calibration and partial least squares for quantitative analysis of brass samples by Laser-induced breakdown spectroscopy". *Spectrochim. Acta - Part B*, **65**: 658, 2010.
 21. VORS, E.; TCHEPIDJIAN, K. & SIRVEN, J.-B. "Evaluation and optimization of the robustness of a multivariate analysis methodology for identification of alloys by Laser induced breakdown spectroscopy". *Spectrochim. Acta - Part B*, **117**: 16, 2016.
 22. GONZAGA, F. B. & PASQUINI, C. "A complementary metal oxide semiconductor sensor array based detection system for Laser induced breakdown spectroscopy: evaluation of calibration strategies and application for manganese determination in steel". *Spectrochim. Acta - Part B*, **63**: 56, 2008.
 23. STIPE, C. B.; HENSLEY, B. D.; BOERSEMA, J. L. & BUCKLEY, S. G. "Laser-induced breakdown spectroscopy of steel: a comparison of univariate and multivariate calibration methods". *Appl. Spectrosc.*, **64** (2): 154, 2010.
 24. NOLL, R.; "Laser-Induced Breakdown Spectroscopy. Fundamentos e Aplicações". 1ª ed., Alemanha, Editora Springer, 2012.
 25. GONZAGA, F. B. & PASQUINI, C. "A compact and low cost laser induced breakdown spectroscopic system: application for simultaneous determination of chromium and nickel in steel using multivariate calibration". *Spectrochim. Acta - Part B*, **69**: 20, 2012.
 26. NOLL, R.; BETTE, H.; BRYSCH, A.; KRAUSHAAR, M.; MÖNCH, I.; PETER, L. & STURM, V. "Laser-induced breakdown spectrometry - applications for production control and quality assurance in the steel industry". *Spectrochim. Acta Part B*, **56**: 637, 2001.
 27. GUPTA, G. P.; SURI, B. M.; VERMA, A.; SUNDARARAMAN, M.; UNNIKRISHNAN, V. K.; ALTI, K.; KARTHA, V. B. & SANTHOSH, C. "Quantitative elemental analysis of nickel alloys using calibration-based Laser-induced breakdown spectroscopy". *J. Alloys Compd.*, **509**: 3740, 2011.
 28. ORTIZ, M. C.; SARABIA, L.; JURADO-LÓPEZ, A. & LUQUE DE CASTRO, M. D. "Minimum value assured by a method to determine gold in alloys by using Laser-induced breakdown spectroscopy and partial least-squares calibration model". *Anal. Chim. Acta*, **515**: 151, 2004.
 29. HAO, Z.; GUO, L.; LI, C.; SHEN, M.; ZOU, X.; LI, X.; LU, Y. & ZENG, X. "Sensitivity improvement in the detection of V and Mn elements in steel using Laser-induced breakdown spectroscopy with ring-magnet confinement". *J. Anal. At. Spectrom.*, **29**: 2309, 2014.
 30. ZHANG, T.; LIANG, L.; WANG, K.; TANG, H.; YANG, X.; DUAN, Y. & LI, H. "A novel approach for the quantitative analysis of multiple elements in steel based on Laser-induced breakdown spectroscopy (LIBS) and random forest regression (RFR)". *J. Anal. At. Spectrom.*, **29**: 2323, 2014.
 31. BOSCO, G. L. "Development and application of portable, hand-held X-ray fluorescence spectrometers". *TrAC - Trends Anal. Chem.*, **45**: 121, 2013.
 32. KENNARD, R. W. & STONE, L. A. "Computer Aided Design of Experiments". *Technometrics*, **11** (1): 137, 1969.
 33. ALLEGRI, F. & OLIVIERI, A. C. "IUPAC-Consistent approach to the limit of

- detection in partial least-squares calibration". *Anal. Chem.*, **86**: 7858, 2014.
34. BROWN, S. D.; TAULER, R. & WALCZAK, B. "Comprehensive Chemometrics". Amsterdam, Editora Elsevier, 2009.

Anexos

Anexo 1 - Função no Matlab para a correção da linha de base e cálculos da área e altura dos sinais selecionados.

```

function [parameters]=libs_par(data,address>window>window_n)
% [parameters]=libs_par(data,address>window>window_n)
% You will need a file (data) where the signals start at line 4
% and column 1 has the wavelengths
% parameters = sb; area; height
% data = dataset with signal (the three first lines must be considered)
% address = address of the wavelength
% window = window for the wavelength (must be an odd number)
% window_n = window for the noise

% sb calculation
X=data(4:end,2:end)';
[m,n]=size(X);
ruido_avg=mean(X(:,[address-((window-1)/2):((address-((window-1)/2))+window_n-
1]),((address+((window-1)/2)-(window_n-1)):address+((window-1)/2)]),2);
altura=max(X(:,address-((window-1)/2):address+((window-1)/2))');
sb=(altura./(((ruido_avg).^2).^0.5))';

% BOC (Baseline offsite correction)
X_teste2=X(:,address-((window-1)/2):address+((window-1)/2))-
(ruido_avg*ones(1>window));

% Area calculation
area = sum(X_teste2,2);

% Height calculation
height = max(X_teste2)';

parameters(:,1)=sb;
parameters(:,2)=area;
parameters(:,3)=height;

figure(1)
clf

```

```
plot (X(:,address-((window-1)/2):address+((window-1)/2)),'-or')
title('Original data')
xlabel('Wavelengths');
ylabel('Intensity');
```

```
figure(2)
clf
plot (X_teste2,'-ob')
title('BOC data')
xlabel('Wavelengths');
ylabel('Intensity');
```

```
figure(3)
clf
subplot (3,1,1)
plot (sb,'-og')
title('S/B')
xlabel('Samples');
ylabel('S/B');
```

```
subplot (3,1,2)
plot (area,'-ok')
title('Area')
xlabel('Samples');
ylabel('Area');
```

```
subplot (3,1,3)
plot (height,'-om')
title('Height')
xlabel('Samples');
ylabel('Height');
```

Anexo 2 – Função no Matlab para o cálculo dos 12 modos de normalização.

```

function [X_out]=libs_treat(data);
% Function [data_out]=libs_treat(data);
% What to do before:
% Right click in the txt file and "import data"
% or
% open('nome.xlsx'). If you have an Excel file;
% What to do after:
% xlswrite('nome_final',data_out');
% clear data
% clear textdata
% output order: X_out (raw data)
% mean raw data
% mean after normalization by norm, area and maximum
% sum data
% sum after normalization by norm, area and maximum
% output order: X_outC193
% mean raw normalized by C193 data
% sum data
% output order: X_outC247
% mean raw normalized by C247 data
% sum data

%var_name=data(11:end,1); % Works with Excel only.
%var_name = textdata(11:end,1); % Work with txt files only.

X=data(11:end,1:end)';

[m,n]=size(X);

teste3=std(X'); % teste3 has the standard deviation of the data

% Normalization by norm
X_n_n=normr(X);

```

```

% Normalization by area
teste=(sum(X,2))*ones(1,n); %matrix with areas.

X_n_a=X./teste;

% Normalization by maximum
teste2=max(X');
teste2=teste2';
teste2=teste2*ones(1,n);

X_n_m=X./teste2; % normalization by maximum

for i =1:m;
teste4(i)=norm(X(i,:));%teste4 has the individual norms
teste4=teste4';
end

figure
subplot(2,2,1)
plot (teste3,'-ob') % Graphic of the standard deviation
title('Standard Deviation')
xlabel('Samples');
ylabel('Standard deviation');
subplot (2,2,2)
plot (teste,'-or') % Graphic of the sum
title('Area')
xlabel('Samples');
ylabel('Area');
subplot(2,2,3)
plot (teste2,'-ok') % Graphic of the maximum
title('Maximum')
xlabel('Samples');
ylabel('Maximum');

```



```
subplot(2,2,4)
plot (teste4,'-og') % Graphic of the maximum
title('Norm')
xlabel('Samples');
ylabel('Norm');

% Finding the maximum for C 193 nm
carbon193_raw=max(X(:,91:95));
carbon193_raw=carbon193_raw';
carbon193_raw=carbon193_raw*ones(1,n);
X_r_C193=X./carbon193_raw;

% Finding the maximum for C 247 nm
carbon247_raw=max(X(:,935:937));
carbon247_raw=carbon247_raw';
carbon247_raw=carbon247_raw*ones(1,n);
X_r_C247=X./carbon247_raw;

% Output1
X_out(1,:)=mean(X);
X_out(2,:)=mean(X_n_n);
X_out(3,:)=mean(X_n_a);
X_out(4,:)=mean(X_n_m);
X_out(5,:)=sum(X,1);
X_out(6,:)=sum(X_n_n);
X_out(7,:)=sum(X_n_a);
X_out(8,:)=sum(X_n_m);
X_out(9,:)=mean(X_r_C193);
X_out(10,:)=sum(X_r_C193,1);
X_out(11,:)=mean(X_r_C247);
X_out(12,:)=sum(X_r_C247,1);
X_out=X_out';
```

Anexo 3 – Artigo publicado no Journal of Analytical Atomic Spectrometry, 2016, 31, 2005-2014.

CrossMark
click for updates

Cite this: DOI: 10.1039/c6ja00224b

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*

This study applies laser-induced breakdown spectroscopy (LIBS) for the direct analysis of 80 metal samples (alloys and steel) for multivariate and univariate regression models, aiming at the determination of 10 analytes (Al, Cr, Cu, Fe, Mn, Mo, Ni, Ti, V and Zn). To optimize the LIBS system, the Doehlert design was used for energy, delay time and spot size adjustment for all samples and analytes. Twelve normalization modes were used to reduce the interference matrix and to improve the calibration models, with error values ranging from 0.27% (Mn) to 14% (Cu and Ni). Models without normalization presented two- to five-fold higher errors. In addition to quantification, classification models (KNN, SIMCA and PLS-DA) were also proposed for sample differentiation. Multivariate and univariate models presented similar performance, and among the classification models, KNN presented the best results, with an accuracy of 100%.

Received 23rd June 2016
Accepted 6th July 2016

DOI: 10.1039/c6ja00224b

www.rsc.org/jaas

1. Introduction

Laser-induced breakdown spectroscopy (LIBS) is a powerful technique, in which solid samples are directly analyzed, and emission line signals are obtained in a short period of time. In general, a LIBS system comprises an energy source (a laser that generates a pulse of high energy radiation) and lenses that focus the laser pulses onto the surface of the sample. In this process, analytes are vaporized and excited, and optical fibers collect the light of plasma, which is registered by a spectrometer, and diffract the plasma emitted light. Finally, the light is detected by, for example, charged coupled devices (CCD) and a computer to process the spectra.^{1,2}

This technique can be used for multivariate or univariate calibration or classification purposes (fingerprint) in direct analysis of a myriad of solid samples, *e.g.*, soils and plastics. On the other hand, problems related to matrix interference, low availability of solid calibration standards with certified values for masses in the μg range, and adequate repeatability are severe. In this case, data pre-processing is necessary to compensate for different interactions between the laser and the sample matrices.^{3,4}

There are several applications of LIBS that confirm its versatility, such as determination of metals in soils,⁵ liquid samples,⁶ polymers,⁷ electric and electronic waste,⁸ metallic alloys,^{9–12} and real time applications in the steel industry.^{13–15}

Alloy and metal analyses are effective in many cases with X-ray fluorescence (XRF), which is an already consolidated technique that presents several advantages for direct fast analysis of solid samples using portable or hand-held equipment.¹⁶ On the other hand, LIBS can also be an alternative because the laser can be focused directly onto slag in the steel industry, and standoff analysis can be performed.¹⁷ Moreover, light elements, such as C, can be easily detected.

Despite the high concentration of analytes, steel and alloy analyses are difficult due to the complexity of the emission spectra obtained (overlap of emission lines) and normalization is mandatory in several cases.¹⁸ Thus, the purpose of this study is to present LIBS as a viable alternative for the direct analysis of alloys and metals using chemometric tools to interpret the amount of spectroscopic data obtained. Twelve normalization modes are compared, and their abilities to improve the accuracy and precision of calibration models (multivariate or univariate), noise compensation, sample heterogeneity and to correct matrix effects are discussed. Actually, signal normalization in laser techniques was already discussed by Sarkar *et al.*¹⁹ and the authors proposed 17 different approaches in laser ablation molecular isotopic spectrometry (LAMIS).

In the present study, the LIBS experimental conditions are optimized using factorial design²⁰ and ten analytes are

Group of Applied Instrumental Analysis, Chemistry Department, Federal University of São Carlos, P. O. Box 676, São Carlos, São Paulo State, Brazil, 13565-905. E-mail: erpf@ufscar.br; Fax: +55 16 3351 8350; Tel: +55 16 3351 8092

[†] Electronic supplementary information (ESI) available. See DOI: 10.1039/c6ja00224b

determined (Al, Cr, Cu, Fe, Mn, Mo, Ni, Ti, V and Zn). Multivariate and univariate calibration strategies²¹ are compared, and classification models are proposed for fast sample identification.

2. Experimental

2.1 Samples

In this study, 80 metallic samples were used in the proposition of classification and regression models (univariate and multivariate). Table 1 shows a description of these samples, which are categorized according to the Fe composition: group 1 has sixteen steel samples, with the Fe composition varying from 71 to 99% (w/w); group 2 has thirty-three steel samples, with the Fe concentration ranging from 16 to 67%; group 3 has thirty-one alloy samples (Hastelloy, brass, steel, aluminum, bronze, copper, titanium and some of them containing Co, Cr and Mo), with the Fe concentration ranging from lower than the limit of detection to 6%.

