

Détermination automatique de fin de réactions à l'aide de méthodes chimométriques

Grégory **Corminboeuf** *, Michael **Amrhein** **, Olivier Naef *, Pascale **Voirin** *,
Ennio **Vanoli** *, Jean-Nicolas **Aebischer** *, Simon **Crelier*****

* *Ecole d'ingénieurs et d'architectes de Fribourg*, Bd de Pérolles 80,
CH-1705 Fribourg (Suisse). Correspondance : olivier.naef@hefr.ch.

** *EPFL, STI I2S LA, ME C2 389 (Bâtiment ME)*, Station 9, CH-1015 Lausanne

*** *Haute école valaisanne, Rte du Rawyl 47, Case Postale 2134, CH - 1950 Sion 2.*

MOTS CLÉS : Evolving Factor Analysis, Partial Least Square Discriminant Analysis, end-point detection, on-line monitoring.

1. Introduction

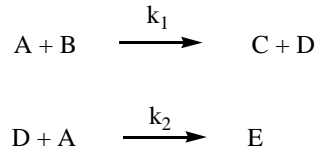
Dans l'optimisation des procédés, l'arrêt d'une réaction au moment opportun peut être d'une importance primordiale. En effet, lors de la post-réaction, des réactions secondaires peuvent survenir, avec comme conséquences la diminution du rendement ou l'introduction d'impuretés dans le produit final. Il est actuellement possible de déterminer la fin d'une réaction, de manière off-line, à l'aide de mesures spectroscopiques complétées par des méthodes chimométriques telles que l'Evolving Factor Analysis (EFA). Cette technique présente l'avantage de ne pas nécessiter de prises d'échantillons ni d'analyses classiques souvent longues et coûteuses (chromatographie liquide haute pression, chromatographie gazeuse, résonance magnétique nucléaire) [1]. Dans le cadre de notre projet, une méthode pouvant s'appliquer on-line a été étudiée. L'analyse discriminante par "partial least squares" (PLS-DA) est utilisée pour classifier les spectres infrarouges selon leur appartenance ou non au domaine de fin de réaction. Au cours de la réaction, le modèle prédit si le spectre appartient à la classe 1 (fin de réaction) ou à la classe 0 (réaction en cours).

La PLS-DA est une méthode de classification supervisée qui utilise la régression PLS pour classifier des échantillons selon leurs propriétés mesurées. En d'autres termes, la PLS-DA est une méthode de décomposition des données qui maximise la covariance entre le bloc X et le bloc Y [2]. Le modèle obtenu est de la forme : $Y = X B$. A chaque spectre de la matrice X correspond une valeur (0 ou 1) dans le vecteur Y . La valeur 1 indiquant l'appartenance au domaine de fin de réaction.

Dans un premier temps, un set de calibration est formé à partir de réactions simulées dans différentes conditions. Le changement de conditions a pour but de reproduire les déviations (dosage plus rapide, changement de température) pouvant apparaître lors d'essais réels. L'influence du prétraitement des spectres et le choix du nombre de variables latentes sont étudiés pour construire le modèle de classification à partir du set de calibration. Chaque modèle ainsi obtenu est testé sur un set de validation situé à l'intérieur du domaine de calibration.

2. Partie pratique

La réaction étudiée est une réaction de la forme:



Equation 1 : Réaction principale et réaction secondaire

Le réactif B est placé initialement dans le réacteur et le réactif A est ajouté selon un débit massique constant pour former le produit D désiré. Ce dernier est consommé par la deuxième réaction pour former le produit secondaire E. De part son coût élevé, la substance B est le réactif limitant. Le but est de détecter la fin de la première réaction, afin d'avoir un minimum de produit E.

Les différentes concentrations ainsi que les spectres infrarouges sont obtenus à l'aide d'une simulation. Un set de calibration comportant 10 expériences simulées est établi sur la base d'un plan d'expériences (D-optimal). Les facteurs retenus sont : la masse de A, la température et la durée d'alimentation. Afin de reproduire les conditions réelles, les constantes de vitesse sont variées selon un plan d'expériences factoriel complet. Du bruit uniforme multiplicatif (1%) est ajouté aux spectres. Le set de validation (5 expériences) est constitué en modifiant aléatoirement les niveaux des facteurs dans le domaine expérimental.

