



**HAL**  
open science

## Détection d'hétérogénéité au sein de mesures de qualité de l'environnement

Carole Langlois-Jacques, Julien Jacques

► **To cite this version:**

Carole Langlois-Jacques, Julien Jacques. Détection d'hétérogénéité au sein de mesures de qualité de l'environnement. La revue MODULAD, 2009, 40, pp.41-52. hal-00375159

**HAL Id: hal-00375159**

**<https://hal.science/hal-00375159>**

Submitted on 14 Apr 2009

**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.

# Détection d'hétérogénéité au sein de mesures de qualité de l'environnement

Carole Langlois-Jacques<sup>1</sup>, Julien Jacques<sup>2</sup>

14 avril 2009

AGLAE (Association Générale des Laboratoires d'Analyse de l'Environnement), 629 Av. de la République, 59000 Lille, France  
Laboratoire Paul Painlevé UMR CNRS 8524, Université Lille I, 59655 Villeneuve d'Ascq Cedex, France

## Résumé

L'accréditation des laboratoires d'analyses de l'environnement nécessite un contrôle qualité externe des mesures qu'ils effectuent. Pour ce faire, AGLAE réalise des campagnes d'essais consistant à envoyer des échantillons à mesurer par les laboratoires. Sous l'hypothèse que les mesures retournées par les laboratoires sont gaussiennes, un intervalle de confiance est construit sur les résultats de mesure de sorte à étudier la performance analytique de chaque laboratoire. Il arrive toutefois que l'hypothèse de normalité ne soit pas satisfaite, ce qu'il est indispensable de prendre en compte lors de la construction de l'intervalle de confiance. L'intérêt de ce travail consiste à détecter une hétérogénéité au sein des mesures des laboratoires, que nous caractérisons par la présence de plusieurs sous-populations gaussiennes au sein des mesures. Le logiciel MIXMOD est alors utilisé pour identifier la présence de ces sous-populations et, le cas échéant, estimer leurs paramètres.

## 1 Introduction

Dans le but d'obtenir leur accréditation, les laboratoires d'analyse de l'environnement participent à des comparaisons inter-laboratoires. Organisée chaque année par l'Association Générale des Laboratoires d'Analyse de l'Environnement (AGLAE), une campagne d'essais d'aptitude est destinée à évaluer les performances des laboratoires. La performance d'un laboratoire est jugée en terme de biais (*justesse*) et de variance (*fidélité*) des mesures

effectuées. Etant donné que la vraie valeur d'un paramètre à mesurer sur un matériau environnemental ne peut pas toujours être connue avec précision, la justesse d'un laboratoire est analysée de façon relative : un laboratoire sera considéré sans biais si ses résultats de mesure coïncident avec la moyenne de tous les laboratoires. La fidélité d'un laboratoire est quant à elle analysée à partir de mesures répétées dans des conditions de répétabilité : même échantillon de matériau mesuré par le même technicien à l'aide du même appareil de mesure et dans un intervalle de temps court.

Nous présentons dans la section 2 le principe des essais inter-laboratoires, c'est-à-dire les consignes à respecter par les laboratoires participant à la campagne d'essai, la nature des mesures rendues par les laboratoires, ainsi que la méthodologie employée par AGLAE pour évaluer la performance de chaque laboratoire. Un exemple d'essai inter-laboratoires illustre ces propos. Nous montrons ensuite qu'en présence d'une hétérogénéité au sein des mesures des laboratoires, cette méthodologie d'évaluation des performances doit être adaptée, et conduit à résoudre un problème de classification automatique. Une application réelle utilisant le logiciel MIXMOD conclut cet article.

## 2 Méthodologie des essais inter-laboratoires

### 2.1 Organisation générale

Le matériau (eau, boue, sédiment, déchet...) soumis à l'analyse des laboratoires est dans un premier temps élaboré par un organisme extérieur sélectionné par AGLAE, dans un souci de stabilité et d'homogénéité. Un lot d'échantillons est préparé à partir de ce matériau, qui est ensuite envoyé aux laboratoires pour analyse : mesures physiques, chimiques (dosages) ou biologiques (dénombrements et mesures de doses létales). Nous nous intéressons dans ce travail uniquement aux dosages chimiques et aux mesures physiques. A réception des échantillons, les laboratoires réalisent leur analyse dans une période raisonnable de traitement spécifiée par AGLAE, durant laquelle la stabilité de la substance est avérée. Afin d'accéder à la fidélité de l'analyse et de prévenir de l'hétérogénéité des échantillons au sein du lot considéré, chaque laboratoire reçoit deux échantillons pour lesquels AGLAE demande d'effectuer deux répliques de l'analyse dans des conditions de répétabilité. Les résultats sont ensuite envoyés à AGLAE qui est chargée de l'analyse statistique.

