

## EVALUATION DES PERFORMANCES INDIVIDUELLES D'UN LABORATOIRE AU MOYEN DES RESULTATS DE CHAINES D'ANALYSES CECALAIT

Les essais interlaboratoires sont utilisés à diverses fins dont les trois principales sont :

- l'évaluation des méthodes d'analyse; dans ce cadre les laboratoires sont supposés être familiarisés et entraînés avec une méthode définie bien décrite au préalable.
- la détermination de "valeurs vraies" garanties pour des matériels de référence; dans ce cadre, la(les) méthode(s) est(sont) parfaitement connue(s) et décrite(s), de même, les laboratoires sont reconnus pour leur compétence.
- l'évaluation des laboratoires dans le contexte général de l'Assurance Qualité Analyse; ici la(les) méthode(s) est(sont) parfaitement connue(s) sans biais de justesse entre elles, officielle(s) ou normalisée(s) ou validée(s), mais les laboratoires sont les objets de l'investigation.

### MISE EN PLACE ET PRINCIPE DE L'ESSAI

L'essai interlaboratoire consiste à envoyer au même moment à chacun des laboratoires inscrits pour la chaîne les mêmes échantillons (6 à 10 dans le cadre de CECALAIT) et régulièrement répartis en taux sur la plage de mesure pour y être analysés par une méthode d'analyse donnée. Le laboratoire organisateur de l'essai, après avoir regroupé l'ensemble des résultats, procède à une analyse statistique des données, mettant en avant les différentes valeurs de précision obtenues par chacun des laboratoires. Les valeurs sont restituées aux participants sous forme de tableaux généraux, où ils ne sont identifiés que par un numéro, de manière à respecter l'anonymat de chacun. Des normes de tolérance accompagnent les résultats et facilitent l'autoévaluation des laboratoires.

Les traitements statistiques décrits ci-dessous sont valables aussi bien pour les chaînes d'analyses physico-chimiques que pour les chaînes d'analyses microbiologiques. Dans ce dernier cas cependant, il sont effectués sur des données transformées en logarithmes décimaux. Les résultats sont présentés à la fois sur ces données transformées et sur les données d'origine en valeur relative, par rapport au niveau de population.

### EVALUATION DES PERFORMANCES

#### **La répétabilité : 2 éléments d'information : comparaison à une valeur référence ou aux valeurs du groupe**

La répétabilité est un critère utile au laboratoire qui tire parti cette information sur la base des conditions effectives de la manipulation.

Elle présente moins d'intérêt pour l'organisme d'accréditation car les "conditions effectives" de répétabilité ne sont pas maîtrisées par le laboratoire organisateur de l'essai et les résultats sont potentiellement biaisables, volontairement (choix des meilleurs doubles parmi plusieurs répétitions) ou non (subjectivité de l'estimation visuelle Gerber, par exemple). Néanmoins, si un

défait apparaît dans ces conditions, il est bien témoin d'un problème analytique et est un élément supplémentaire de diagnostic.

La plupart des chaînes d'analyse CECALAIT recommandent des analyses en double. Celles-ci permettent de calculer à partir de 10 échantillons habituels 10 écarts entre doubles individuels ainsi qu'un écart-type de répétabilité par laboratoire. Ces informations sont reportées dans le tableau "répétabilité - écarts entre doubles" (une ligne par laboratoire, une colonne par échantillon).

Les éléments de diagnostic sont :

- 1) la comparaison de chaque écart  $w$  entre doubles à la valeur  $r$ , normalisée (le cas échéant), ou tirée de la bibliographie CECALAIT.  $r$  est la valeur limite en-dessous de laquelle doivent se trouver 95% des écarts entre doubles. Elle ne devrait donc pas être dépassée.
- 2) la comparaison de chacune des valeurs d'écarts entre doubles à la valeur d'écart limite  $L$ , découlant du test de Cochran (voir note 1), calculée pour chacun des échantillons à partir de l'ensemble des écarts entre doubles des laboratoires (voir norme FIL 135B, p6 et note 1)