Après simulation, les spectres sont filtrés et l'Evolving Factor Analysis est réalisée avec le logiciel PCM_Toolbox [3]. La fin de la réaction est déterminée sur la base de la backward-EFA et de l'autocorrélation des scores [1,4]. La matrice X et le vecteur Y sont ainsi formés en adjoignant les matrices des 10 expériences. Les différents facteurs pris en compte lors du calcul des modèles sont :

- Le nombre de variables latentes (#LV)
- X et Y centrés sur la moyenne (prétraitement A)
- Standard Normal Variate sur X (prétraitement B)

Les paramètres des modèles sont estimés à l'aide de la PLS-Toolbox [5]. Selon l'Equation 1, on s'attend à un rang de 3 pour les matrices centrées sur la moyenne et un rang de 4 pour SNV sur X [6]. La qualité des modèles est déterminée à l'aide de l'efficacité, du taux de faux positifs et du taux de faux négatifs [7]. Ces indicateurs se calculent de la manière suivante :

Indicateur	Expression
Efficacité :	$\frac{TP + TN}{n_{tot}}$
Taux de faux positifs :	$\frac{FP}{TN + FP}$
Taux de faux négatifs :	$\frac{FN}{TP + FN}$

TP = vrai positif ; TN = vrai négatif ; FN = faux négatif ;
 FP = faux positif ; n_{tot} = nombre total d'observations

Table 1: Calcul des performances des modèles

Le taux de faux positifs permet de mesurer le taux de spectres pour lesquels le modèle indique de manière erronée la fin de la réaction principale. La présence de réactif B n'étant pas souhaitée, il faut donc que ce taux soit le moins élevé possible. Si l'on attend suffisamment longtemps, ce taux atteindra la valeur 0. A l'opposé, le taux de faux négatifs permet de mesurer le taux de spectres pour lesquels le modèle n'indique pas la fin de la réaction alors que cette dernière est déjà terminée.

On peut ainsi formuler un problème d'optimisation comme suit: minimiser le taux de faux positifs en limitant le taux de faux négatifs à un seuil supérieur donné.

3. Résultats

Formation du set de calibration

a) Détermination de la fin de réaction

La Figure 1 représente la b-EFA et l'autocorrélation des scores pour l'expérience n° 1 du set de calibration. L'événement de fin de réaction ($t = 52$ min) est indiqué par le trait noir vertical. A noter que la fin réelle de la réaction est à $t = 54$ min. Cette différence dépend fortement du rapport signal-bruit.

b) Formation des matrices

La matrice X , composée des spectres IR filtrés des 10 expériences, est de taille 1210×101 et le vecteur Y de taille 1210×1 .

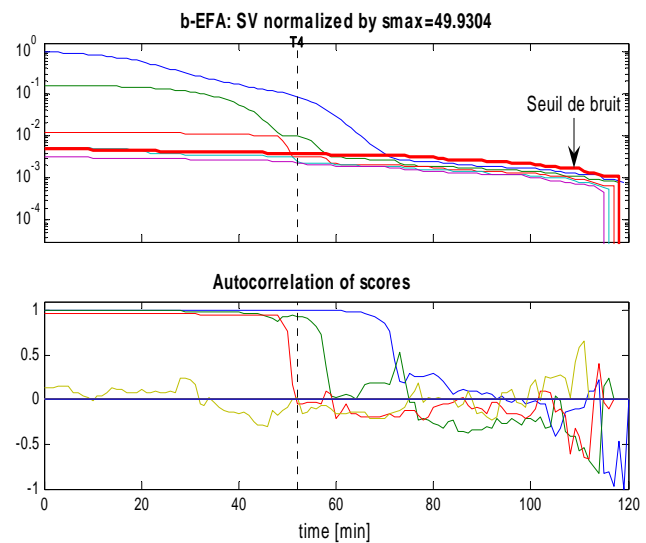


Figure 1: b-EFA et autocorrélation des scores

Construction du modèle de classification

Le seuil entre deux classes est déterminé à l'aide d'une fonction basée sur la statistique Bayésienne. Une limite séparant au mieux les classes, avec la probabilité minimale d'effectuer une fausse classification lors d'une prochaine prédiction, est ainsi recherchée [5].