## 2.2 Les données

Soit  $x_{ijl}$  la  $j$ -ème réplique de l'analyse du  $i$ -ème échantillon par le  $l$ -ème laboratoire ( $1 \leq j \leq 2$ ,  $1 \leq i \leq 2L$ ,  $1 \leq l \leq L$  où  $L$  est le nombre de laboratoires ayant participé à l'essai). On suppose que l'effet des sources d'erreur sur cette mesure est additif et sans effet d'interaction. Ainsi,  $x_{ijl}$  est supposé se décomposer suivant le modèle linéaire suivant :

$$x_{ijl} = \mu + \alpha_l + \beta_i + \epsilon_{ijl}$$

où

- $\mu$  est la vraie valeur du paramètre à mesurer, souvent difficile à connaître de façon précise,
- $\alpha_l$  est le biais du  $l$ -ème laboratoire (erreur inter-laboratoire),
- $\beta_i$  est l'erreur due à l'hétérogénéité du lot d'échantillons,
- $\epsilon_{ijl}$  est l'erreur de mesure aléatoire induite par la  $j$ -ème réplique de l'analyse.

Dans cet article, nous restreignons notre étude à une comparaison inter-laboratoires. Ainsi, plutôt que de travailler sur les valeurs individuelles  $x_{ijl}$ , nous travaillons avec les valeurs moyennes rendues par chaque laboratoire :

$$x_l = \frac{1}{4} \sum_{i,j=1}^2 x_{ijl}.$$

## 2.3 Evaluation des performances des laboratoires

Dans le domaine des analyses chimiques, l'hypothèse de base est que les moyennes  $x_l$  des laboratoires sont des réalisations indépendantes d'une variable aléatoire gaussienne de moyenne  $\mu$ , la vraie valeur du paramètre à mesurer, et de variance  $\sigma^2$ . Sous cette hypothèse, il devient naturel de définir des intervalles de confiances autour de la valeur de référence  $\mu$  dans lesquels devraient se trouver les mesures des laboratoires :

- $IC_1 = [\mu - 2\sigma; \mu + 2\sigma]$ ,
- $IC_2 = [\mu - 3\sigma; \mu + 3\sigma]$ .

Les intervalles  $IC_1$  et  $IC_2$  sont des intervalles de confiances respectives 95.45% et 99.73%.

En pratique, nous avons vu que la vraie valeur  $\mu$  d'un paramètre sur un matériau environnemental ne pouvait pas toujours être connue de façon précise. Pour les essais réalisés par AGLAE, la valeur  $\mu$  est estimée par une valeur de consensus issue des données retournées par les laboratoires ayant respecté la période raisonnable de traitement de l'échantillon : en pratique,

cette valeur est la moyenne  $m$  des moyennes  $x_l$  des laboratoires, calculée par un algorithme d'estimation robuste [1] que nous ne détaillons pas dans ce papier. Ainsi, la fidélité d'un laboratoire est jugée relativement aux résultats des autres laboratoires. De même, la variance  $\sigma^2$  est estimée par la variance  $s^2$  des moyennes  $x_l$  des laboratoires, calculée également de façon robuste [1]. Un score est ensuite construit pour chaque laboratoire :

$$z_l = \frac{x_l - m}{s}.$$

Ce score est interprété de la façon suivante par AGLAE :

- $|z| < 2$  : le résultat du laboratoire est non différent de ceux de l'ensemble des laboratoires. Nous écrivons dans cet article que le laboratoire est dans les *limites d'acceptabilité*.
- $2 \leq |z| < 3$  : le résultat du laboratoire est légèrement différent de ceux de l'ensemble des laboratoires. Nous écrivons que le laboratoire est dans les *limites de tolérance*.
- $|z| \geq 3$  : le résultat du laboratoire est notablement différent de ceux de l'ensemble des laboratoires.

## 2.4 Vérification de l'hypothèse de normalité sur un exemple d'essai inter-laboratoires

Nous présentons dans cette section l'analyse de l'essai inter-laboratoires n°1 consistant à mesurer une teneur en cuivre sur une eau destinée à la consommation humaine, auquel 119 laboratoires ont participé et rendu des résultats dans les dates imparties. La fonction de répartition empirique des résultats moyens des laboratoires est présentée sur la figure 1.

Avant d'estimer la valeur de consensus  $m$  ainsi que la variance  $s^2$  pour en déduire les scores des laboratoires, il convient de s'assurer de la normalité des données.

L'examen visuel de l'histogramme ainsi que de la droite de Henry (figure 2) peut nous faire douter de cette hypothèse puisque quatre données semblent particulièrement élevées par rapport aux autres (mesures moyennes supérieures à 300).