2.2 LIBS system

Commercial LIBS equipment (Applied Spectra, Fremont – CA, USA, model J200) was used for direct solid sample analysis of metallic alloys (see details in Table 1). This instrument comprises an Nd:YAG laser emitting at a fundamental wavelength (1064 nm). The laser energy can be varied from 0 to 100 mJ, and the delay time ranged from 0 to 2 μ s. In the optical system, the laser beam goes through turning mirrors and a beam expander and then goes through an objective lens (5 times objective lens). This lens has a clear aperture of approximately 10 mm and a working distance of around 35 mm. The spectrometer has 6 channels, with a resolution from 0.08–0.1 nm in channels 1 (186.940 to 311.195 nm) to 4 (591.426 to 693.690 nm) and a resolution of 0.11–0.14 nm in channels 5 (693.807 to 884.407 nm) and 6 (884.509 to 1042.026 nm). A CCD records spectral information from 186 to 1042 nm. During the spectrum acquisition, the spot size could be changed from 50 to 250 μ m. The pulse duration was approximately 10 ns, and the acquisition time of the signal was fixed at 1.05 ms.

2.3 XRF systems (reference value acquisition)

To obtain analyte reference concentration values, XRF, which employs little or no sample preparation, was selected. In this case, the reference concentration values of 10 analytes (Al, Cr, Cu, Fe, Mn, Mo, Ni, Ti, V and Zn) were obtained using two different instruments: Innov-X alfa 6500 model and Niton XL3T

985 ULTRA of Thermo Scientific. With the former, it was possible to detect light elements (in the case of Al). Both instruments had a library with the composition of several alloy types, that also helped in sample identification.

2.4 LIBS parameter optimization

In the LIBS system used, it was possible to adjust three parameters: delay time, laser energy and spot size. These three variables play an important role in the quality of the spectral information acquired. In this case, a Doehlert design²² was used; Table 2 shows the levels studied. With a Doehlert design, the variables can be tested at different levels simultaneously. The first variable (delay time) was tested at 5 different levels from 0 (coded as –1) to 2 μ s (coded as 1). The laser energy (mJ) plays an important role in the signal intensity and was tested at 7 levels, ranging from 30 (coded as –0.866) to 80 mJ (0.866). The last variable, spot size (μ m), was tested at 3 different levels (from 50 μ m to 150 μ m). The first three experiments in Table 2 have the following working conditions: delay time of 1 μ s (coded in 0), laser energy of 55 mJ (coded in 0) and spot size of 100 μ m (also coded in 0). These three experiments represent the central point of the Doehlert design and they were used to estimate the pure error (PE) of the models proposed.

Nine samples of the 3 different groups (see Table 1) were chosen for the optimization part. In the beginning, 100 spectra

Table 2 Doehlert design used for delay time, laser energy and spot size experimental condition optimization

Experiments	Delay time (μ s)		Laser energy (mJ)		Spot size (μ m)	
	Coded	Real	Coded	Real	Coded	Real
1	0	1	0	55	0	100
2	0	1	0	55	0	100
3	0	1	0	55	0	100
4	1	2	0	55	0	100
5	0.5	1.5	0.866	80	0	100
6	0.5	1.5	0.289	63	0.817	150
7	–1	0	0	55	0	100
8	–0.5	0.5	–0.866	30	0	100
9	–0.5	0.5	–0.289	47	–0.817	50
10	0.5	1.5	–0.866	30	0	100
11	0.5	1.5	–0.289	47	–0.817	50
12	–0.5	0.5	0.866	80	0	100
13	0	1	0.577	72	–0.817	50
14	–0.5	0.5	0.289	63	0.817	150
15	0	1	–0.577	38	0.817	150

Table 1 Steel and alloy sample descriptions used for multivariate and univariate calibration models and classification

Groups of samples	Number of samples	Remarks	Major components
(1) Fe 71 to 99%	16	Steel	Fe
(2) Fe 16 to 67%	33	Steel	Cr, Fe, Ni and Zn
(3) Fe < LOD to 6%	31	Hastelloy, brass, steel, aluminum, bronze, copper, titanium and Co–Cr–Mo alloy	Al, Cr, Cu, Fe, Ni, Ti and Zn
Total	80		

were recorded in a line (raster) to clean the sample surface. Afterward, 15 experiments (see Table 2) were performed for each sample (100 spectra for each experiment, raster), and signals for Al (I 394.400 nm), Cr (II 283.563 nm), Cu (I 324.754 nm), Fe (II 259.940 nm), Mn (II 294.920 nm), Mo (I 386.411 nm), Ni (I 361.939 nm), Ti (II 334.941 nm), V (I 437.924 nm) and Zn (I 481.053 nm) were recorded. These emission lines represent the most intense wavelengths and were preliminarily used for LIBS optimization conditions.

Iron models were calculated for three different samples (one from group 1 and two from group 2). The spectra obtained for each experiment were averaged, and the signal area and height were calculated.

In the optimization process, 24 regression models were calculated (12 for signal area and 12 for height), and 10 coefficients were considered: b_0 (constant), b_1 , b_2 and b_3 (linear coefficients), b_{11} , b_{22} , and b_{33} (quadratic coefficients), and b_{12} , b_{13} and b_{23} (interaction coefficients). The significance of the models was evaluated using an ANOVA table (analysis of variance), and all valid models proposed with the 95% confidence level were considered to obtain a commitment condition for all samples and analytes simultaneously. Data treatment was performed using Matlab 2009 (The Mathworks, Natick, USA) and Microsoft Excel (mainly for matrices organization). The statistical confidence level used in all cases was 95%. With the models calculated, surface and contour plots were visualized to simultaneously identify compromise operational conditions for all analytes.

2.5 Normalization mode

In this study, twelve different normalization modes were tested to compensate for signal variations and sample matrix differences. The goal was to improve the predictability of the regression models used to calculate analyte concentrations (multivariate and univariate) and the figures of merit (accuracy, sensitivity, false alarm rate and specificity) of the classification models.

Table 1S† shows each normalization mode and a pictorial description of data transformation made in each case. The first row of Table 1S† shows three raw replicate spectra with two carbon emission lines (C I 193.091 and 247.856 nm) and signals of a hypothetical analyte. The normalizations descriptions are as follows:

- (1) Spectra are averaged over the number of measurements;
- (2) Individual norm (vector length) is calculated for each spectrum, the raw data are divided by this value. The norm of the normalized spectrum is now equal to 1. Later, normalized spectra are averaged;
- (3) Individual spectral area (sum of all signals) is calculated, the raw data are divided by this value. The area of the normalized spectrum is 1. Later normalized spectra are averaged;
- (4) Individual spectral maximum (the highest signal) is identified for each spectrum, the raw data are divided by this value. The signal intensity for the highest emission line is 1. Later, normalized spectra are averaged;

(5) Individual spectrum is divided by the signal intensity of C I 193.091 nm. Later normalized spectra are averaged. The resulting signal intensity for C I 193.091 nm is 1;

(6) Individual spectrum is divided by the signal intensity of C I 247.856 nm. Later normalized spectra are averaged. The resulting signal intensity for C I 247.856 nm is 1;

(7) Spectral sum is calculated over the number of measurements;

(8) Individual norm (vector length) is calculated for each spectrum, the raw data are divided by this value. Later the sum of normalized spectra is calculated;

(9) Individual spectral area (sum of all signals) is calculated for each spectrum, the raw data are divided by this value. Later the sum of normalized spectra is calculated;

(10) Individual spectral maximum (the highest signal) is identified for each spectrum, the raw data are divided by this value. Later the sum of normalized spectra is calculated;

(11) Individual spectrum is divided by the signal intensity of C I 193.091 nm. Later the sum of normalized spectra is calculated;

(12) Individual spectrum is divided by the signal intensity of C I 247.856 nm. Later the sum of normalized spectra is calculated.

2.6 Emission line selection

Preliminary partial least squares (PLS) and univariate models were calculated for each normalization mode (previously described), and ten analytes (Al, Cr, Cu, Fe, Mn, Mo, Ni, Ti, V and Zn) were considered. Partial least squares was already applied in LIBS data in several studies dedicated to alloy and steel analysis,²³ classification of slag,²⁴ analysis of fertilizers²⁵ and jewelry characterization.²⁶

In PLS, the total peak profile was used (12 288 variables, emission lines from 186 to 1042 nm), and for univariate models, only the most intense lines for each analyte were preliminary selected (see details in Section 2.4 LIBS parameter optimization). The goal was to identify among the 12 normalization modes (see details in Table 1S, ESI†) those with the lowest standard error of calibration (SEC) for PLS and univariate models and the SE of cross-validation (SECV)^{27,28} for PLS only.

Moreover, through PLS regression vectors, the emission lines of the analytes that influenced most in the analysis predictions were observed and selected. This procedure was used because there are several and strong problems related to spectral interference in the emission spectra.² These emission lines were used to calculate later univariate models using signal area and height.

2.7 Univariate and multivariate calibration models

The dataset was divided into 64 samples for calibration and 16 for validation. This selection was made using the Kennard–Stone²⁹ routine, implemented in Matlab.

The normalization modes were again evaluated in the divided dataset (calibration and validation).

With the normalization mode chosen for each analyte, the multivariate calibration models were also calculated using the PLS tool, but with all variables. The data were mean centered

and cross-validation was used. The parameters observed were the number of latent variables, SEC, SECV for the calibration dataset (64 samples) and standard error of validation (SEV) for the validation dataset (16 samples). With the emission lines selected after observing the most important regression vectors, the same procedure was made for univariate calibration models, and SEC and SEV were calculated for both signal area and height.

2.8 Figures of merit for multivariate, univariate and classification models

A repeatability calculation was performed after obtaining approximately 100 spectra for each sample (raster mode). In this case, the signal/background ratio (SBR), area and height were calculated for several emission lines of the analytes determined. Then, the Relative Standard Deviation (RSD) was obtained.

Particularly for the calculation of limits of detection (LOD), it was necessary to obtain an estimate of the standard deviation (SD) of the blank signal. Then, SD values were calculated from the noise signal surrounding emission lines of the analytes determined (see details in Fig. 1S in the ESI†). For the SD calculation, at least 100 spectra were obtained for each raster (10 rasters).

The limit of detection for univariate models was calculated using the following equation:

$$\text{LOD} = \frac{3 \times \text{SD}_{\text{blank}}}{s} \quad (1)$$

where SD_{blank} is the standard deviation of spectral noise near the emission lines of a specific analyte and s is the slope of the calibration curve.

In the case of multivariate calibration models, the same SD_{blank} values were used, and the sensitivity (SEN) was calculated according to:

$$\text{SEN}_{\text{multivariate}} = \frac{1}{\|b\|} \quad (2)$$

where $\|b\|$ is the norm (vector length) of the regression vector coefficient.³⁰ The new IUPAC (International Union of Pure and Applied Chemistry) definition for the LOD in multivariate calibration was used and minimum (LOD_{min}), maximum (LOD_{max}) and pseudounivariate (LOD_{pu}) LOD values were calculated. The main contribution of this new approach for LOD_{min} and LOD_{max} calculation is the fact that it takes into consideration the variance of the analyte concentration ($\text{var}(y_{\text{cal}})$). This parameter was calculated through replicates of the samples available with 26 degrees of freedom. More details about this IUPAC definition can be found in the paper published by Allegrini and Olivieri.³¹ With this new approach, the analyte is absent and present if the concentration is lower than LOD_{min} and higher than LOD_{max} , respectively.

For classification models, four figures of merit were calculated: accuracy, sensitivity, false alarm rate and specificity.³²

2.9 Classification models

As described in section 2.1, the sample dataset was divided into three groups to calculate classification models (see Table 1).

Three types of models were calculated: K -Nearest Neighbor (KNN), Soft Independent Modeling of Class Analogy (SIMCA) and Partial Least Squares-Discriminant Analysis (PLS-DA)³² in three different matrixes (XRF concentrations, signal area and height). The combination of classification models and LIBS was also presented in the literature for several purposes like explosive detection³³ and classification of geological materials.³⁴

3. Results and discussion

3.1 LIBS parameter optimization

Energy, delay time and spot size were evaluated through the Doehlert design (see description in Table 2). In the specific case of Ti, it was not possible to obtain a model with valid coefficients at a 95% confidence level. Other emission lines (less sensitive) were tested, but no model was obtained.

Table 3 shows all 22 models calculated (11 for area and 11 for height) for 9 samples from the 3 different groups. For aluminum alloys (group 3), *e.g.*, the emission line used was 394.400 nm and two models (signal area and height) are shown. Only significant coefficients and their confidence intervals at a confidence level of 95% are presented after backward elimination. In the case of Al, for instance, the model obtained for area was:

$$\begin{aligned} \text{Area} = & (11\,873 \pm 1427) - (6265\text{DT} \pm 2764) + (8744E \pm 2764) \\ & - (2980\text{SS} \pm 2762) \end{aligned}$$

where DT, E and SS are the delay time, energy and spot size, respectively. Coefficient b_0 was equal to 11 873 and its confidence interval was 1427 (this is a valid coefficient with the 95% confidence level). Table 2S in the ESI† shows the other parameters of the models calculated: sum of square for regression (SSR), residues (SSres), total (SST), pure error (SSPE) and lack of fit (SSLoF). In addition, this table shows also the mean of square and the F values calculated when mean values of the square of regression and residues are compared (test F_1 , MSR/MSres) and when mean values of the square of lack of fit and pure error are compared (test F_2 , MSLoF/MSPE). With these values, the readers can have a general idea about the quality of the models calculated. If F_1 is higher than the tabulated F , a significant model is obtained. On the other hand, the model presents lack of fit if F_2 is higher than the tabulated F . In the case of the Al area model (see Table 2S†), the F_1 value (32.61) is higher than the tabulated one (3.587), but the model presents lack of fit because the F_2 value (19.99) was higher than the tabulated one (19.38). In this particular case, the mean of the square of lack of fit (MSLoF) was used as variance to calculate the confidence interval of the coefficients. The same statistical analysis was made for the models obtained with signal height (see Tables 3 and 3S†). Table 4S (at the ESI†) shows the equations used in the calculation of ANOVA parameters.³⁵

Some models presented only two significant coefficients (in the case of Hastelloy, group 3) for the Ni model (b_0 and b_2 , energy). Other models (in the case of stainless steel, group 2) were more complex, with 6 significant coefficients (interaction among them, in the case of Fe).