Prédiction sur le set de validation

Résultats de la classification des spectres IR pour le set de validation :

#LV	A			B		
	Efficacité [%]	Faux positifs [%]	Faux négatifs [%]	Efficacité [%]	Faux positifs [%]	Faux négatifs [%]
2	97.9	0.00	3.81	94.4	12.9	0.00
3	99.0	0.76	1.17	98.0	0.00	3.52
4	99.0	0.38	1.47	99.0	1.14	0.88
5	99.0	0.00	1.76	98.5	1.13	1.76

Table 2 : Efficacités et taux de faux négatifs/positifs des modèles

En supposant un retard d'une minute lors de la détection de la fin de réaction, le seuil supérieur pour le taux de FN vaudra 1.54 %. Selon les critères d'optimisation définis

précédemment, c'est le modèle obtenu avec les données centrées (prétraitement A) et 4 LVs qui doit être retenu. Pour le prétraitement B, le meilleur modèle est également obtenu avec 4 LVs. En observant les pourcentages de faux positifs, on constate que le modèle avec le prétraitement A prédira moins fréquemment une fin de réaction anticipée que le modèle avec le prétraitement B.

La Figure 2 représente les valeurs de références et les valeurs prédites pour l'expérience 1 du set de validation. Le nombre LVs retenu est de 4 (prétraitement B). Les valeurs de références étant symbolisées par la ligne bleue et les valeurs prédites par les points verts. Le seuil est indiqué à l'aide de la ligne horizontale rouge. L'instant de fin de réaction correspond au passage de la valeur 0 à la valeur 1. On remarque que les erreurs de prédiction dépendent de la valeur du seuil.

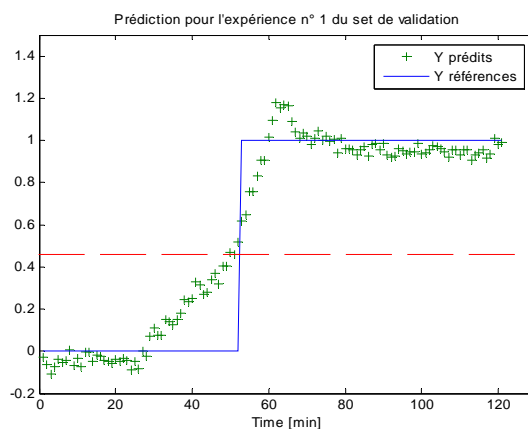


Figure 2: Y références et Y prédits pour le modèle à 4 LVs (SNV sur X)

4. Conclusion

L'utilisation de la PLS-DA pour estimer la fin de réaction à partir des spectres IR simulés a permis d'obtenir une bonne efficacité de classification. Le modèle basé sur les données centrées donne de très faibles taux de faux positifs/négatifs. Un retard d'une minute (taux de FN = 1.54%) dans la détection de la fin d'une réaction chimique est tout à fait acceptable. Parallèlement à cette étude basée sur des simulations, une réaction chimique témoin (synthèse du cinnamaldéhyde) de même qu'une réaction de fermentation vont être utilisées pour tester la méthode proposée.

L'application de cette technique n'est pas uniquement destinée à la détection de fin de réaction, mais elle est également applicable à la détection des débuts de réactions principales ou secondaires. Un transfert dans un autre domaine que les réactions, comme par exemple la cristallisation (début/fin) ou le séchage (fin), peut être envisagé. Actuellement, il existe des applications de la PLS-DA dans l'industrie des semi-conducteurs, notamment dans la détection de la fin de processus de gravure sèche [8].

5. Références

- [1] M. Amrhein, G. Corninboeuf, O. Naef, 'Etude chimiométrique sans calibration des données spectrales et analytiques provenant de réactions chimiques', *Chimiométrie* **2004**, Paris
- [2] L. Chiang, E. Russell, R. Braatz, *Chemometrics and Intelligent Laboratory Systems* **50** (2000), 243-252
- [3] PCM_Toolbox, Online Control Sàrl, **1003** Lausanne, Suisse
- [4] E. R. Malinowski, 'Factor Analysis in Chemistry', 2nd edition, John Wiley & Sons, New York, **1991**
- [5] PLS_Toolbox, version 3.5, Eigenvector Research, Inc
- [6] M. Amrhein, 'Reaction and Flow Variants/Invariants for the Analysis of Chemical Reaction Data', thèse No 1861, EPF Lausanne, **1998**
- [7] S.L.R. Ellison, T. Fearn, 'Characterizing the performance of qualitative analytical methods: Statistics and terminology', *Trends in Analytical Chemistry*, Vol. 24, No. 6, **2005**, 468-476
- [8] K. Harvey, J. Hosch, N. Gallagher, B. Wise, 'System and Method for Determining Endpoint in Etch Processes using Partial Least Squares Discriminant Analysis in the Time Domain of Optical Emission Spectra', US Patent n° 2004/0045934 A1, **2004**