En effet, la présence de 4 données extrêmes parmi 119 reflète une distribution à queues plus lourdes que la loi normale. Ceci est confirmé par les tests<sup>1</sup> de Shapiro-Wilk et de Kolmogorov-Smirnov qui rejettent la normalité avec des *p-value* respectives de 0.003 et  $2e - 14$ . Toutefois, en supprimant les quatre données décelées à l'examen visuel, la normalité est acceptée par les deux

---

1. réalisés à l'aide des fonction `shapiro.test` et `ks.test` sur le logiciel R

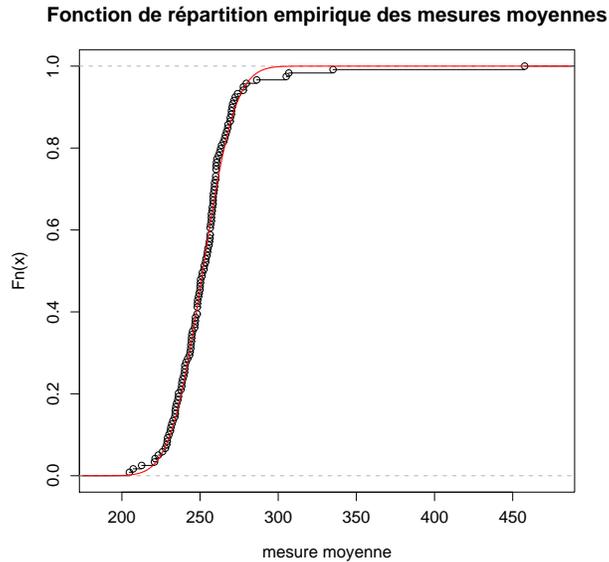


FIG. 1 – Essai inter-laboratoires n° 1 : mesure de la teneur en cuivre ( $\mu\text{g/l}$ ) sur une eau destinée à la consommation humaine.

tests ( $p$ -value de 0.59 et 0.29). Sachant que les algorithmes d'estimation de  $m$  et  $s^2$  utilisés par AGLAE sont des algorithmes robustes, l'effet de ces quatre laboratoires atypiques sera atténué. Nous décidons donc d'appliquer la méthodologie d'évaluation des performances présentée précédemment.

Les estimations des valeurs de consensus  $m = 251.78$  et variance  $s^2 = 16.28^2$  conduisent à classer les laboratoires de la façon suivante :

- 111 laboratoires sont dans les limites d'acceptabilité,
- 4 laboratoires sont hors des limites d'acceptabilité mais dans les limites de tolérance,
- 4 laboratoires sont hors des limites de tolérance.

Les quatre laboratoires hors limites de tolérance sont les quatre laboratoires que l'on avait repérés comme atypiques dans l'étude de la normalité des données, et que l'on avait dû supprimer de la série pour qu'elle vérifie l'hypothèse de normalité.

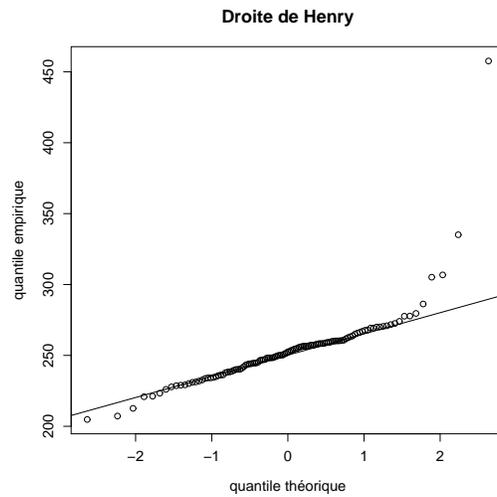
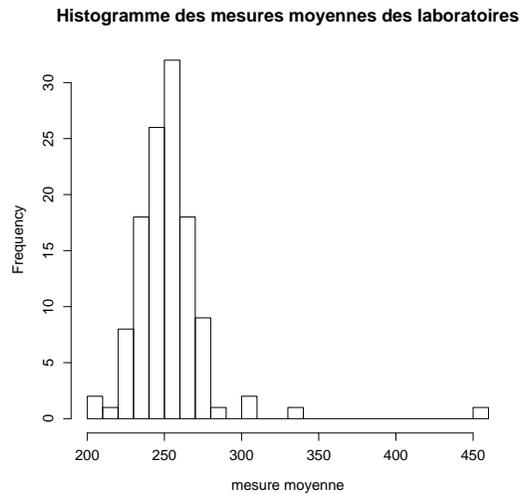


FIG. 2 – Histogramme et droite de Henry des données de l’essai inter-laboratoires n°1.