Table 3 Models for signal area and height obtained for optimization of laser energy, delay time and spot size

Sample name	Group (see details in Table 1)	Signal identification	Equation for signal area ^b	Equation for signal height ^b
Aluminum alloy	3	Al I 394.400	Area = (11 873 ± 1427) – (6265DT ± 2764) + (8744E ± 2764) – (2980SS ± 2762)	Height = (4068 ± 628) – (1757DT ± 1217) + (2851E ± 1217)
Co–Cr–Mo alloy	3	Cr II 283.563	Area = (5855 ± 1801) – (12 263DT ± 2093) + (9794E ± 2093) + (7407DT ² ± 3543) + (4459E ² ± 3544) – (12 307DT × E ± 4584)	Height = (3585 ± 909) – (5378DT ± 1332) + (4640E ± 1332) + (2400DT ² ± 2226) – (5615DT × E ± 2919)
Unidentified alloy	3	Cu I 324.754	Area = (1286 ± 211) – (1132DT ± 246) + (1425E ± 246) + (492DT ² ± 416) + (938E ² ± 416)	Height = (528 ± 67) – (389DT ± 98) + (541E ± 98) + (338E ² ± 164)
Steel	1	Fe II 259.940	Area = (49 113 ± 5301) – (82 764DT ± 7772) + (45 617E ± 7772) + (58 053DT ² ± 12 984) – (39 363DT × E ± 17 027)	Height = (15 942 ± 1236) – (23 034DT ± 1812) + (13 604E ± 1813) + (13 936DT ² ± 3028) – (10 623DT × E ± 3971)
Stainless steel	2	Fe II 259.940	Area = (17 352 ± 4007) – (14 597DT ± 7760) + 20 056E ± 7761) – (19 932DT × E ± 17 000)	Height = (5481 ± 1243) – (4552DT ± 2407) + (6395E ± 2407) – (6425DT × E ± 5272)
Stainless steel ^a	2	Fe II 259.940	Area = (10 992 ± 2630) – (18 750DT ± 3857) + (12 665E ± 3857) – (5108SS ± 3854) + (13 324DT ² ± 6443) – (12 914DT × E ± 8449)	Height = (3392 ± 888) – (6175DT ± 1031) + (4510E ± 1031) – (1747SS ± 1031) + (4267DT ² ± 1746) + (1932E ² ± 1746) – (4515DT × E ± 2259)
Stainless steel	2	Mn II 294.920	Area = (1256 ± 237) – (2378DT ± 348) + (1595E ± 348) + (1810DT ² ± 581) – (1699DT × E ± 761)	Height = (758 ± 129) – (1279DT ± 189) + (882E ± 189) + (879DT ² ± 316) – (965DT × E ± 415)
Hastelloy	3	Mo I 386.411	Area = (439 ± 162) + (514E ± 314)	Height = (239 ± 49) + (224E ± 96)
Hastelloy	3	Ni I 361.939	Area = (611 ± 265) + (758E ± 513)	Height = (256 ± 83) + (315E ± 162)
6Al 4V titanium alloy	3	V I 437.924	Area = (43 ± 14) + (29E ± 27)	Height = (58 ± 17) + (102E ± 71)
Steel (two layers of Cr)	2	Zn I 481.053	Area = (155 142 ± 13 445) – (90 883DT ± 26 037) + (138 788E ± 26 038) – (114 063DT × E ± 57 039)	Height = (55 418 ± 3525) – (7389DT ± 5169) + (28 240E ± 5169) – (22 439E ² ± 8637)

^a Contour plot in Fig. 1. ^b E = energy, DT = delay time and SS = spot size.

To identify the commitment instrument operational condition, a combination of variable values that offer a high signal area and height and do not saturate the detector (the highest signal intensity was 70 000) was selected. For the two signal types studied (area and height), the models proposed were very similar. Therefore, the Doehlert design was very useful for LIBS optimization, enabling the efficient use of time and resource, producing models having more accuracy and being possible to study the variables and at the same time the interactions among them.

Fig. 1 shows contour plots of the Fe area model proposed for the stainless steel sample (group 2):

$$\begin{aligned} \text{Area} = & (10\,992 \pm 2630) - (18\,750\text{DT} \pm 3857) \\ & + (12\,665E \pm 3857) - (5108\text{SS} \pm 3854) \\ & + (13\,324\text{DT}^2 \pm 6443) - (12\,914\text{DT} \times E \pm 8449) \end{aligned}$$

In Fig. 1a, it is observed that when using high laser energy, high signal area and, consequently, height are obtained. On the other hand, in several cases (see Table 3), a high energy promotes signal saturation (mainly for Zn). The spot size (variable 3) did not have a huge influence on the signals

obtained for the tested samples (see Fig. 1b and c). The delay time, on the other hand, needed to be shorter than 1.0 μs. Finally, the best conditions selected for use in all 80 samples was an energy of approximately 60 mJ, 0.9 μs delay time and 100 μm spot size (see star in Fig. 1). With these conditions, signal saturation or self-absorption was not observed in any sample for all analytes, as presented in Table 4, which shows the SBR, signal area and height for the same samples after optimization. The fluence was around 70 J cm⁻². Sarkar *et al.*³⁶ also applied a similar optimization procedure for Pd, Rh and Ru signal LIBS condition optimization and the best results were obtained with 100 mJ energy and 3.5 μs delay time. It is important to mention that both equipment setups (this study and that presented by Sarkar *et al.*³⁶) are different.

In Table 4, the first row, for instance, shows the results for Al in an aluminum alloy. The area range observed in the laboratory varied from 6421 to 17 165 (average of 10 627), and the predicted area (using the model described in Table 3) range with the 95% confidence level was from 11 990 to 14 894. As can be observed, there is an overlap between both ranges. The same tendency was noted for height and in all cases, the number of spectra obtained was at least 100 (*n* = 100). For the other

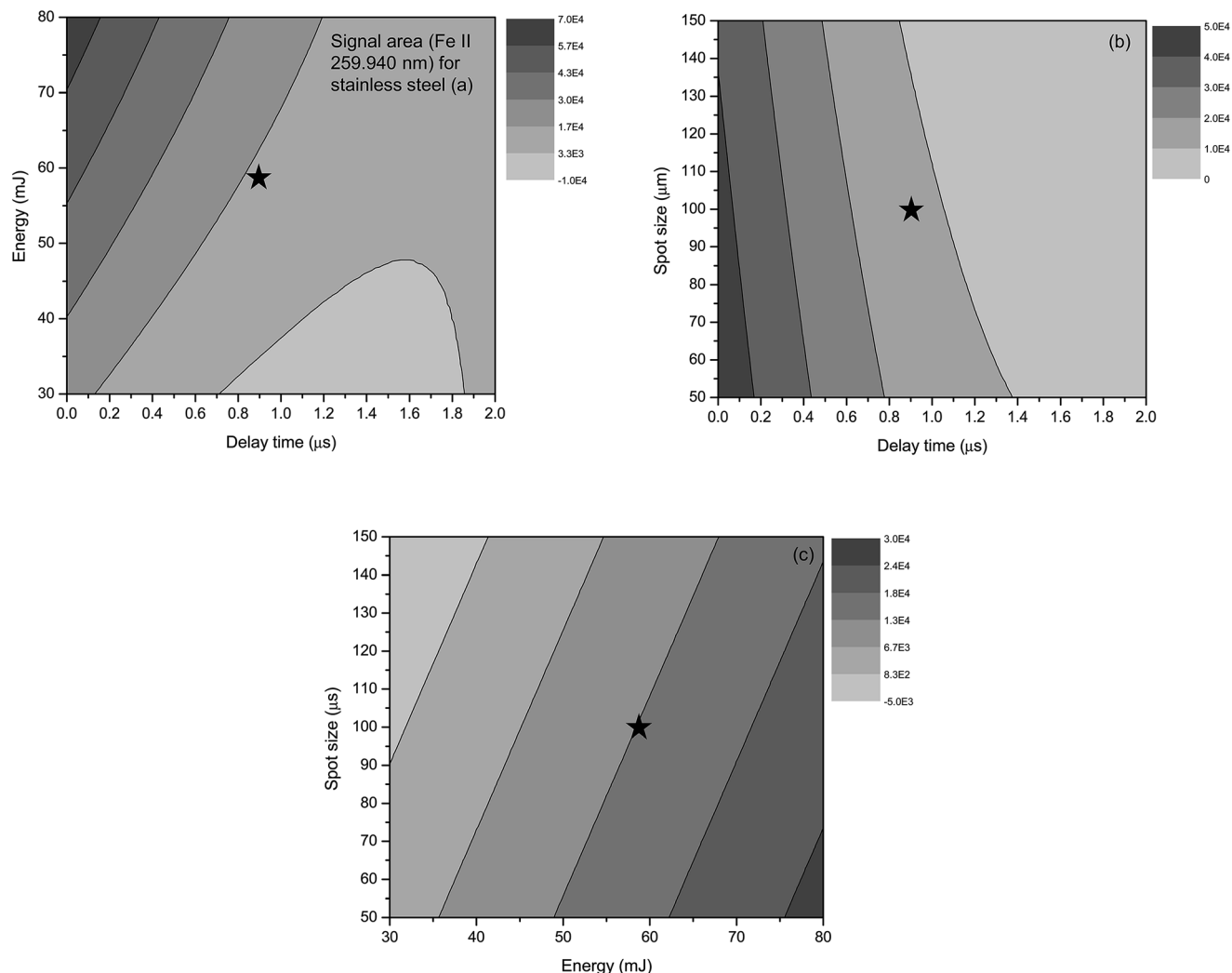


Fig. 1 Contour plots of signal area optimization, considering the Fe II 259.940 nm emission line and a stainless steel sample (group 2): delay time versus energy (a), delay time versus spot size (b) and energy versus spot size (c). The star represents the selected working conditions.

Table 4 SBR, signal area and height (range and average) obtained ($n = 100$) and predicted (95% confidence level) with the optimized conditions for the 9 alloys used

Sample name (Signal identification)	SBR	Obtained area (range and average)	Predicted area range (95%)	Obtained height (range and average)	Predicted height range (95%)
Aluminum alloy (Al I 394.400 nm)	82	6421–17 165, 10 627	11 990–14 894	2268–6002, 3776	3912–5191
Co–Cr–Mo alloy (Cr II 283.563 nm)*	16	21 024–39 182, 31 251	6545–10 238	11 625–21 635, 17 297	3795–5617
Unidentified alloy (Cu I 324.754 nm)	4	1043–6587, 3655	1346–1791	598–2780, 1471	5609–699
Steel (Fe II 259.940 nm)	22	36 693–65 180, 52 701	57 957–68 586	12 253–22 058, 17 384	18 716–21 194
Stainless steel (Fe II 259.940 nm)	27	14 239–58 344, 40 478	17 294–25 080	5569–20 436, 14 498	5488–7902
Stainless steel (Fe II 259.940 nm)	27	6905–21 675, 12 800	11 859–17 133	2596–8208, 4928	3697–5517
Stainless steel (Mn II 294.920 nm)	3	646–3014, 1756	1464–1939	390–2025, 1114	870–1129
Hastelloy (Mo I 386.411 nm)	4	351–1228, 760	298–690	158–600, 351	203–323
Hastelloy (Ni I 361.939 nm)	12	196–968, 593	373–1014	90–462, 260	189–391
6Al 4V titanium alloy (V I 437.924 nm)	3	29–853, 462	29–63	64–335, 189	41–73
Steel – two layers of Cr (Zn I 481.053 nm)	46	89 600–226 049, 112 942	167 357–193 480	34 678–59 687, 42 351	55 273–62 618

analytes investigated, the same tendency was seen and the only exception was for Cr (marked with an asterisk, second row). In fact, the model proposed for Cr did not present lack of fit (see Tables 2S and 3S at the ESI†), but due to the complexity of the samples, this type of test can present a discrepancy.

3.2 Dataset emission spectra

Fig. 2S (see ESI†) shows some representative spectra after the optimization for different samples, with several emission lines highlighted. In this case, the readers can have a look at the spectral environment in the different groups of samples. Fig. 2Sa† shows a group 1 sample spectrum and emission lines for C and Fe. Fig. 2Sb–d† show some representative samples from group 2, with emission lines for C, Fe, Mn and Zn identified. Fig. 2Se–i† (group 3) show emission lines for Al, Co, Cr, Cu, Mo, Ni, Ti and V. The RSDs for the monitored lines were

calculated; for signal height, RSD varied from 10% for Zn (analyte concentration of 83% w/w) to 38% for V (analyte concentration of 4.0% w/w).

3.3 Multivariate and univariate calibration models

Before calculating the models, the spectra were submitted to several normalization modes, as previously described in Table 1S (ESI†) and Section 2.5 Normalization mode. Multivariate calibration models were calculated for each normalization mode using the whole signal profile. The best models were evaluated according to SEC, SECV, SEV, LOD_{min}, LOD_{max} and LOD_{pu} (LOD pseudounivariate).³¹ In this case, the dataset was split into calibration (64 samples) and validation (16 samples) using the Kennard–Stone algorithm.²⁹ Both datasets were mean-centered.

Carbon emission lines (related to normalization modes 5 and 11) were not observed in any sample. In this case, Fig. 2 shows the SEV values obtained for 10 normalization modes (1, 2, 3, 4, 6, 7, 8, 9, 10 and 12) for Cr. The lowest SEV value was observed for normalization 10, marked with an asterisk, (see details in Section 2.5 Normalization mode and Table 1S†) and this mode was selected for further calculation of this analyte. This SEV value was compared with those of the other normalization modes using an *F* test, with the 95% confidence level, and differences of normalization 10 (selected one) were observed only when compared with normalizations 6 and 12 (marked with two asterisks). The same procedure was repeated for the other analytes (please see Fig. 3S, ESI†) and in the majority of the cases, statistical differences were not observed among the normalization modes.

On the other hand, we always selected the normalization with the lowest SEV values (marked with an asterisk), because when comparing these values with those of the other normalizations, the errors presented were two- to five-fold lower.