#### *Note 1 : Test de Cochran*

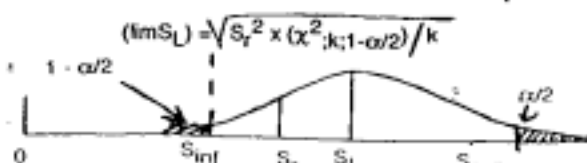
Chaque écart maximal observé par échantillon est testé par rapport à la somme des écarts au carré selon un rapport  $w^2_{max}$  sur somme des  $w^2$  ( $w^2/Sw^2$ ) qui doit rester inférieur à une valeur critique  $C$  de Cochran liée au nombre de laboratoires et au nombre de répétitions (tables de Cochran). Dans le cas contraire, il y a élimination de l'écart et itération du procédé. Si  $w^2/Sw^2 < C$ , on peut en déduire un seuil critique d'écart maximal entre doubles après arrêt du processus :

$$L = w^2_{max} = C \times Sw^2$$

Cette limite  $L$  fixe le seuil au-delà duquel on peut supposer, avec un risque d'erreur de 5%, que les écarts entre doubles n'appartiennent plus à la population des écarts entre doubles de cette méthode, pour le groupe de laboratoires en question. C'est à dire qu'ils sont a priori suspects, voire anormaux.

Contrairement à la valeur  $r$  qui reste fixe, en tant que référence dans l'absolu,  $L$  est liée au degré de performance du groupe. On constate que ces deux éléments d'information,  $r$  et  $L$ , sont complémentaires.

- c) la comparaison des écarts-types de répétabilité laboratoire  $S_L$  à la valeur normalisée ou tirée de la bibliographie-  $S_r$ . Compte tenu du fait que  $S_L$  est une estimation à partir de 10 doubles, on admet une tolérance, liée à l'intervalle de confiance de  $S_L$ , fonction du nombre de doubles et directement tirée de  $S_r$ .



La valeur  $S_r$  doit appartenir à l'intervalle de confiance de  $S_L$ , ou encore  $S_{inf} \leq S_r$ .

$$S_{sup} = \sqrt{SCE/x^2 \alpha/2} \quad \text{et} \quad S_{inf} = \sqrt{SCE/x^2 1-\alpha/2}$$

$$\Rightarrow S_r \geq \sqrt{SCE/x^2 1-\alpha/2} \Rightarrow S_r^2 \geq S_L^2 \times k/x^2 1-\alpha/2$$

$$\Rightarrow S_L^2 \leq S_r^2 \times x^2 1-\alpha/2 / k$$

Exemple : Méthode Gerber  $S_r = 0,18 \Rightarrow \lim_{S_L} = 0,26$  pour  $k = 10$  et  $\alpha = 0,05$

#### - d) évaluation graphique

L'histogramme de distribution des écarts types  $S_L$  permet d'identifier visuellement les laboratoires dont la valeur se distingue nettement du groupe. Les limites normalisées, situées sur le graphique permettent également de faire le point.

\* L'histogramme de l'ensemble des écarts types, laboratoires et échantillons confondus permettent de visualiser la répétabilité globale de la méthode utilisée par le groupe (information d'ordre général pour l'analyste)

### Le critère majeur : la justesse

La justesse d'un laboratoire est le critère majeur intéressant à la fois les laboratoires et les organismes d'accréditation. Elle est mesurée par l'écart systématique d'un laboratoire aux valeurs "vraies" correspondant aux échantillons envoyés.

#### 1) VALEUR "VRAIE"

Elle est généralement la valeur de plus grande probabilité mise en évidence par le groupe, mais peut également être déterminée par un sous-groupe de laboratoires de compétence connue et vérifiée par ailleurs, ou de compétence ponctuelle démontrée par la qualité des résultats obtenus à partir de produits parfaitement connus (matériels de référence, solutions aux concentrations connues par pesée).

La valeur de plus grande probabilité (mode) correspond, pour une distribution gaussienne, à la moyenne et à la médiane. L'élimination des laboratoires ou groupes de laboratoires nuisant à la symétrie de la distribution théorique est nécessaire pour estimer correctement cette valeur. Les éliminations peuvent être visuelles pour l'asymétrie, d'après des histogrammes de distribution des résultats, ou statistiques pour les taux "anormaux". On utilise le test de Grubbs.

#### 2) TEST DE GRUBBS (CF NORME FIL 135b, P7)

Il mesure la réduction de la variance des résultats pour chacun des échantillons, qui résulte du retrait -dans l'ordre- :

- du résultat le plus faible, puis du plus fort ( $G_{HL}$ )
- des 2 résultats les plus faibles, puis des 2 plus forts ( $G_{LL}$ )

$G_{HH}$ )

- en même temps, du résultat le plus faible et du plus fort ( $G_{HL}$ ).

$$G_{HL} = (1 - S_{(-1)}/S_0) \times 100$$

$$G_{HH,HL,LL} = (1 - S_{(-2)}/S_0) \times 100$$

La réduction de variance la plus élevée, supérieure à un seuil critique, fixé dans une table statistique et lié au nombre de résultats, donne lieu à l'élimination des échantillons en cause. Le processus est ensuite renouvelé sur le nouveau jeu de résultats. Le tout s'arrête naturellement lorsque la population redevient homogène (groupée).

