



HAL
open science

Evaluation et optimisation de la robustesse d'une méthode d'analyse multivariée pour l'identification d'alliages par LIBS

E. Vors, K. Tchepidjian, J.-B. Sirven

► To cite this version:

E. Vors, K. Tchepidjian, J.-B. Sirven. Evaluation et optimisation de la robustesse d'une méthode d'analyse multivariée pour l'identification d'alliages par LIBS. Journées LIBS (Laser induced breakdown spectroscopy) 2015, Jun 2015, Verneuil En Halatte, France. cea-02489579

HAL Id: cea-02489579

<https://cea.hal.science/cea-02489579v1>

Submitted on 24 Feb 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



EVALUATION ET OPTIMISATION DE LA ROBUSTESSE D'UNE MÉTHODE D'ANALYSE MULTIVARIÉE POUR L'IDENTIFICATION D'ALLIAGES PAR LIBS

E. Vors, K. Tchepidjian, J.-B. Sirven

CEA, DEN, DPC, SEARS, LANIE, F-91191 Gif-sur-Yvette, France

Objectifs

Cette étude, qui traite de l'utilisation de la LIBS et la chimimétrie pour la classification et l'identification d'alliages, concerne plus particulièrement l'étude de la robustesse du modèle, motivée par la question suivante : pendant quelle durée un modèle prédictif (avec apprentissage supervisé), construit avec une base de données établie à une certaine date, peut-il être utilisé pour faire des prédictions? Pendant la période entre l'apprentissage et les tests de prédictions, les conditions instrumentales ont évolué de manière « naturelle ». Nous présentons ici l'optimisation d'une méthode de classification prédictive : la méthode SIMCA (« Soft Independent Modeling of Class Analogy »). Le modèle a été construit avec un jeu de données d'apprentissage issu de 13 alliages métalliques. Ces mesures ont ensuite été reproduites, sur les mêmes échantillons et avec le même instrument LIBS, huit mois après l'apprentissage : l'optimisation des paramètres du modèle SIMCA a été effectuée et un taux d'erreur de prédiction de 0% a ainsi pu être atteint après cette période. L'étude a ensuite été complétée par un test de la robustesse à plus long terme du modèle optimisé : deux ans après l'apprentissage, des mesures ont été effectuées sur des échantillons de test avec le même instrument LIBS et les prédictions correspondantes ont été effectuées avec ce modèle.

Expériences

DISPOSITIF LIBS

Instrument LIBS transportable (IVEA)

- Laser (Quantel) : $\lambda = 266$ nm, impulsions de 5 ns à 20 Hz, énergie par impulsion ~6-4 mJ. Diamètre du cratère sur cible ~50 μ m.
- Spectromètre à échelle (LLA) de bande spectrale 200-780 nm et de résolution spectrale $\lambda/\Delta\lambda \sim 10000$

Ex : spectre d'un acier inox

~50000 points par spectre

DÉMARCHE SUIVIE

Familles des échantillons	Aciers non inox				Aciers inox			Inconels			Monel	Autres	
Numéros des échantillons	1	2	5	12	9	10	11	6	7	8	13	4	3

Base de données de 2011 :
13 échantillons d'aciers et autres alliages, 25 spectres par échantillon

80% des spectres → apprentissage
20% des spectres → test du modèle

Données de 2012 :
les 13 mêmes échantillons, 50 spectres par échantillon

100% des spectres → validation et optimisation de la robustesse du modèle

↓

Données de 2014 :
3 échantillons de test, 25 spectres par échantillon

100% des spectres → **test final de la robustesse du modèle**

Résultats et discussion

MODÈLE SIMCA ET ÉVALUATION DE LA ROBUSTESSE

SIMCA

Apprentissage :

Test sur un spectre inconnu :
Suivant le critère de distance retenu, le spectre inconnu est attribué à la classe la plus proche (au-dessus d'un seuil).

Données de test de 2011 :
taux d'erreur de 3,1%

Matrice de confusion des données de test de 2011

Classe prédite	Classe réelle												
	1	2	3	4	5	6	7	8	9	10	11	12	13
1	4	0	0	0	0	0	0	0	0	0	0	0	0
2	1	5	0	0	0	0	0	0	0	0	0	0	0
3	0	0	5	0	0	0	0	0	0	0	0	0	0
4	0	0	0	5	0	0	0	0	0	0	0	0	0
5	0	0	0	0	4	0	0	0	0	0	0	0	0
6	0	0	0	0	0	5	0	0	0	0	0	0	0
7	0	0	0	0	0	0	5	0	0	0	0	0	0
8	0	0	0	0	0	0	0	5	0	0	0	0	0
9	0	0	0	0	0	0	0	0	5	0	0	0	0
10	0	0	0	0	0	0	0	0	0	5	0	0	0
11	0	0	0	0	0	0	0	0	0	0	5	0	0
12	0	0	0	0	0	0	0	0	0	0	0	5	0
13	0	0	0	0	0	0	0	0	0	0	0	0	5