Using the PLS regression vectors, the most important emission lines for each analyte were selected, and univariate models were also evaluated. Fig. 3 shows the selected emission lines for each analyte. In the univariate calibration, the best normalization mode (see Fig. 3S†) was used. In the case of Al, the best univariate

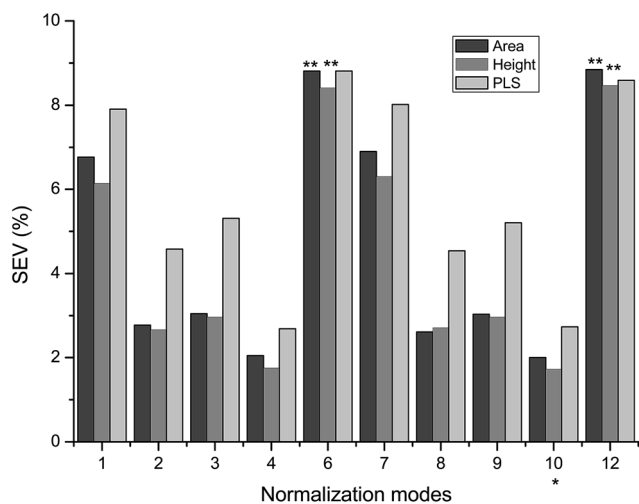


Fig. 2 Standard error of validation for Cr (univariate and multivariate models) using the normalization modes proposed. An asterisk represents the mode selected and type of signal. Normalization modes with two asterisks presented significant difference with the selected one with the 95% confidence level.

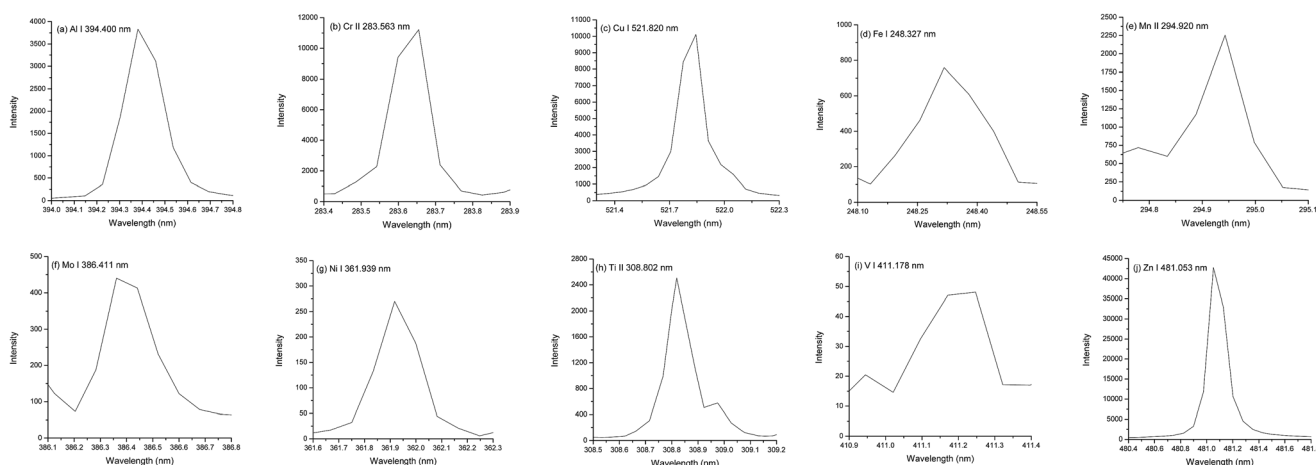


Fig. 3 Selected emission lines of the 10 analytes determined: (a) Al, (b) Cr, (c) Cu, (d) Fe, (e) Mn, (f) Mo, (g) Ni, (h) Ti, (i) V and (j) Zn.

Table 5 Comparison between univariate (area and height) and multivariate models for the 10 analytes using the selected normalization modes: SEV, LOD, LOD_{min}, LOD_{max} and LOD_{pu} (all in % w/w)

	Emission lines (nm)	Normalization mode	Area		Height		Multivariate				
			SEV	LOD	SEV	LOD	SEV	VL	LOD _{min}	LOD _{max}	LOD _{pu}
Al	394.400	2	5.82	0.28	5.87	0.81	13.0	3	0.61	2.0	16
Cr	283.563	10	2.00	0.65	1.72	1.4	2.73	3	1.5	3.5	9.9
Cu	521.820	4	5.12	0.25	5.25	0.65	14.2	4	0.28	0.65	16
Fe	248.327	9	10.2	6.7	11.1	16	8.75	3	3.1	8.2	43
Mn	294.920	4	0.37	0.12	0.27	0.14	0.46	1	0.15	0.19	1.5
Mo	386.411	12	1.46	1.5	4.33	10	2.59	3	0.13	0.39	2.8
Ni	361.939	12	13.3	1.2	14.2	2.7	13.5	3	1.8	9.4	15
Ti	308.802	8	7.91	1.6	8.44	4.0	7.53	3	1.4	4.7	29
V	411.178	10	1.05	14	1.30	16	0.90	1	0.16	0.34	1.2
Zn	481.053	12	2.50	4.2	2.13	9.2	2.51	4	5.5	8.7	12

model was obtained for emission line number 23 (see Fig. 2Se,† Al I 394.400 nm). At least 2 (in the case of Ni) to 8 (in the case of Cr and Fe) emission lines were tested in this part of the study.

Table 5 shows SEV, latent variables (LV), LOD_{min}, LOD_{max} and LOD_{pu} for the calculated PLS models and SEV and LOD for the univariate models (area and height). In this table it is also possible to see the normalization modes selected for each analyte, as well as, the emission lines for the univariate models (see details in Fig. 2S† and 3). An additional *F* test was made in order to compare the univariate and multivariate models and in all cases, no difference was observed with the 95% confidence level.

SEV values varied from 0.27% (Mn) to 14% (Cu and Ni), while the LOD_{min} values ranged from 0.13% (Mo) to 5.5% (Zn). Zinc was not divided into two sets because few samples presented this element and SEC values were used in all cases. Note that some analytes that presented low concentration (in the case of Mo and V) required some type of sum in the normalization process (normalizations 10 and 12).

The calculated LOD values for multivariate and univariate models for Cr are presented in Fig. 4. The LOD values for

univariate models were systematically lower than those obtained for multivariate models (LOD_{min}, LOD_{max} and LOD_{pu}). This observation is expected because in the multivariate calculation, the contribution of *y* values, var(*y*_{cal}) (variance for the analytes

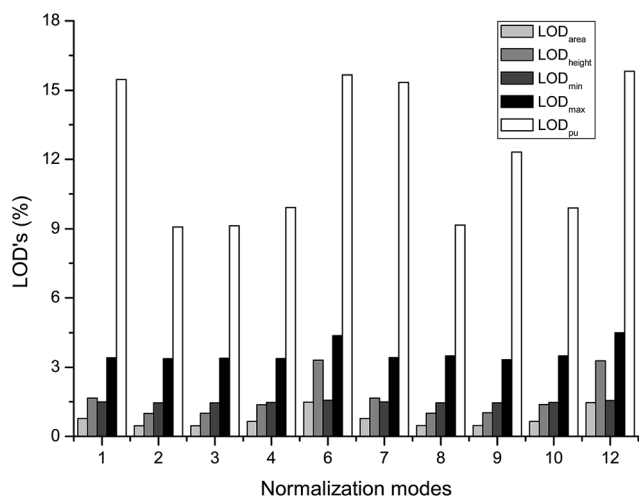
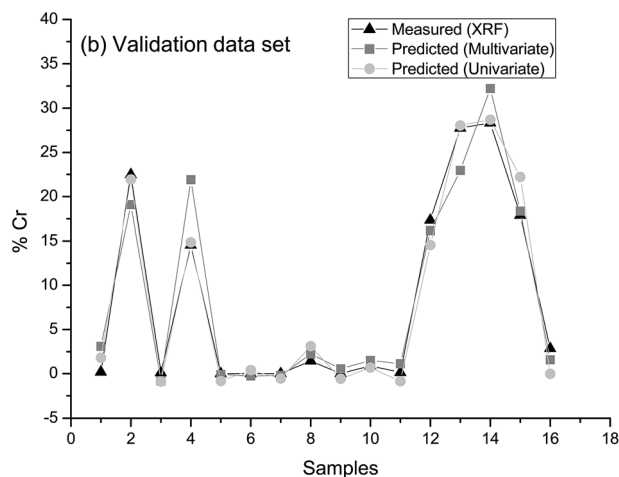
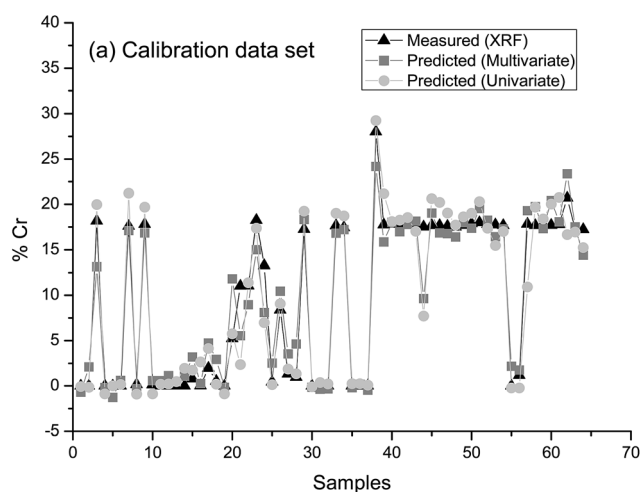


Fig. 4 Limit of detection (LOD) values for univariate models (LOD_{area} and LOD_{height}) and multivariate models (LOD_{min}, LOD_{max} and LOD_{pu}) for Cr determination in the normalization modes used.

Fig. 5 Reference (XRF), predicted using multivariate (PLS) and univariate (height) concentration for chromium (w/w%) models obtained in calibration (a) and validation (b) datasets.

concentration) is considered.³¹ A general comparison among the LOD values for the other analytes is presented in Fig. 4S.†

As an example of the quality of the calibration models proposed, Fig. 5 shows the comparison of the two models for Cr (multivariate with a total peak profile and univariate, using height, with emission line 36, Cr II 283.563 nm, in both cases normalized by normalization number 10). Fig. 5a and b show the results for calibration and validation datasets, respectively. In both cases, there was no significant difference between models, and the error values (SEC and SEV) ranged from 2.0–2.7 for multivariate models and from 1.7 (signal height) to 2.5% (signal height) for univariate models.

In general, for the other analytes, there was a good correlation between the predicted and reference values for the majority of analytes, with correlation (R^2) ranging from 0.42 (for V) to 0.99 (for Ti) for multivariate calibration and 0.23 (for Ni) to 0.99 (for Al) for univariate calibration.

Taking into consideration all model parameters, the univariate calibration using area or height generally presented better results than the multivariate one, mainly when SEV values are considered. In the case of the LOD, the values obtained with PLS, despite being high, were more realistic because the y errors were considered (variance).

3.4 Classification models

Fig. 6 shows the figures of merit for the classification models proposed. Three types of datasets were selected: XRF concentrations, signal area and height for all emission lines described in Fig. 2S† (57 emission lines) in the normalization modes selected for the multivariate and univariate models (see Table 5). For each data, three classification models were proposed: KNN, SIMCA and PLS-DA, and in all cases, the matrices were autoscaled.

In the case of KNN, one neighbor was selected for the XRF and LIBS dataset. For SIMCA, the number of principal components was 2 or 3 for the 3 classes considered. For PLS-DA, one latent variable was selected in all cases.

Fig. 6 shows the accuracy, sensitivity, false alarm rate and specificity of XRF (Fig. 6a); LIBS selected signal height (Fig. 6b) and area (Fig. 6c) for the validation dataset (16 samples). Among the three different models, KNN presented the best results, with an accuracy of approximately 100%, followed by SIMCA (accuracy ranging from 80–100%). Among the three different types of data, the XRF concentration was better, followed by the signal area (LIBS). Therefore, it is possible to conclude that all the models (except PLS-DA) were efficient for sample classification.

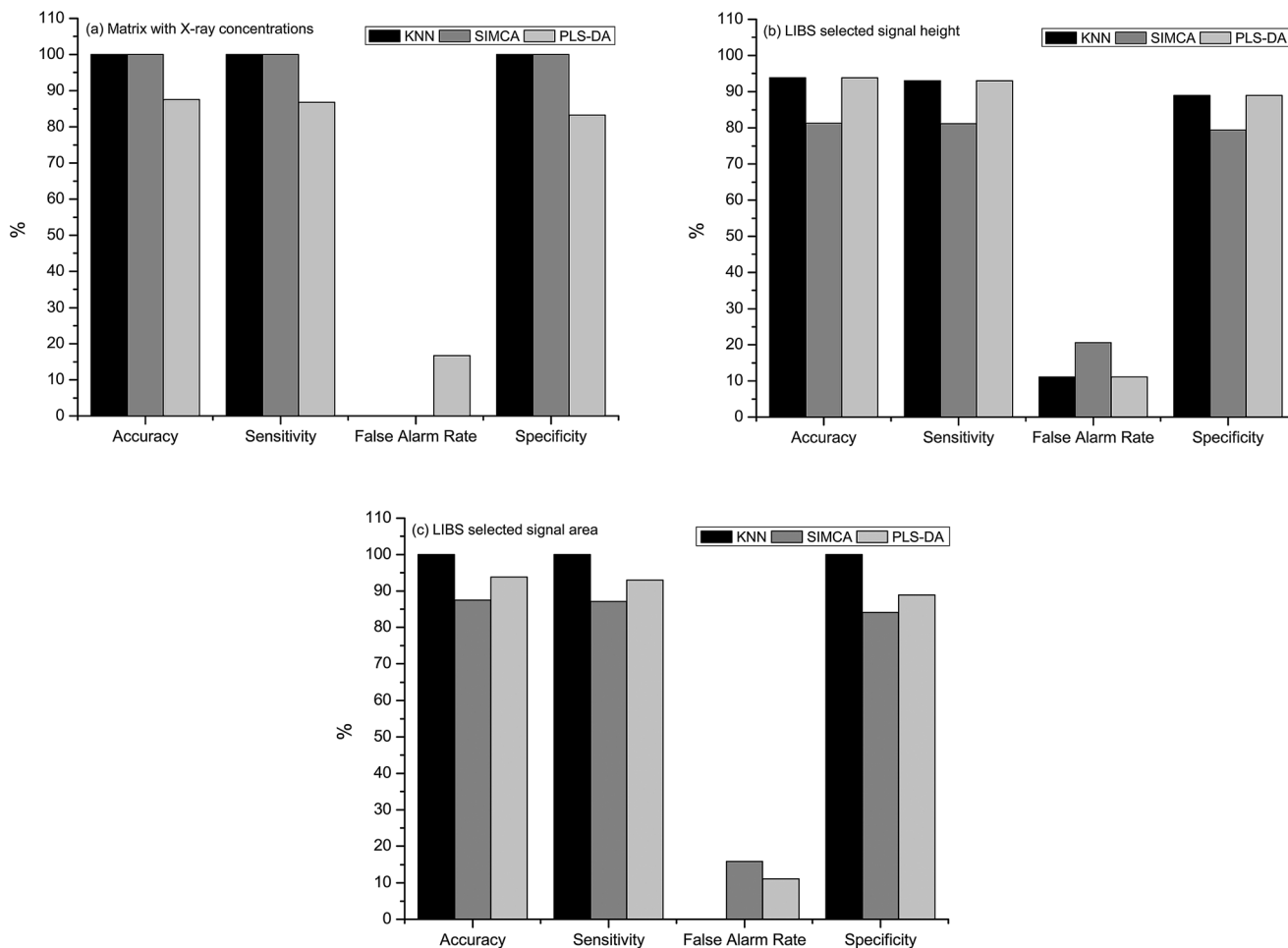


Fig. 6 Accuracy, sensitivity, false alarm rate and specificity obtained for KNN, SIMCA and PLS-DA models for XRF (a), LIBS signal height (b) and LIBS signal area (c).