### 3 Détection d’hétérogénéité à l’aide de modèles de mélange

#### 3.1 Problématique

Comme on peut le constater pour l’essai inter-laboratoires n°2 qui sera détaillé dans la section 4 (mesure de la teneur en aluminium sur des sédiments),

l'hypothèse de normalité faite précédemment ne peut pas toujours être vérifiée même en supprimant quelques valeurs atypiques. Les résultats de l'essai n°2 sont présentés sur la Figure 3.

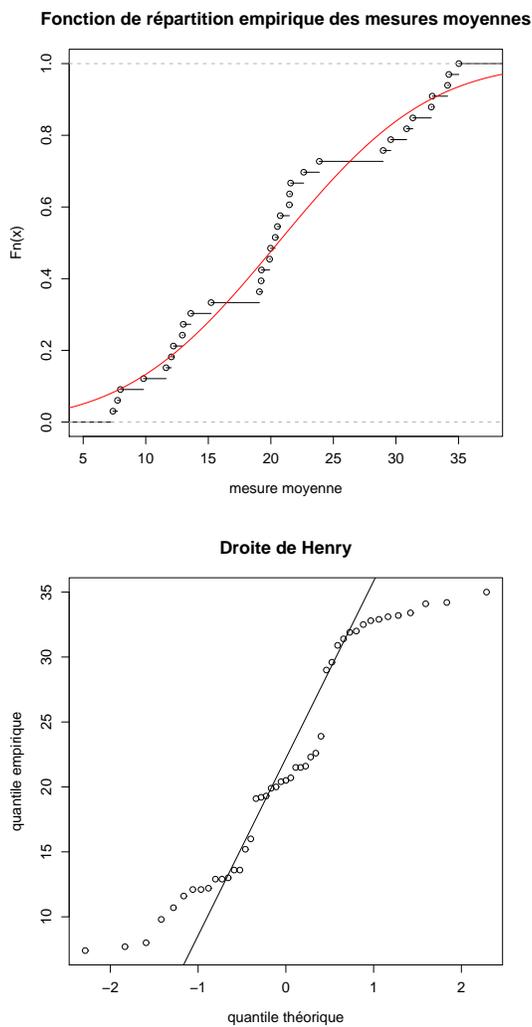


FIG. 3 – Essai inter-laboratoires n°2 : mesure de la teneur en aluminium (g/kg) sur une matrice de sédiments de rivière. Fonction de répartition et droite de Henry.

Il semble à l'examen de la droite de Henry que trois sous-populations soient

présentes au sein des données. Dans ce cas, la méthodologie d'évaluation des performances des laboratoires présentée précédemment ne peut plus être employée. En effet, l'importante disparité des données conduirait à construire des intervalles de confiance trop grands donc trop conservateurs, ce qui ne refléterait pas la réalité : des laboratoires risquent de se retrouver dans les limites d'acceptabilité alors qu'ils devraient être dans les limites de tolérance ou même en dehors de ces dernières.

La recherche d'explications à ce phénomène de non normalité a conduit les spécialistes des essais inter-laboratoires à l'explication suivante : il semblerait que cette hétérogénéité au sein des données puisse parfois être expliquée par des différences entre méthodes d'analyse employées par les laboratoires (cf. section 4 pour des exemples de méthodes). Ainsi, les données ne seraient plus expliquées par une loi normale mais par un mélange de lois normales, dont chaque composante serait expliquée par une ou un ensemble de méthodes d'analyse. Toutefois, il est fréquent que la présence d'hétérogénéité au sein des données ne puisse pas être expliquée par la différence entre méthodes d'analyse. Dans ce cas, l'identification de groupes au sein des données doit être faite en utilisant uniquement les résultats de mesures. On est alors en présence d'un problème de classification automatique dans lequel le nombre de groupes n'est pas connu *a priori*.

L'objectif de ce travail consiste donc à modéliser les résultats de mesure par un mélange de gaussiennes, puis à appliquer la stratégie d'évaluation des performances des laboratoires au sein de chaque composante du mélange obtenu.

### 3.2 Modèles de mélange gaussien univariés

Aux moyennes des laboratoires  $x_l$  s'ajoutent désormais les appartenances  $z_l$  des laboratoires à l'une des  $K$  composantes du mélange. Ces appartenances sont supposées être la réalisation d'une variable multinomiale  $Z$  :

$$Z \sim \mathcal{M}(1, \pi_1, \dots, \pi_K)$$

où  $\pi_k$  est la proportion de la  $k$ -ème composante du mélange.