Par sécurité, on fixe un pourcentage maximal d'élimination (20 à 50% selon le nombre de participants).

### Apprécier les performances de justesse

#### 1) APPRECIATION QUALITATIVE.

Un premier tableau présente les moyennes des doubles des laboratoires, accompagnées des 10 moyennes générales relatives aux 10 échantillons envoyés avec les écarts-types correspondants. Le test de Grubbs a été effectué et les résultats identifiés comme "anormaux" (risque de 5% d'erreur) sont marqués à l'aide d'un astérisque.

#### 2) APPRECIATION QUANTITATIVE

Pour chaque échantillon et chaque laboratoire est calculé l'écart à la valeur de référence ("vraie") correspondante, d'où 10 écarts par laboratoire et le calcul d'une moyenne, d'un écart-type des écarts et d'un test de Student de la moyenne par rapport à l'écart optimal, nul. Ces valeurs sont reportées dans un second tableau. Aucune élimination de résultat anormal n'est effectuée et, de ce fait, les "anormaux" pointés par le test de Grubbs se signalent par des écarts importants. Ceux-ci vont induire un effet plus ou moins important sur la moyenne et l'écart-type (augmentation) et éloigner le laboratoire de l'objectif idéal (laboratoire parfait !) de moyenne 0 et d'écart-type 0.

Pour chaque méthode et chaque produit, on fixe des tolérances sur la moyenne et l'écart-type des écarts. Elles sont indicatives et correspondent à des objectifs de qualité, souhaitables au plan technico-économique. Des objectifs moins ambitieux peuvent être envisagés pour des utilisations de moindre implication économique.

#### a) classement

L'ensemble des résultats : moyennes et écarts-types  $\bar{d}$  et  $s_d$  est repris dans un tableau de classement des laboratoires. Ce classement est réalisé par l'intégration -à pondération égale- de ces 2 paramètres estimés en un seul critère synthétique :

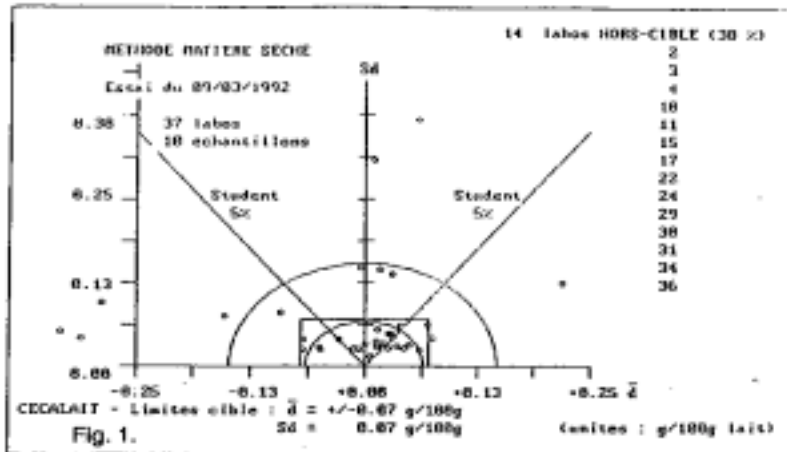
$$R = \sqrt{\bar{d}^2 + s_d^2}$$

R représente la distance qui sépare le laboratoire du laboratoire idéal de  $\bar{d} = 0$  et  $s_d = 0$ .