4. Conclusion

The use of factorial design made it possible to obtain a commitment condition for all analytes simultaneously. This study shows that the application of normalization modes was useful to compensate for the differences among sample matrices. These results were compared with those of the F tests and for the majority of the cases, significant differences were not observed, but the lowest SEV values were two- to five-fold lower when compared with the others. Moreover, the combination of LIBS with chemometric strategies was a good alternative for proposing reliable regression models. The most important emission lines were selected using regression vectors from PLS.

Finally, univariate models were more efficient, with a lower LOD (but the variance of the y values is not considered), when compared with multivariate models. Classification models were useful for identifying the samples, with KNN being the best classification model compared with SIMCA and PLS-DA.

Acknowledgements

This study was supported by the São Paulo Research Foundation (FAPESP, grants 2012/01769-3, 2012/50827-6 and 2014/22408-4) and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq, 401074/2014-5 and 305637/2015-0). The authors are also grateful to Prof. Dr Ronei J. Poppi and Dr José Augusto Dacol for XRF determinations, Dr Marcos Monteiro for some samples supplied and Thermo Scientific for the Niton XL3T 985 ULTRA measurements.

References

- 1 C. Pasquini, J. Cortez, L. M. C. Silva and F. B. Gonzaga, *J. Braz. Chem. Soc.*, 2007, **18**, 463–512.
- 2 D. A. Cremers and L. J. Radziemski, *Handbook of Laser-Induced Breakdown Spectroscopy*, John Wiley & Sons, Chichester, 2013.
- 3 J. El Haddad, L. Canioni and B. Bousquet, *Spectrochim. Acta, Part B*, 2014, **101**, 171–182.
- 4 G. Galbács, *Anal. Bioanal. Chem.*, 2015, **407**, 7537–7562.
- 5 M. J. C. Pontes, J. Cortez, R. K. H. Galvão, C. Pasquini, M. C. U. Araújo, R. M. Coelho, M. K. Chiba, M. F. de Abreu and B. E. Madari, *Anal. Chim. Acta*, 2009, **642**, 12–18.
- 6 P. Fichet, P. Mauchien, J. F. Wagner and C. Moulin, *Anal. Chim. Acta*, 2001, **429**, 269–278.
- 7 R. J. Lasheras, C. Bello-Gálvez and J. Anzano, *Polym. Test.*, 2010, **29**, 1057–1064.
- 8 M. A. Aguirre, M. Hidalgo, A. Canals, J. A. Nóbrega and E. R. Pereira-Filho, *Talanta*, 2013, **117**, 419–424.
- 9 G. P. Gupta, B. M. Suri, A. Verma, M. Sundararaman, V. K. Unnikrishnan, K. Alti, V. B. Kartha and C. Santhosh, *J. Alloys Compd.*, 2011, **509**, 3740–3745.
- 10 M. C. Ortiz, L. Sarabia, A. Jurado-López and M. D. Luque de Castro, *Anal. Chim. Acta*, 2004, **515**, 151–157.
- 11 T. Zhang, L. Liang, K. Wang, H. Tang, X. Yang, Y. Duan and H. Li, *J. Anal. At. Spectrom.*, 2014, **29**, 2323–2329.
- 12 Z. Hao, L. Guo, C. Li, M. Shen, X. Zou, X. Li, Y. Lu and X. Zeng, *J. Anal. At. Spectrom.*, 2014, **29**, 2309–2314.
- 13 R. Noll, H. Bette, A. Brysch, M. Kraushaar, I. Monch, L. Peter and V. Sturm, *Spectrochim. Acta, Part B*, 2001, **56**, 637–649.
- 14 S. Palanco and J. J. Laserna, *J. Anal. At. Spectrom.*, 2000, **15**, 1321–1327.
- 15 C. Lopez-Moreno, K. Amponsah-Manager, B. W. Smith, I. B. Gornushkin, N. Omenetto, S. Palanco, J. J. Laserna and J. D. Winefordner, *J. Anal. At. Spectrom.*, 2005, **20**, 552–556.
- 16 G. L. Bosco, *Trends Anal. Chem.*, 2013, **45**, 121–134.
- 17 V. Stum, R. Fleige, M. Kanter, R. Leitner, K. Pilz, D. Fischer, G. Hubmer and R. Noll, *Anal. Chem.*, 2014, **86**, 9687–9692.
- 18 S. M. Zaytsev, A. M. Popov, E. V. Chernykh, R. D. Voronina, N. B. Zorov and T. A. Labutin, *J. Anal. At. Spectrom.*, 2014, **29**, 1417–1424.
- 19 A. Sarkar, X. Mao and R. E. Russo, *Spectrochim. Acta, Part B*, 2014, **92**, 42–50.
- 20 L. C. Nunes, G. A. da Silva, L. C. Trevizan, D. S. Júnior, R. J. Poppi and F. J. Krug, *Spectrochim. Acta, Part B*, 2009, **64**, 565–572.
- 21 C. B. Stipe, B. D. Hensley, J. L. Boersema and S. G. Buckley, *Appl. Spectrosc.*, 2010, **64**, 154–160.
- 22 S. L. C. Ferreira, W. N. L. Santos, C. M. Quintella, B. Barros Neto and J. M. Bosque-Sendra, *Talanta*, 2004, **63**, 1061–1067.
- 23 A. Sarkar, V. Karki, S. K. Aggarwal, G. S. Maurya, R. Kumar, A. K. Rai, X. Mao and R. E. Russo, *Spectrochim. Acta, Part B*, 2015, **108**, 8–14.
- 24 T. Zhang, S. Wu, J. Dong, J. Wei, K. Wang, H. Tang, X. Yang and H. Li, *J. Anal. At. Spectrom.*, 2015, **30**, 368–374.
- 25 S. Yao, J. Lu, J. Li, K. Chen, J. Li and M. Dong, *J. Anal. At. Spectrom.*, 2010, **25**, 1733–1738.
- 26 L. E. García-Ayuso, J. Amador-Hernández, J. M. Fernández-Romero and M. D. Luque de Castro, *Anal. Chim. Acta*, 2002, **457**, 247–256.
- 27 K. Kjeldahl and R. Bro, *J. Chemom.*, 2010, **24**, 558–564.
- 28 F. Westad and F. Marini, *Anal. Chim. Acta*, 2015, **893**, 14–24.
- 29 R. W. Kennard and L. A. Stone, *Technometrics*, 1969, **11**, 137–148.
- 30 J. W. Braga and R. J. Poppi, *J. Pharm. Sci.*, 2004, **93**, 2124–2134.
- 31 F. Allegrini and A. C. Olivieri, *Anal. Chem.*, 2014, **86**, 7858–7866.
- 32 S. D. Brown, R. Tauler and B. Walczak, *Comprehensive Chemometrics*, Elsevier, 2009.
- 33 F. C. De Lucia Jr, J. L. Gottfried, C. A. Munson and A. W. Miziolek, *Appl. Opt.*, 2008, **47**, G112–G121.
- 34 J. L. Gottfried, R. S. Harmon, F. C. De Lucia Jr and A. W. Miziolek, *Spectrochim. Acta, Part B*, 2009, **64**, 1009–1019.
- 35 R. H. Myers, D. C. Montgomery and C. M. Anderson-Cook, *Response Surface Methodology*, John Wiley & Sons, Chichester, 2009.
- 36 A. Sarkar, V. M. Telmore, D. Alamelu and S. K. Aggarwal, *J. Anal. At. Spectrom.*, 2009, **24**, 1545–1550.

Anexo 4 – Material Suplementar do Artigo publicado no Journal of Analytical Atomic Spectrometry, 2016, 31, 2005-2014.

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	<p>p = number of coefficients</p> <p>n = number of experiments</p> <p>m = number of independent experiments</p> <p>\hat{y}_i : predicted response (area or height)</p> <p>\bar{y} : average of experimental responses (area or height)</p>
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	
Pure Error, SSPE	$\sum_{i=1}^m \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2$	n-m	
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.

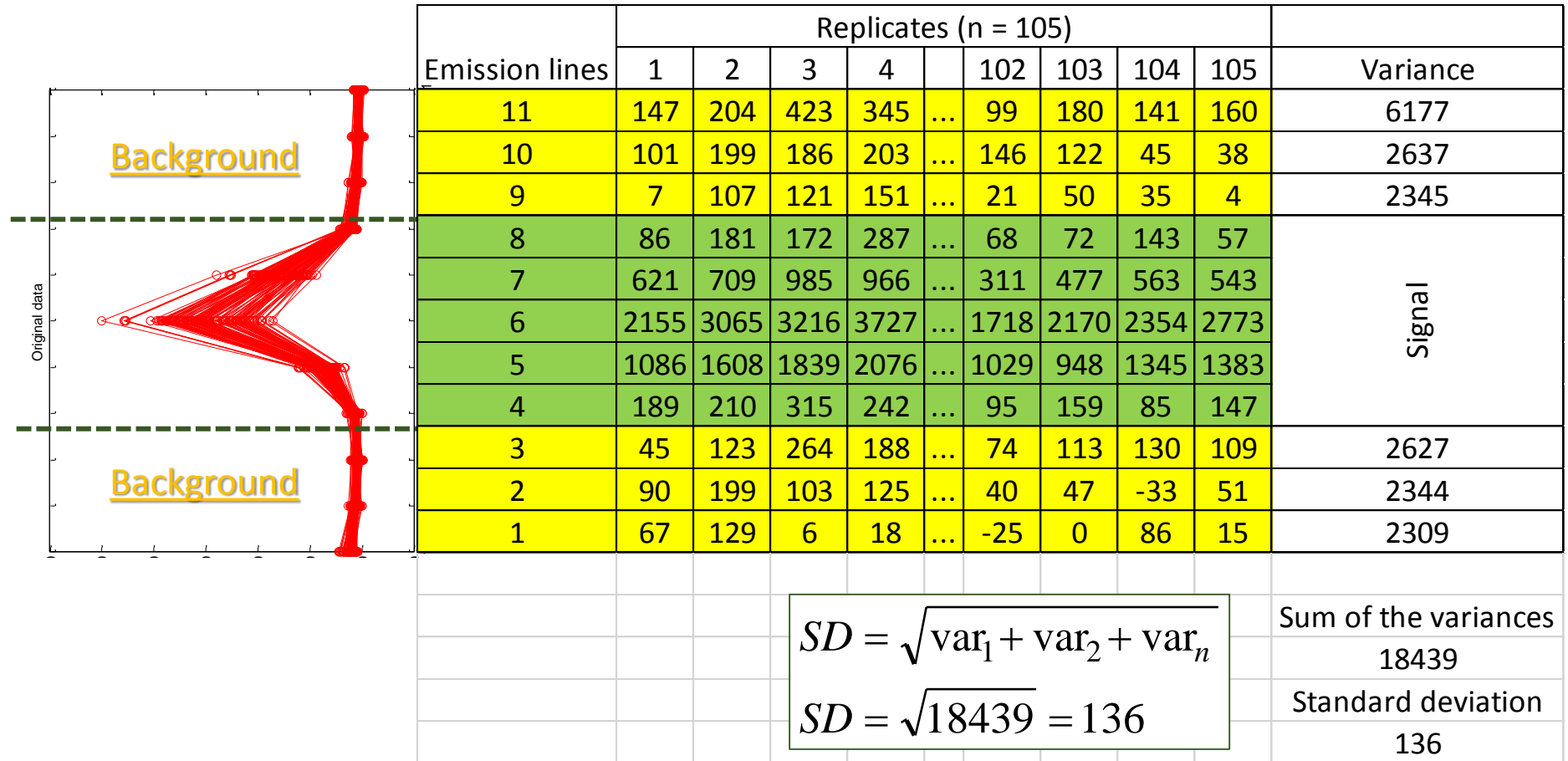


Fig. 2Sa Emission spectra for some samples and selected emission lines: steel with C and Fe lines (a), stainless steel with C and Fe lines (b), stainless steel with Mn lines (c), steel – two layers of Cr with Zn lines (d), aluminum alloy with Al lines (e), unidentified sample with Cu lines (f), Co-Cr-Mo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