La moyenne  $x_l$  du  $l$ -ème laboratoire est la réalisation d'une variable  $X$  de densité mélange

$$f(x) = \sum_{k=1}^K \pi_k \phi(x; \mu_k, \sigma_k^2),$$

où  $\phi(\cdot; \mu_k, \sigma_k^2)$  est la densité gaussienne unidimensionnelle de la  $k$ -ème composante, d'espérance  $\mu_k$  et de variance  $\sigma_k^2$ . L'estimation de ce modèle de

mélange par maximum de vraisemblance via l'algorithme EM est réalisée à l'aide du logiciel MIXMOD [5], dédié entre autre à l'estimation de modèles de mélange gaussien pour la classification automatique.

La détermination du nombre de composantes du mélange, qui est le point crucial de notre travail puisque c'est ce nombre qui déterminera s'il faut évaluer les laboratoires ensemble ou séparément, sera effectué par une approche choix de modèle. Le logiciel MIXMOD a l'avantage de proposer plusieurs critères de choix de modèles pour les problématiques de classification automatique : le critère BIC (Bayesian Information Criterion, [7]), le critère ICL (Integrated Completed Likelihood, [4]) et le critère NEC (Normalized Entropy Criterion, [6]). D'après la synthèse présentée dans [5], le critère BIC doit être préféré lorsqu'il s'agit d'estimer une densité, ce qui ne constitue pas notre préoccupation première dans ce travail. Nous nous tournons donc vers les critères ICL et NEC, avec une préférence pour ICL qui est particulièrement dévolu à mettre en évidence des groupes bien séparés au sein des données ([4]).

### 3.3 Evaluation de la performance des laboratoires en présence d'hétérogénéité

Que les groupes soient formés par les différences de méthode ou par classification automatique, les laboratoires sont regroupés en classes homogènes d'un point de vue de leurs résultats de mesure. Ainsi, l'évaluation des performances des laboratoires décrite dans la section 2.3 est réalisée séparément pour chaque classe de laboratoires. C'est-à-dire qu'AGLAE effectue les comparaisons inter-laboratoires par groupe, et calcule les scores des laboratoires en utilisant des valeurs de consensus  $m_k$  et des variance  $s_k^2$  propres à chaque groupe homogène  $k$ .

## 4 Utilisation de MIXMOD sur l'essai inter-laboratoires n°2

### 4.1 Présentation de l'essai

Cet essai inter-laboratoires est basé sur la mesure de la teneur en aluminium (g/kg) sur une matrice de sédiments de rivière. Les résultats moyens des 45 laboratoires ayant participé à cet essai sont présentés par la figure 4, en fonction de la méthode de mesure utilisée.

Pour cet essai, les 7 méthodes de mesures utilisées sont les suivantes :

1. extraction avec acide chlorhydrique et acide nitrique sous reflux,

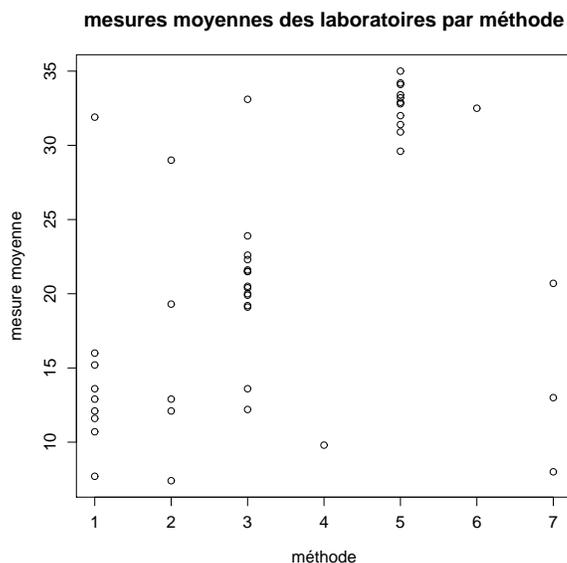


FIG. 4 – Essai inter-laboratoires n°2 : mesure de la teneur en aluminium sur des sédiments de rivière.

2. extraction avec acide chlorhydrique et acide nitrique sous ébullition dans un tube en verre,
3. extraction avec acide chlorhydrique et acide nitrique en four micro-ondes fermé,
4. extraction avec acide chlorhydrique et acide nitrique en four micro-ondes ouvert,
5. attaque avec acide fluorhydrique et acide perchlorique,
6. fusion alcaline avec fondants oxydants,
7. autres.