**b) Visualisation**

Dans un système d'axes orthonormés (cf Fig 1 ci-dessous), où  $\bar{d}$  est placé en abscisse et  $s_d$  en ordonnée, R est la distance entre l'origine des axes et le point de coordonnées  $(\bar{d}, s_d)$  correspondant à un laboratoire. Un rectangle délimité par les limites de tolérance pour  $\bar{d}$  et  $s_d$  représente la cible "objectif" ou cible de conformité (conformité à relativiser en fonction de l'usage fait avec les résultats de routine du laboratoire). Deux demi-droites symétriques par rapport à l'axe des ordonnées délimitent les seuils significatifs pour les moyennes en fonction de la valeur des écarts-types (significatifs en-dessous, non significatifs au-



dessus). On peut délimiter par des demi-cercles des zones limites pour des pourcentages donnés de populations de laboratoires, généralement 50% et 75%. Les rayons de ces demi-cercles peuvent être retrouvés sur le tableau de classement à l'aide des colonnes "%", rang en % des laboratoires et R (cf Fig. 2 ci-dessous). Par ailleurs, un histogramme de distribution des valeurs de  $\bar{d}$  permet de confirmer les laboratoires hors du groupe et d'identifier d'éventuels sous-groupes de laboratoires, cause de non-normalité de la distribution.

Labo	R	%	$\bar{d}$	$s_d$	R	%
1	2	3.3	+0.00	0.02	0.02	3.3
2	5	7.1	+0.01	0.03	0.03	7.1
3	8	11.4	+0.01	0.03	0.03	11.4
4	11	15.7	+0.01	0.03	0.03	15.7
5	14	19.9	+0.00	0.03	0.03	19.9
6	18	25.0	+0.02	0.03	0.03	25.0
7	19	26.4	+0.01	0.04	0.04	26.4
8	22	31.4	+0.02	0.03	0.03	31.4
9	24	33.3	+0.02	0.03	0.03	33.3
10	27	37.9	+0.03	0.04	0.04	37.9
11	30	42.9	+0.04	0.04	0.04	42.9
12	32	45.7	+0.03	0.04	0.04	45.7
13	35	50.0	+0.02	0.05	0.05	50.0
14	38	54.3	+0.03	0.05	0.05	54.3
15	41	58.6	+0.03	0.05	0.05	58.6
16	43	62.1	+0.04	0.05	0.05	62.1
17	44	63.6	+0.03	0.05	0.05	63.6
18	48	68.6	+0.04	0.05	0.05	68.6
19	51	72.9	+0.04	0.05	0.05	72.9
20	54	77.1	+0.07	0.03	0.07	77.1
21	57	81.4	+0.04	0.04	0.07	81.4
22	59	85.7	+0.02	0.04	0.05	85.7
23	62	89.9	+0.00	0.04	0.05	89.9
24	65	94.3	+0.07	0.04	0.09	94.3
25	68	98.6	+0.09	0.08	0.12	98.6
26	70	100.0	+0.05	0.14	0.14	100.0
27	74	100.0	+0.01	0.14	0.14	100.0
28	76	100.0	+0.04	0.12	0.15	100.0
29	78	100.0	+0.15	0.08	0.17	100.0
30	81	100.0	+0.22	0.12	0.20	100.0
31	84	100.0	+0.29	0.10	0.20	100.0
32	86	100.0	+0.01	0.21	0.21	100.0
33	89	100.0	+0.11	0.24	0.24	100.0
34	92	100.0	+0.23	0.05	0.22	100.0
35	95	100.0	+0.06	0.27	0.27	100.0
36	97	100.0	+0.15	0.21	0.24	100.0
37	100	100.0	+0.00	0.47	0.47	100.0

Fig. 2

**3) INFORMATION FOURNIE**

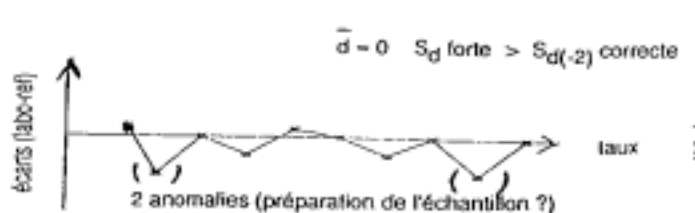
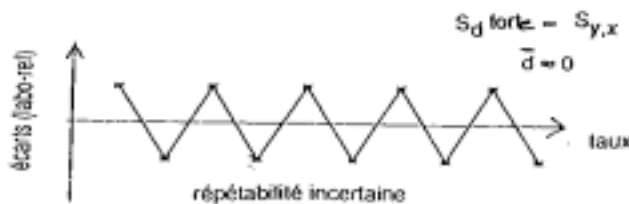
**a) par la moyenne des écarts :**

Elle indique la justesse moyenne du laboratoire, qui peut être constante à tous les niveaux de mesure (écart-type faible) ou variable en fonction des taux mesurés (écart-type des écarts supérieur à ce qui est attendu). Les causes sont à rechercher dans la procédure d'analyse (modifications ou biais par rapport à la procédure normalisée, nouveau produit, étalonnage...)