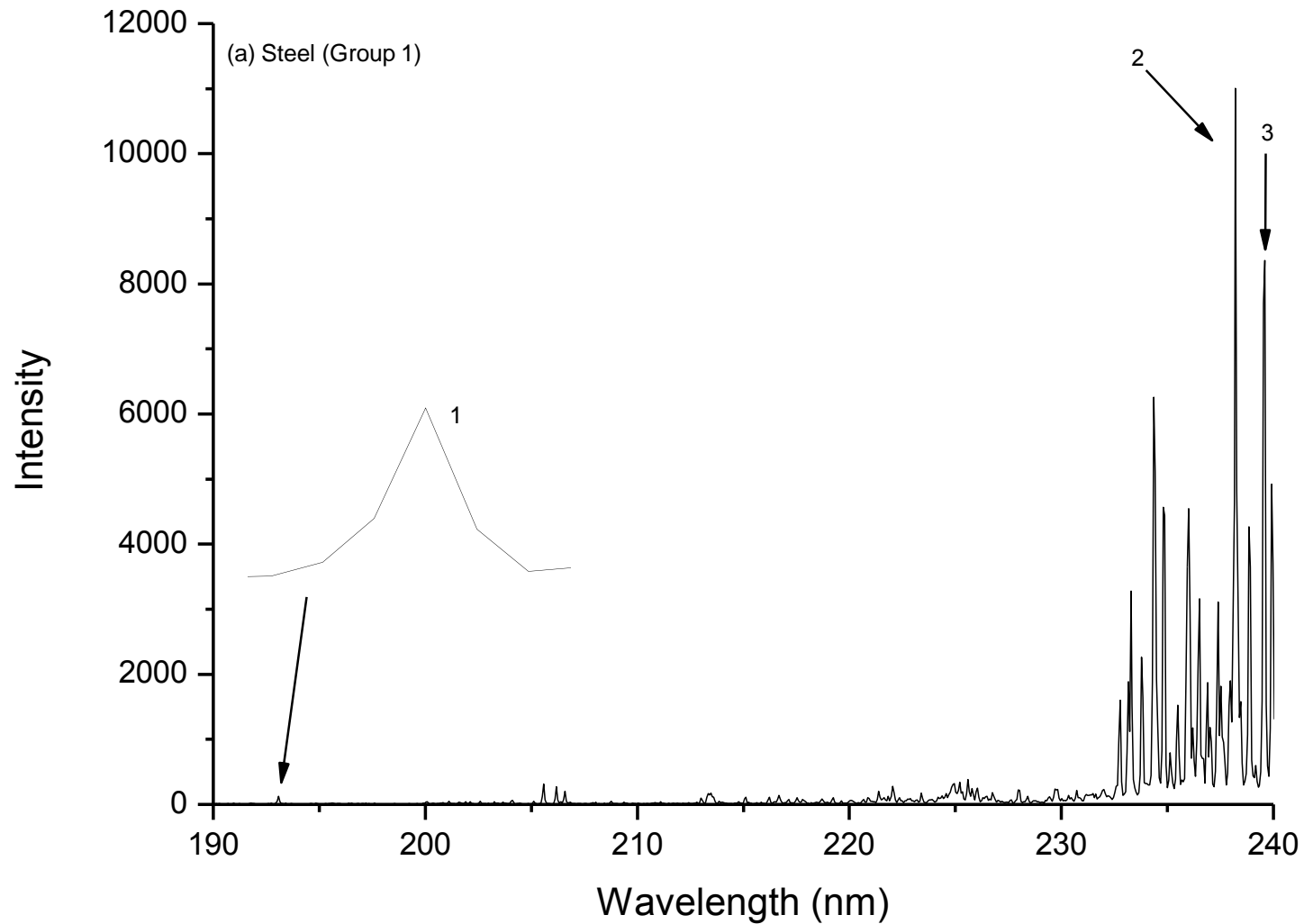


Fig. 2Sb

- 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

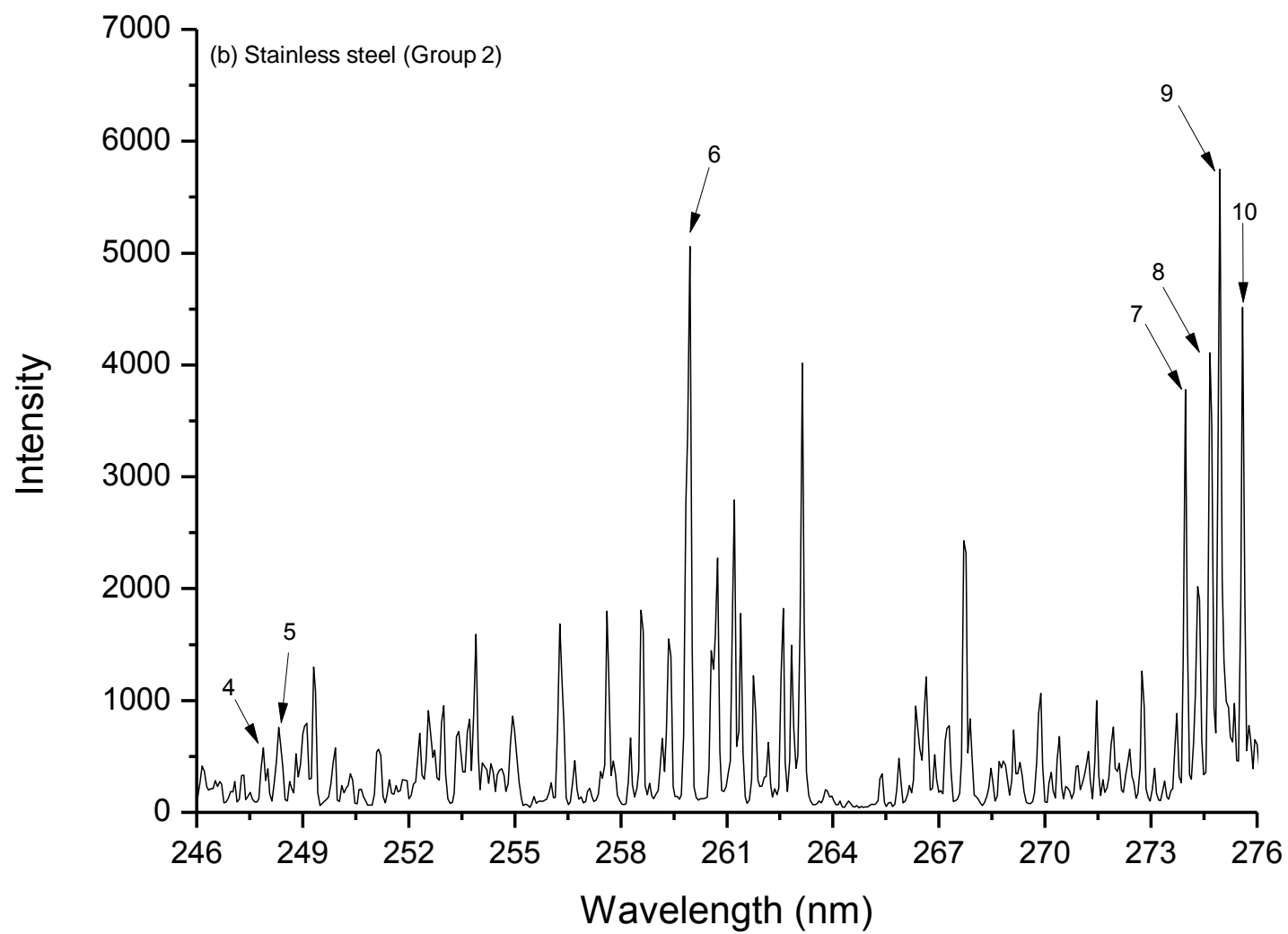


Fig. 2Sc

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

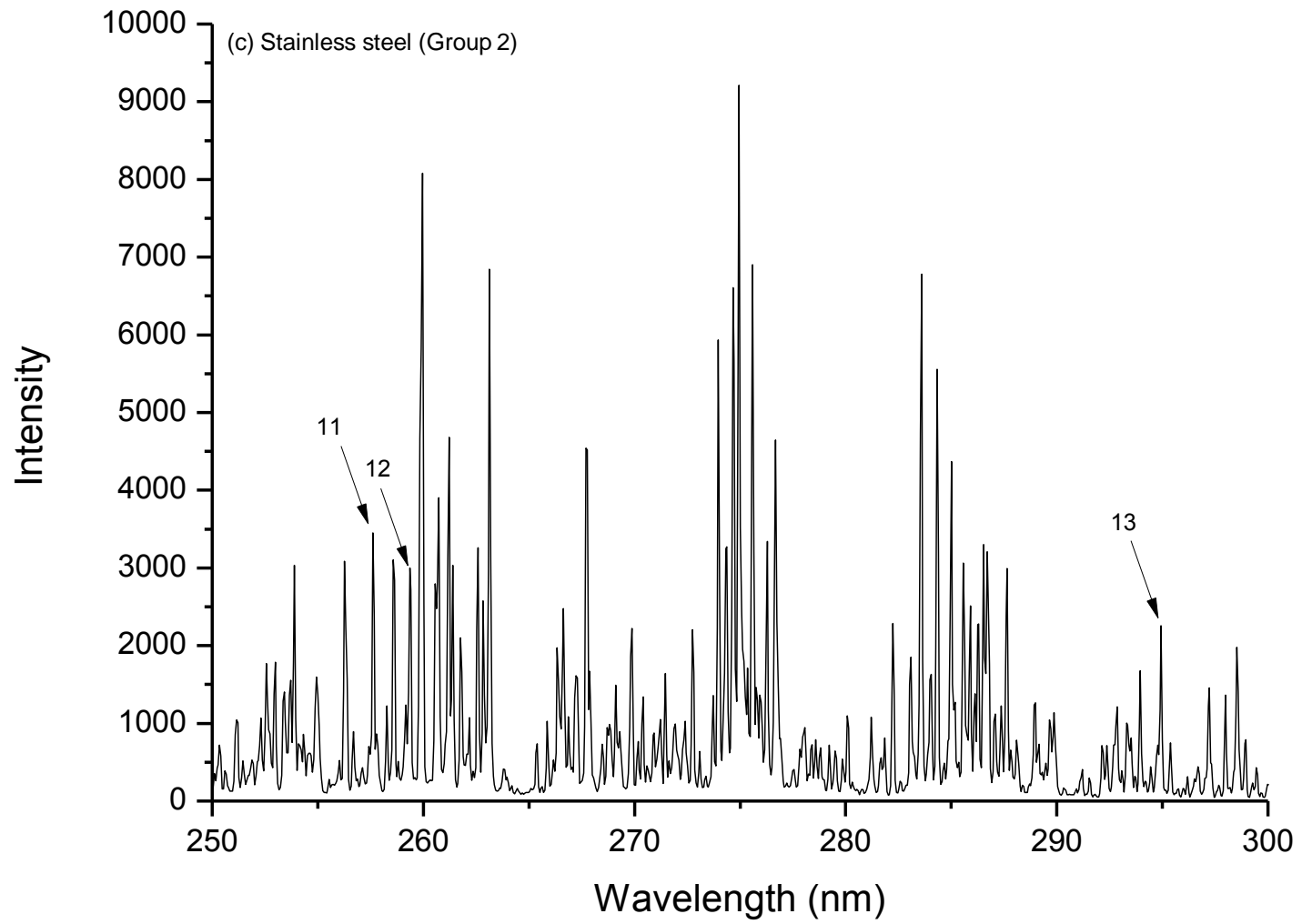


Fig. 2Sd

- 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

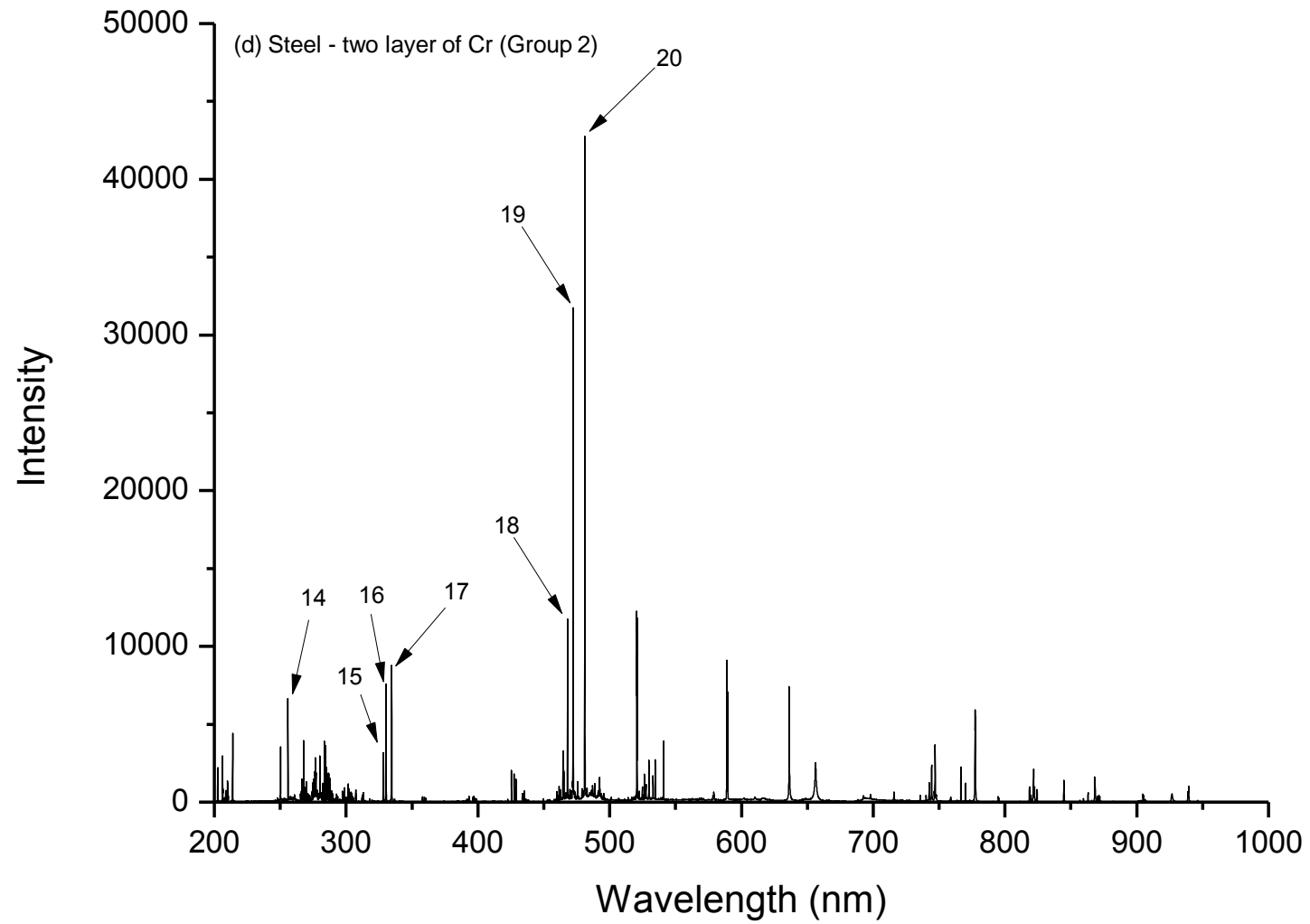


Fig. 2Se

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