Dans les sédiments, l'aluminium est lié à divers composés (en particulier des phyllosilicates hydratés, constituants principaux des argiles), desquelles il est nécessaire de l'extraire avant d'analyser l'échantillon. Pour cela, les laboratoires utilisent différentes méthodes d'extraction. Ces méthodes consistent à utiliser des acides plus ou moins forts et de chauffer d'une façon ou d'une autre. Ainsi, au sein d'une même méthode, il peut y avoir des dispersions des résultats s'expliquant par des puissances de niveaux énergétiques différents. Par exemple pour la méthode numéro 3, il est possible que les laboratoires utilisent des micro-ondes de puissances différentes.

## 4.2 Ajustement de la série de données de l'essai n°2 par mélange gaussien

La non-normalité des données dont on peut avoir l'intuition visuellement, est confirmée par un test de Shapiro-Wilk avec une *p-value* de 0.002. Toutefois, un test de Kolmogorov-Smirnov infirme cette hypothèse de non-normalité : on accepte la normalité avec une *p-value* de 0.30. Bien que le test de Shapiro-Wilk soit connu pour être le plus fiable, surtout pour un échantillon de taille peu importante comme c'est le cas ici (45), cette différence nous encourage à aller plus en avant dans l'analyse.

L'examen visuel de la répartition des moyennes des laboratoires en fonction des méthodes de mesures (figure 4) semble montrer que les méthodes de mesures ont une influence sur les résultats des laboratoires : en effet, la méthode de mesure numéro 5, utilisée par 11 laboratoires sur 45, donne des résultats plus élevés que les autres méthodes, avec une variabilité relativement réduite. Les deux autres méthodes majoritaires, numéros 1 et 3, donnent des résultats plus faibles mais de variabilités plus importantes. En outre, les résultats sont encore plus faibles pour la méthode numéro 1 que la méthode numéro 3. Une analyse de variance<sup>2</sup> permet de confirmer l'influence du facteur méthode d'analyse sur le résultat, avec un test d'indépendance donnant une *p-value* de 0.001.

Ainsi, les analyses préliminaires semblent conclure que les méthodes de mesures ont une influence sur les résultats de mesure, qui ne peuvent donc pas être ajustés par une simple loi normale.

En parallèle, le logiciel MIXMOD est utilisé pour réaliser une classification automatique des données à l'aide d'un modèle de mélange gaussien à nombre de composantes inconnu. Le paramétrage utilisé pour le logiciel MIXMOD est le suivant :

- aucune contrainte sur le modèle de mélange gaussien n'est imposée.
- bien que nous portions notre préférence sur le critère de choix de modèles ICL, le critère NEC est également utilisé.
- nous permettons au nombre de composantes du mélange d'évoluer entre 1 et 5.
- l'initialisation de l'algorithme EM est réalisée par la stratégie *small\_EM*, qui consiste à utiliser comme valeur initiale la meilleure obtenue suite à l'exécution sur seulement quelques itérations de plusieurs algorithmes EM, initialisés eux-mêmes aléatoirement.

Le modèle choisi par le critère ICL est un modèle à trois composantes à proportions et variances différentes par groupe. Ce modèle est noté  $p_k L_k C$

---

2. réalisée à l'aide de la fonction `gm` sur le logiciel R

dans MIXMOD et  $[p_k \lambda_k D A D']$  dans [5], où  $D$  et  $A$  sont les matrices d'orientation et de forme, ici réduites à des réels ( $= 1$ ) puisque les mesures  $X$  sont en dimension 1. La figure 5 représente l'ajustement du mélange sur l'histogramme des données, ainsi que la répartition des résultats des laboratoires dans les 3 groupes obtenus par MIXMOD.

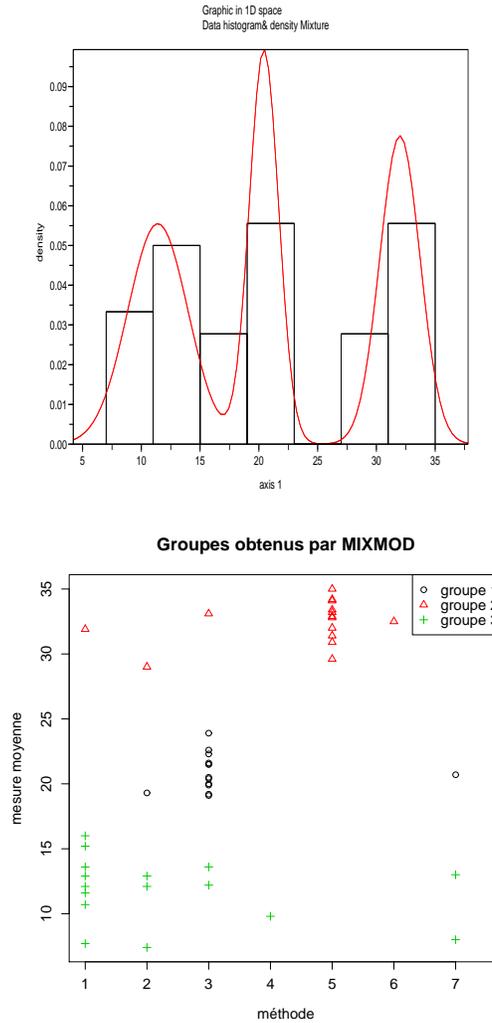


FIG. 5 – Résultats MIXMOD de l'ajustement d'un mélange gaussien sur les mesures de l'essai inter-laboratoires n°2.