**b) par l'écart-type des écarts**

Il caractérise la qualité du travail réalisé en intégrant :

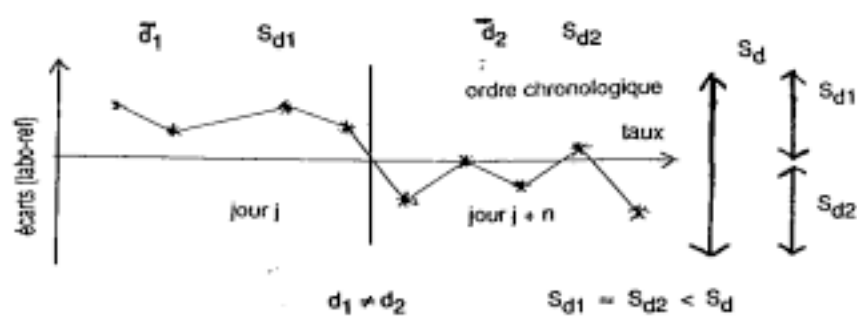
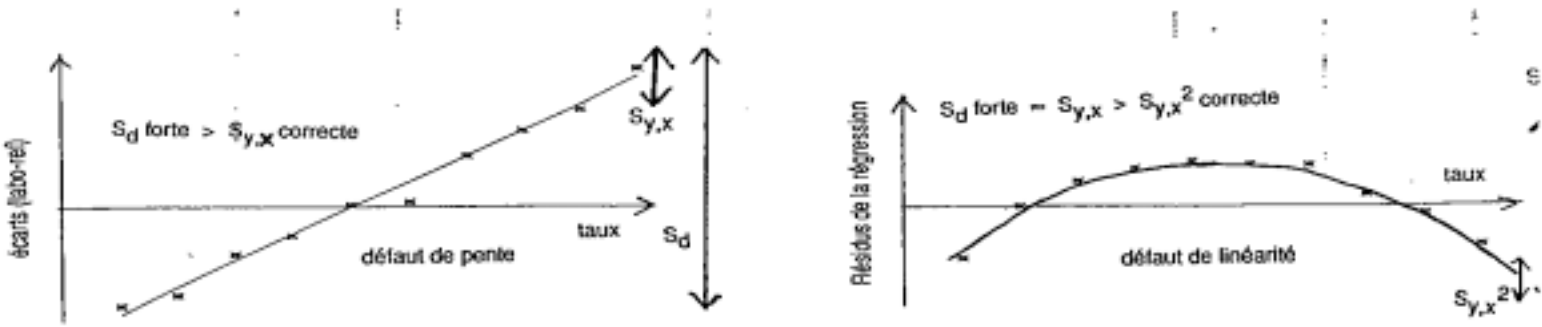
- les erreurs de préparation d'échantillon
- de répétabilité (ou de reproductibilité) si les 10 échantillons sont analysés en plusieurs fois
- d'étalonnage (biais fonction du taux)
- de linéarité du signal.



Le biais d'étalonnage est apprécié par la comparaison de l'écart-type des écarts à l'écart-type résiduel  $S_{y,x}$  de la régression ( $x =$  laboratoire,  $y =$  référence). La première valeur est nettement supérieure à la seconde en cas de biais de pente et équivalente en cas de pente égale à 1.

Le biais de linéarité est apprécié en comparant l'écart-type résiduel de la même régression linéaire avec l'écart-type résiduel de la régression polynomiale du 2e degré, effectuée sur les mêmes données. Le rapport des deux variances doit rester inférieur à 3,73 ( $F_{0,95}$ ) pour 10 résultats. Le manque de linéarité est visible également par l'examen des écarts à la régression linéaire en fonction des taux (ou résidus) dans le tableau de linéarité (uniquement en chaînes Noir Amido) en fonction des taux.

Ci-dessous et ci-contre, sont donnés des exemples des anomalies les plus remarquables qui peuvent ainsi être détectées.



**Conclusion**

Les chaînes d'analyses sont des outils essentiels dans la détection des anomalies analytiques. Le traitement statistique permet de mettre en évidence des problèmes que les laboratoires ne peuvent pas toujours appréhender isolément et d'y

apporter des solutions. Vis à vis des utilisateurs, elles peuvent servir de garantie quant à la fiabilité des résultats. C'est cette utilisation qui en est faite par les organismes d'accréditation.