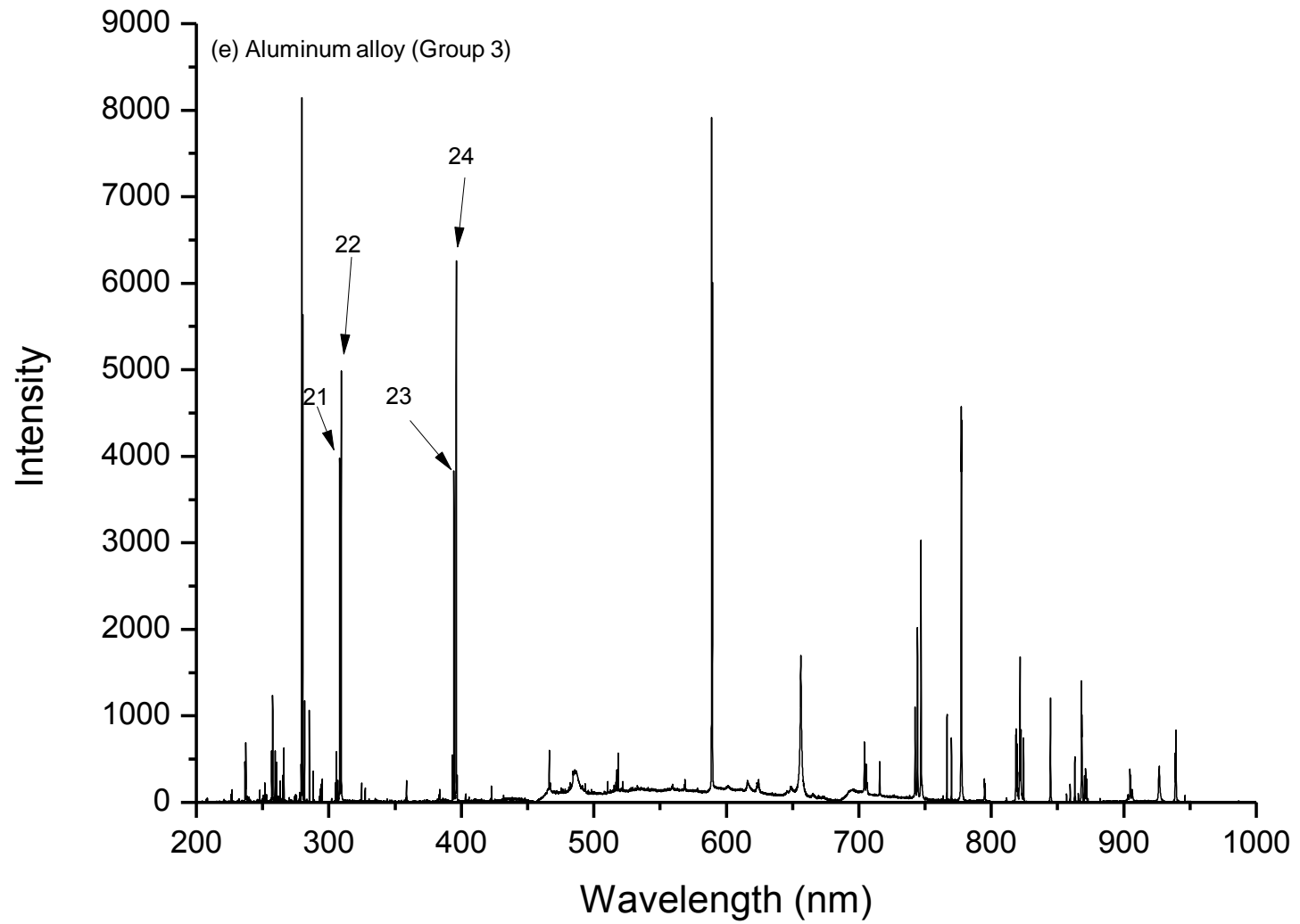


Fig. 2Sf

- 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

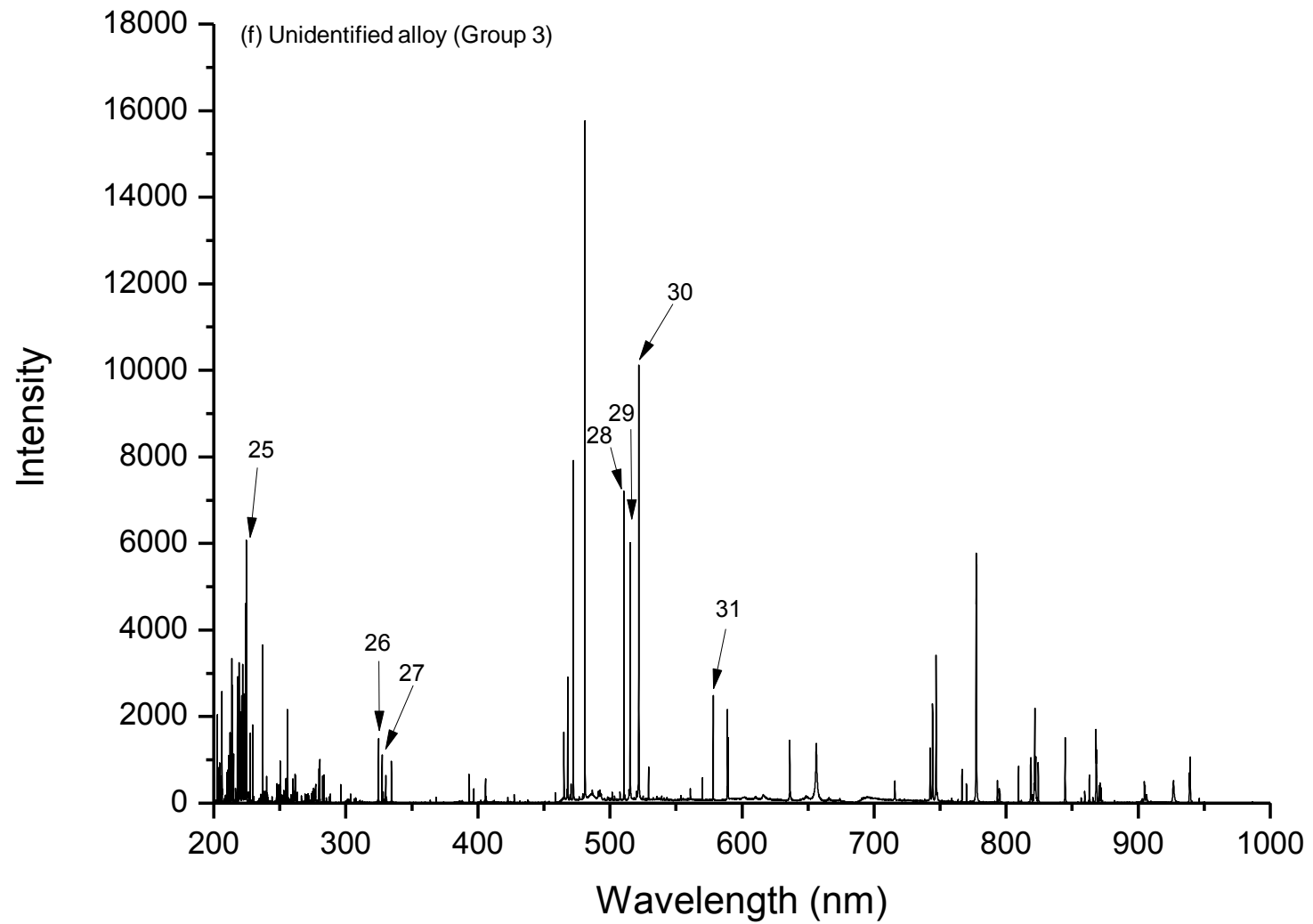


Fig. 2Sg

- 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

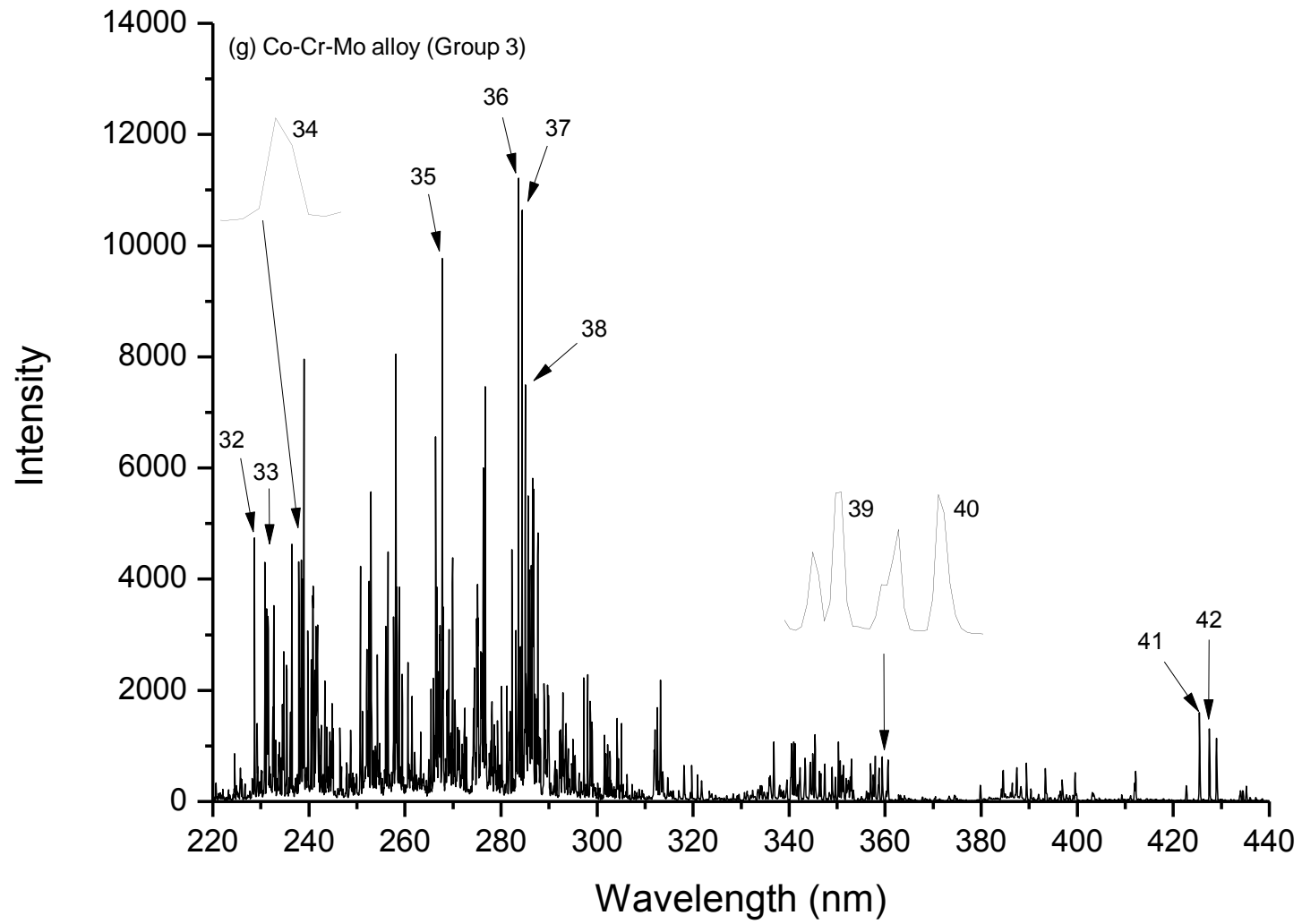


Fig. 2Sh

- 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

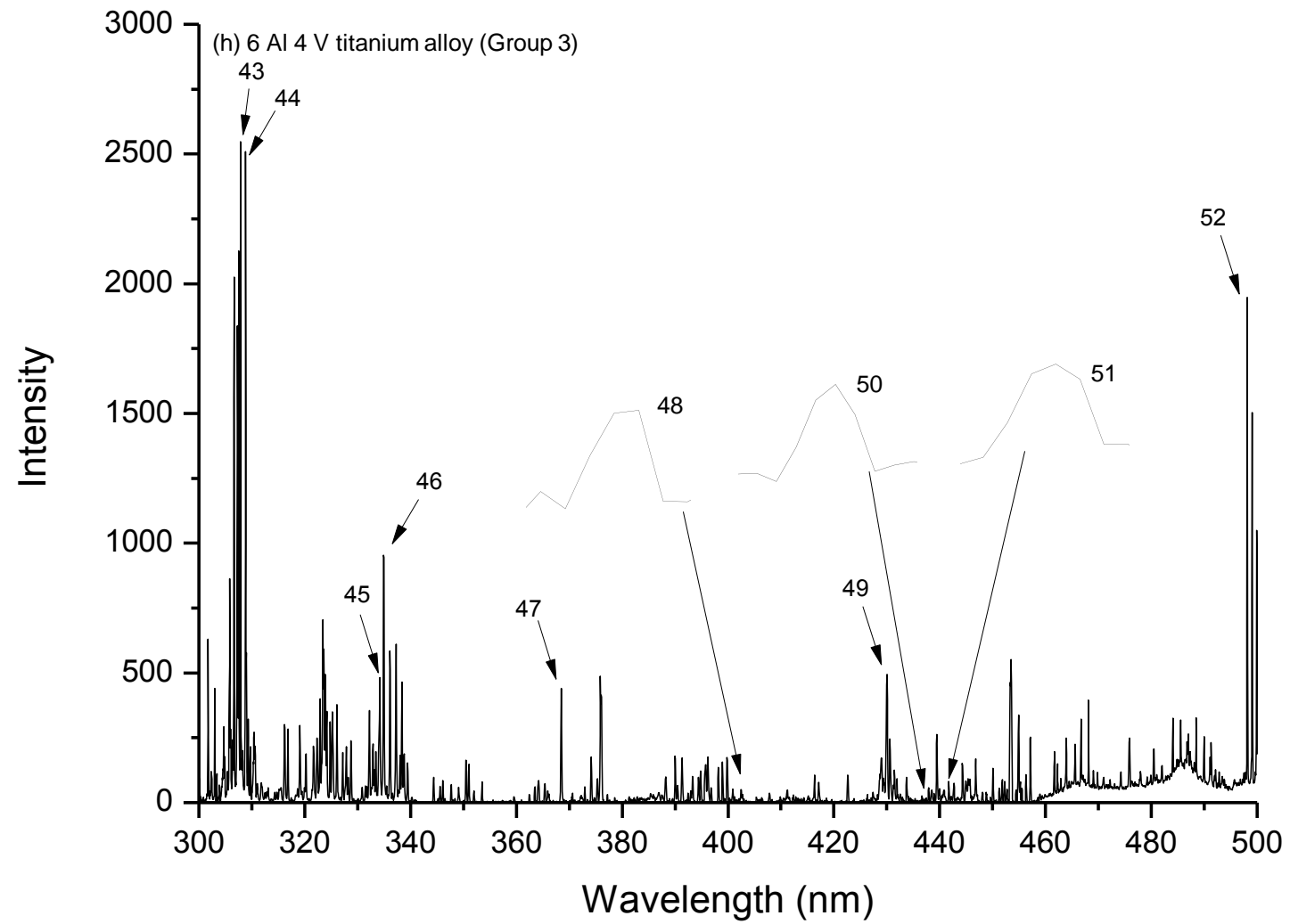


Fig. 2Si