On constate que les 3 groupes obtenus par classification automatique correspondent essentiellement aux 3 principales méthodes de mesures (représentant 77% des laboratoires) : le premier groupe correspond à la méthode 3, le deuxième groupe à la méthode 5 et le troisième groupe à la méthode 1. Cette explication est confirmée par un test d'indépendance<sup>3</sup> du  $\chi^2$  entre les labels des laboratoires obtenus par classification automatique et les méthodes de mesures, qui conclut au rejet de l'hypothèse d'indépendance avec une *p-value* de  $7e - 08$ .

Les paramètres de chacune des composantes du mélange sont présentés dans le tableau 1.

	composante 1	composante 2	composante 3
proportion	0.31	0.33	0.36
moyenne	20.43	32	11.40
variance	1.55	2.93	6.56

TAB. 1 – Paramètres du mélange gaussien estimé par MIXMOD pour l'essai inter-laboratoires n°2.

Remarque : si ICL choisit un modèle à trois composantes, le critère NEC choisit un modèle à une seule composante. On retrouve l'incertitude mise en évidence par les conclusions différentes des deux tests de normalité effectués. Néanmoins, pour les raisons que nous avons évoquées précédemment, nous privilégions le critère ICL. De plus, le critère NEC n'étant pas défini pour un mélange à une seule composante, le choix d'un tel modèle par NEC est fait suivant une règle empirique parfois criticable selon ses auteurs [3].

### 4.3 Evaluation des performances des laboratoires pour l'essai inter-laboratoires n°2

Le meilleur modèle de mélange gaussien pour les mesures moyennes des laboratoires étant un modèle à 3 composantes, les laboratoires sont évalués par groupe. Comme les groupes obtenus par classification automatique s'expliquent relativement bien par les méthodes d'analyse, deux options sont envisagées : évaluer les laboratoires par groupe de méthodes ou par groupe obtenus par classification automatique. Naturellement, si la composition de ces groupes est similaire pour une majorité des laboratoires, c'est au niveau des laboratoires les plus atypiques qu'apparaissent les différences.

---

3. réalisée à l'aide de la fonction `chisq.test` sur le logiciel R

En utilisant les groupes obtenus par classification automatique, seul un laboratoire sur les 45 sort des limites d'acceptabilité tout en restant dans les limites de tolérance, avec un score=2.77. Les résultats des laboratoires ainsi que les limites d'acceptabilité et de tolérance sont représentés sur la figure 6.

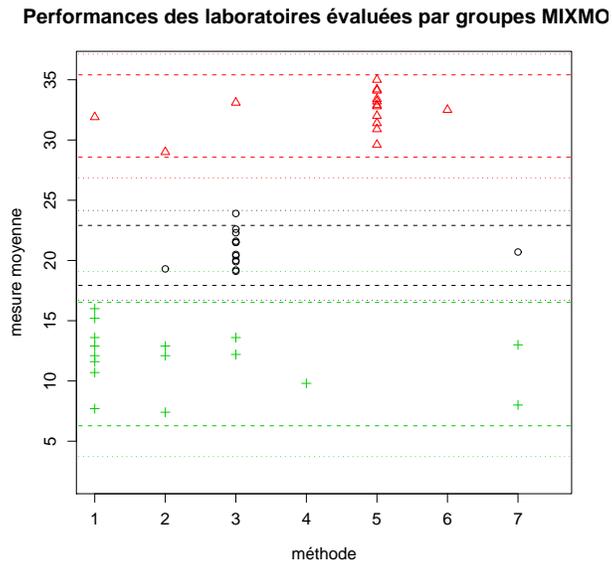


FIG. 6 – Performances des laboratoires évaluées en les regroupant par classification automatique (limites d'acceptabilité en tiret et de tolérance en pointillé).

Cette analyse n'est pas jugée satisfaisante par les spécialistes d'AGLAE parce que la classification automatique sépare des laboratoires qui ont utilisé la même méthode d'analyse sachant que l'effet méthode est significatif. En s'appuyant sur les résultats de classification automatique et sur les spécificités techniques des méthodes d'analyse, les 3 groupes de méthodes de mesures suivants sont créés :

- groupe 1 : méthode numéro 3,
- groupe 2 : méthodes numéro 5 et 6,
- groupe 3 : méthodes numéro 1, 2, 4 et 7.