Données de 2012 :
taux d'erreur de 16% → faible robustesse

Matrice de confusion des données de test de 2012

Classe prédite	Classe réelle												
	1	2	3	4	5	6	7	8	9	10	11	12	13
1	47	32	0	0	0	0	0	0	0	0	0	0	0
2	3	18	0	0	0	0	0	0	0	0	0	0	0
3	0	0	45	0	0	0	0	0	0	0	0	0	0
4	0	0	0	50	0	0	0	0	0	0	0	0	0
5	0	0	0	0	34	0	0	0	0	0	0	0	9
6	0	0	0	0	0	50	0	0	0	0	0	0	0
7	0	0	0	0	0	0	50	0	0	0	0	0	0
8	0	0	0	0	0	0	0	47	0	0	0	0	0
9	0	0	0	0	0	0	0	0	29	0	1	0	0
10	0	0	0	0	0	0	0	0	3	20	50	14	0
11	0	0	0	0	0	0	0	0	0	1	0	35	0
12	0	0	0	0	0	0	0	0	0	0	0	0	41
13	0	0	0	0	0	0	0	0	0	0	0	0	50

OPTIMISATION DE LA ROBUSTESSE DU MODÈLE SIMCA

Optimisation des paramètres du modèle avec un plan d'expérience :

Facteurs	Niveaux
Sélection de variables	4
Prétraitement	6
Définition de la distance	3
Seuil de classification	6

↓

Paramètres optimaux :

- Sélection de variables par sélection de raies sur les composantes ACP, avec un seuil d'intensité
- Prétraitement : normalisation par la somme des raies + centrage
- Définition de la distance basée sur les résidus
- Tout seuil de classification

Données de test de 2011 :
taux d'erreur de 0%

Matrice de confusion des données de test de 2011

Classe prédite	Classe réelle												
	1	2	3	4	5	6	7	8	9	10	11	12	13
1	5	0	0	0	0	0	0	0	0	0	0	0	0
2	0	5	0	0	0	0	0	0	0	0	0	0	0
3	0	0	5	0	0	0	0	0	0	0	0	0	0
4	0	0	0	5	0	0	0	0	0	0	0	0	0
5	0	0	0	0	5	0	0	0	0	0	0	0	0
6	0	0	0	0	0	5	0	0	0	0	0	0	0
7	0	0	0	0	0	0	5	0	0	0	0	0	0
8	0	0	0	0	0	0	0	5	0	0	0	0	0
9	0	0	0	0	0	0	0	0	5	0	0	0	0
10	0	0	0	0	0	0	0	0	0	5	0	0	0
11	0	0	0	0	0	0	0	0	0	0	5	0	0
12	0	0	0	0	0	0	0	0	0	0	0	5	0
13	0	0	0	0	0	0	0	0	0	0	0	0	5

Données de 2012 :
taux d'erreur de 0%

Matrice de confusion des données de test de 2012

Classe prédite	Classe réelle												
	1	2	3	4	5	6	7	8	9	10	11	12	13
1	50	0	0	0	0	0	0	0	0	0	0	0	0
2	0	50	0	0	0	0	0	0	0	0	0	0	0
3	0	0	50	0	0	0	0	0	0	0	0	0	0
4	0	0	0	50	0	0	0	0	0	0	0	0	0
5	0	0	0	0	50	0	0	0	0	0	0	0	0
6	0	0	0	0	0	50	0	0	0	0	0	0	0
7	0	0	0	0	0	0	50	0	0	0	0	0	0
8	0	0	0	0	0	0	0	50	0	0	0	0	0
9	0	0	0	0	0	0	0	0	50	0	0	0	0
10	0	0	0	0	0	0	0	0	0	50	0	0	0
11	0	0	0	0	0	0	0	0	0	0	50	0	0
12	0	0	0	0	0	0	0	0	0	0	0	50	0
13	0	0	0	0	0	0	0	0	0	0	0	0	50

TEST DE LA ROBUSTESSE À LONG TERME APRÈS OPTIMISATION

Test de la robustesse à long terme du modèle deux ans après l'apprentissage :

Prédictions sur 3 échantillons de test issus du lot précédent (mesures effectuées avec le même instrument LIBS). Pour chaque échantillon, le pourcentage de spectres attribués à chaque classe de la base de donnée est calculé. En vert : valeurs correspondants aux identifications correctes.

Prédictions de 2014

Classe prédite	Echantillon de test		
	1	2	3
1	0	76%	16%
2	0	0	44%
3	0	0	0
4	0	0	0
5	92%	0	16%
6	0	0	0
7	0	20%	20%
8	0	0	0
9	0	0	0
10	0	0	0
11	0	0	0
12	8%	0	4%
13	0	4%	0

→ Identification encore correcte

Conclusion

L'optimisation des paramètres du modèle prédictif au moyen d'un plan d'expérience a permis d'atteindre un taux d'erreur de prédiction nul pour le lot de données de validation acquies huit mois après l'apprentissage. Les performances des prédictions du modèle restent encore bonnes après deux ans, même si elles sont légèrement dégradées.

→ La démarche d'optimisation utilisée permet donc de rendre un modèle multivarié très robuste sur une longue durée.

Un prolongement intéressant de ce travail serait de faire une étude similaire avec des spectromètres compacts, dont les pouvoirs de résolution sont inférieurs (d'un facteur au moins 3). Dans ce cas, il serait intéressant de caractériser la robustesse et de déterminer la période durant laquelle les performances de prédiction restent dans des limites acceptables.