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

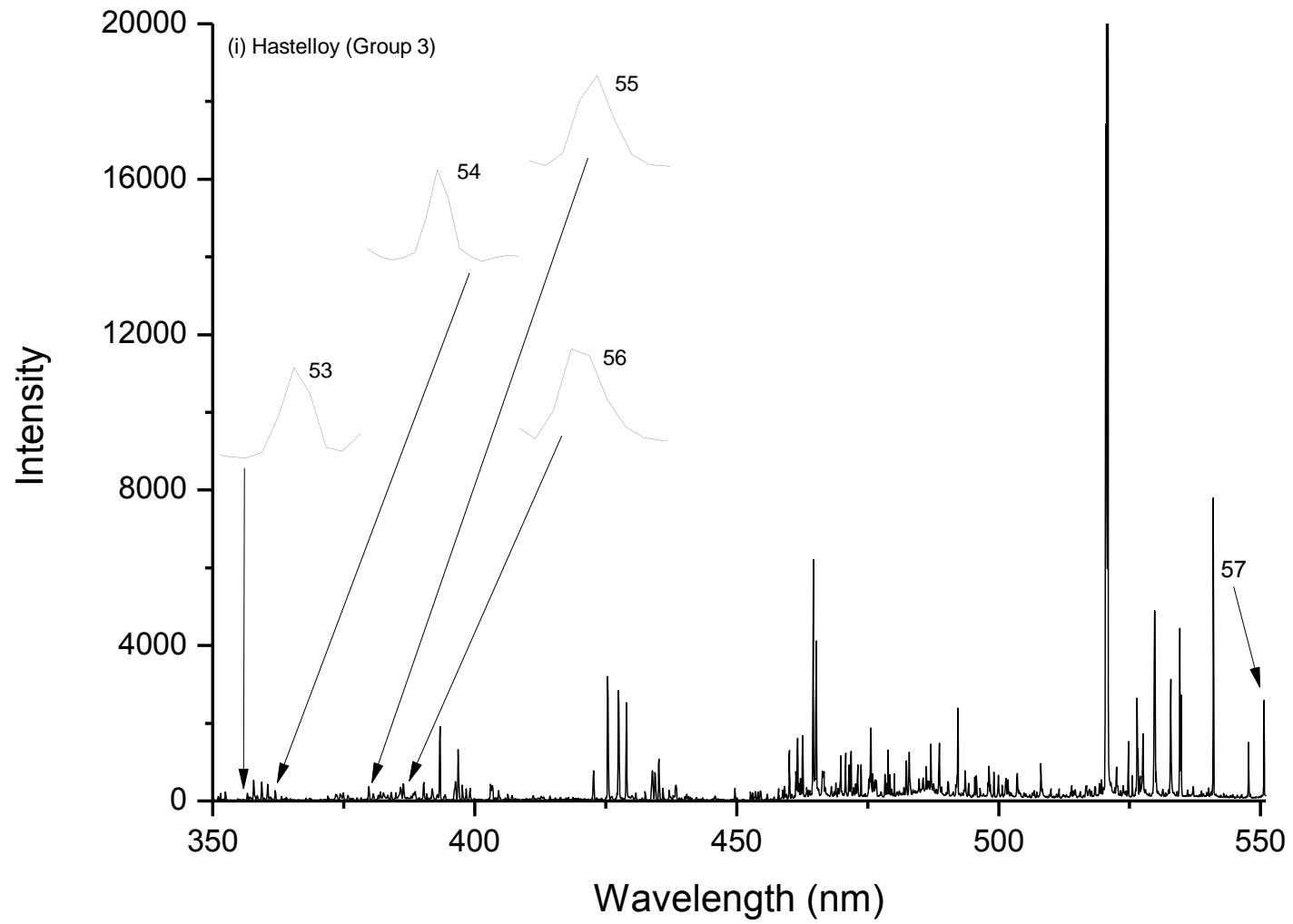
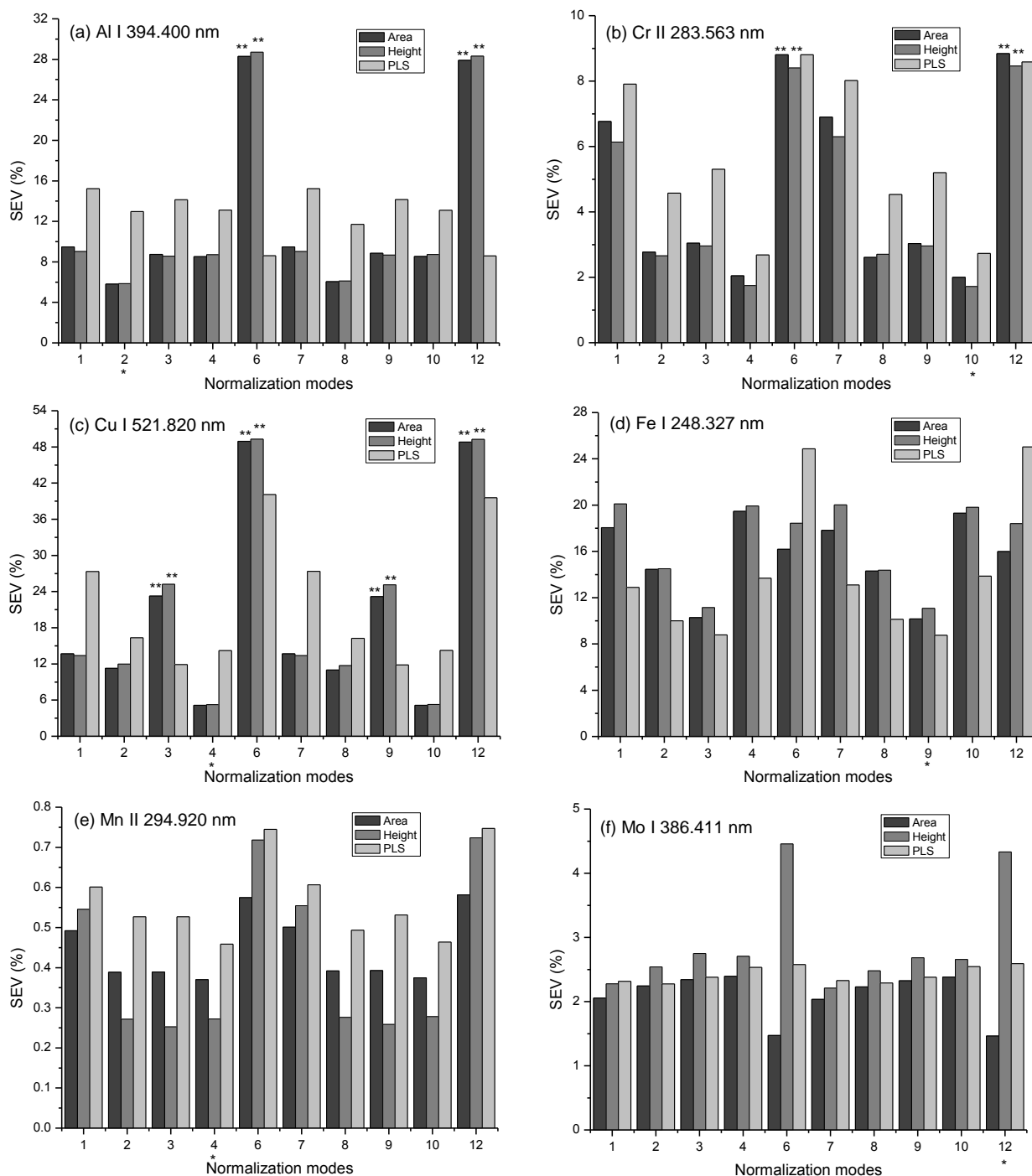


Fig. 3S

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.



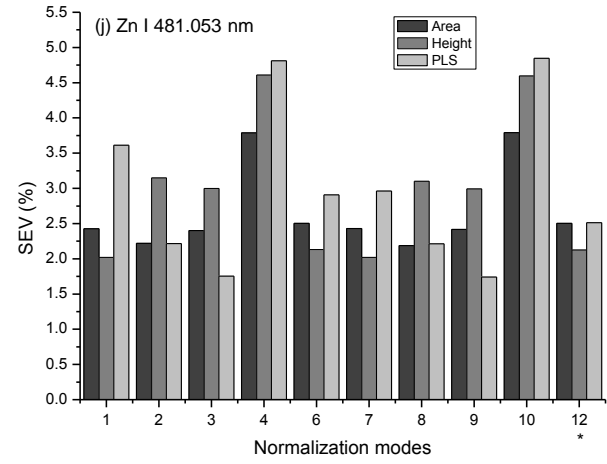
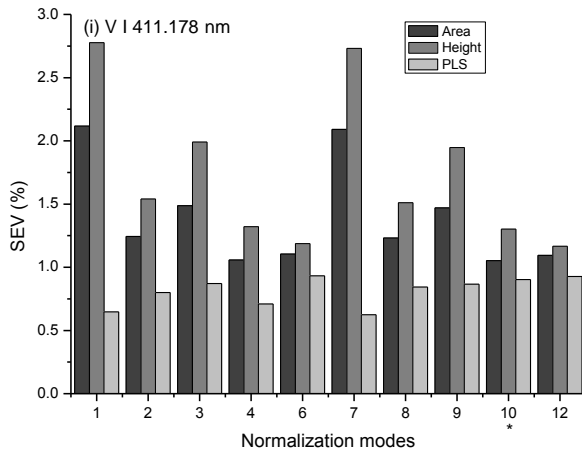
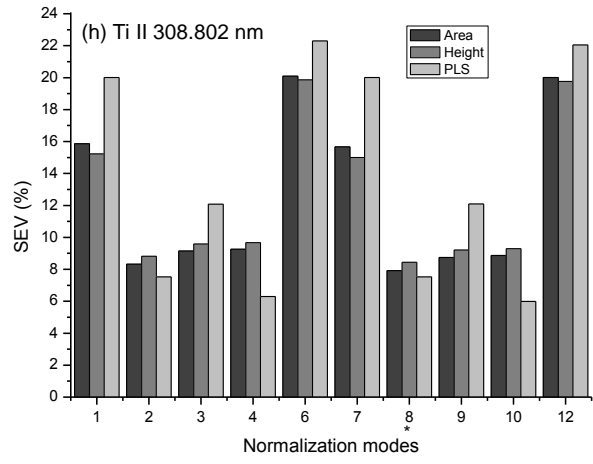
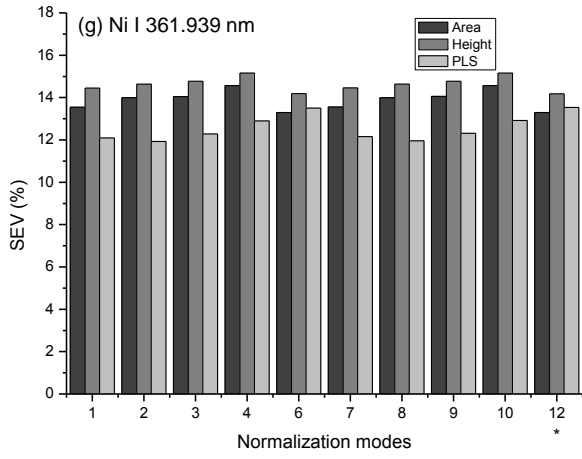


Fig. 4S

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