Les résultats d'évaluation des performances des laboratoires, présentés par la figure 7, conduisent à classer deux laboratoires dans les limites d'accep-

tabilité et deux laboratoires dans les limites de tolérance. Ce résultat est satisfaisant pour les spécialistes des essais inter-laboratoires dans le domaine de la chimie.

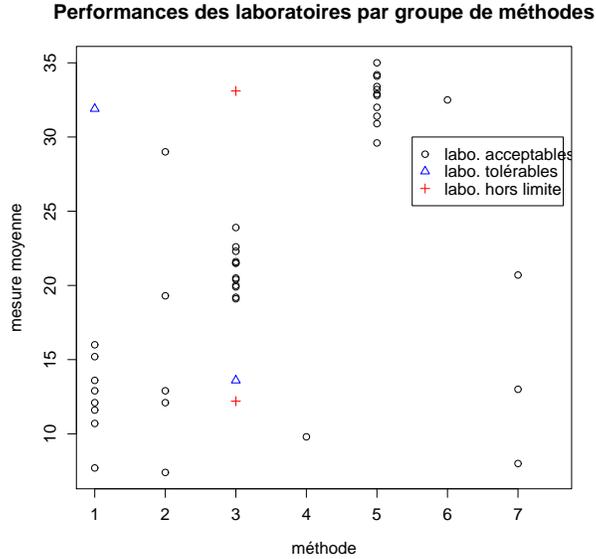


FIG. 7 – Performances des laboratoires évaluées par groupe de méthodes de mesure.

## 5 Conclusion

Nous avons montré dans cet article comment l'utilisation des modèles de mélanges gaussiens *via* le logiciel MIXMOD pouvait apporter une réponse à un problème d'hétérogénéité au sein de données d'essais inter-laboratoires. En effet, l'évaluation des performances analytiques des laboratoires, qui consiste en la construction d'intervalles de confiance, suppose la normalité des données. En présence d'hétérogénéité, que nous avons caractérisée par l'ajustement d'un mélange gaussien sur les résultats de mesure, les laboratoires doivent être évalués par groupe homogène. Si parfois la différence de méthode de mesures explique en partie cette hétérogénéité, ce n'est pas toujours le cas. Des *artefacts* tel l'arrondi pratiqué par les laboratoires sur les résultats de mesure peuvent également être à la source de distributions

multimodales. Il est dès lors indispensable d'avoir recours à une procédure de classification automatique pour identifier les grandeurs mesurées (teneur en aluminium, teneur en chrome, ou autre...) affectées de manière récurrente par des hétérogénéités ; puis de tenter de mettre en relation les composantes du mélange gaussien avec les méthodes de mesure. Enfin, le logiciel MIXMOD propose plusieurs critères de choix de modèles, dont l'utilisation est primordiale dans notre étude puisque c'est sur ces critères que se base la décision de traiter les laboratoires ensemble ou séparément.

## Remerciements

Nous adressons nos remerciements à toute l'équipe MIXMOD pour leur disponibilité et pour les discussions fructueuses que nous avons pu avoir ensemble. Nous remercions également Philippe Guarini et Ronan Charpentier d'AGLAE pour leur collaboration à ce travail.

## Références

- [1] NF ISO 5725-5 : Exactitude (justesse et fidélité) des résultats et méthodes de mesure, AFNOR (1998).
- [2] Biernacki, C. and Govaert, G. : Choosing Models in Model-based Clustering and Discriminant Analysis. *Journal of Statistical Computation and Simulation*, **64** (1999) 49–71.
- [3] Biernacki, C., Celeux, G. and Govaert, G. : An Improvement of the NEC Criterion for Assessing the Number of Clusters in a Mixture Model. *Pattern Recognition Letters*, **20**[3] (1999) 267–272.
- [4] Biernacki, C., Celeux, G. and Govaert, G. : Assessing a Mixture Model for Clustering with the Integrated Completed Likelihood. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **22**[7] (2000) 719–725.
- [5] Biernacki, C., Celeux, G., Anwuli, A., Govaert, G. and Langrognet, F. : Le logiciel MIXMOD d'analyse de mélange pour la classification et l'analyse discriminante. *La Revue de Modulad*, **35** (2006) 25–44.
- [6] Celeux, G. and Soromenho, G. : An entropy criterion for assessing the number of clusters in a mixture model. *Journal of Classification*, **13** (1996) 195–212.
- [7] Schwarz, G. : Estimating the dimension of a mode. *The Annals of Statistics*, **6**[2] (1978), 461–464.