

REVUE DE STATISTIQUE APPLIQUÉE

M. F. ANCEY

F. BASTENAIRE

R. BOULIN

T. CHPILIOTOFF

A. FEILLOLAY

L. REGNIER

**Enseignements tirés des circuits d'analyses de
comparaison quant à la certification et à l'utilisation
des matériaux de référence**

Revue de statistique appliquée, tome 30, n° 1 (1982), p. 5-23

http://www.numdam.org/item?id=RSA_1982__30_1_5_0

© Société française de statistique, 1982, tous droits réservés.

L'accès aux archives de la revue « *Revue de statistique appliquée* » (<http://www.sfds.asso.fr/publicat/rsa.htm>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques
<http://www.numdam.org/>

ENSEIGNEMENTS TIRES DES CIRCUITS D'ANALYSES DE COMPARAISON QUANT A LA CERTIFICATION ET A L'UTILISATION DES MATERIAUX DE REFERENCE

M.F. ANCEY*, F. BASTENAIRE**, R. BOULIN*
T. CHPILIOFF**, A. FEILLOLAY*, L. REGNIER**

RESUME

Les résultats des circuits d'analyses périodiquement réalisés par un grand nombre de laboratoires européens dans le cadre de l'Association Technique de la Sidérurgie Française ont été soumis à une étude statistique approfondie pour déterminer la répétabilité et la reproductibilité de telles analyses chimiques.

Ces deux caractéristiques mesurées par des écarts-types correspondants ont été étudiées en fonction de la teneur pour l'ensemble des laboratoires participants et pour le sous groupe de ceux qui coopèrent à la certification de Matériaux de Référence. La répétabilité et la reproductibilité ont été ainsi exprimées pour 15 éléments chimiques et pour les deux ensembles de laboratoires.

Les valeurs centrales déduites de l'étude statistique pour ces deux ensembles ont aussi été comparées deux à deux.

Outre les données quantitatives qu'elle fournit quant à la répétabilité et à la reproductibilité des analyses chimiques, cette étude, basée sur près de 50 000 résultats d'analyse, souligne l'importance de l'emploi de méthodes statistiques adaptées au traitement de telles données où les valeurs aberrantes sont fréquentes.

SUMMARY

The results of the round robin analytical tests periodically carried out by a large number of European laboratories under the sponsorship of the Association Technique de la Sidérurgie Française have been thoroughly examined by statistical analysis to assess the repeatability and reproducibility of such chemical analyses.

These two characteristics measured by corresponding standard deviations were studied as a function of element content for the whole population of laboratories participating in these tests and also for the subgroup of those involved in the certification of Reference Materials. Repeatability and reproducibility was thus expressed for 15 chemical elements and the two groups of laboratories.

The central values deduced from the statistical analysis for both groups were also compared.

In addition to giving quantitative data on the repeatability and reproducibility of chemical analyses, this investigation, based on nearly 50 000 individual analytical test results, stresses the importance of using adequate statistical methods in the treatment of data such as these frequently containing outliers.

(*) Département "Analyses chimiques" de l'I.R.S.I.D. (Institut de Recherches de la Sidérurgie Française.

(**) Département "Statistique Appliquée" de l'I.R.S.I.D.

1. INTRODUCTION

Dans le monde technologique complexe dans lequel nous vivons, la production de la plupart des produits exige un grand nombre de matériaux, de produits semi-finis et de pièces. Dans de nombreux cas, la combinaison complexe de tous ces éléments rend difficile l'obtention d'un produit satisfaisant aux normes de qualité exigées par ses utilisateurs. Il est donc nécessaire d'exercer une étroite surveillance sur les éléments d'entrée habituellement par la mesure d'un grand nombre de leurs propriétés.

Bien qu'en théorie, des unités de longueur, de masse et de temps définies de façon précise dussent suffire à l'exécution de toutes les mesures physiques, on a trouvé commode d'utiliser d'autres références et, parmi celles-ci des matériaux de référence (MR).

Beaucoup de caractéristiques des MR telles que leur coût, leur facilité d'emploi etc. ont leur importance mais, de toutes leurs qualités, la principale est l'uniformité de déterminations répétées de leurs propriétés, celles, du moins, qui doivent servir de références. Ceci est nécessaire parce que la justesse et la précision des instruments et des méthodes de mesures s'apprécient en les faisant opérer sur des MR. Il doit donc exister au moins une méthode (et même plusieurs, si possible) dont les résultats présentent, de tous les points de vue possibles, une uniformité satisfaisante.

Dans le présent article, nous examinons la justesse et la précision d'analyses chimiques faites sur des matériaux de référence par des laboratoires chargés de leur certification et par des laboratoires utilisateurs de ces matériaux.

2. EVALUATION DE LA QUALITE EFFECTIVE DE MESURES FAITES SUR DES MATERIAUX DE REFERENCE

Nous nous occuperons ici des deux aspects suivants de cette qualité : la justesse et la fidélité.

Un troisième et important aspect de cette qualité est la stabilité, notamment à long terme, que des données du même type que celles utilisées ici permettent aussi d'apprécier mais cette question fera l'objet d'un autre mémoire.

Nous référant à des définitions reconnues [1] [2], la justesse s'entend ici comme caractérisant "l'étroitesse de l'accord d'une moyenne de mesures avec une valeur de référence reconnue". Dans l'évaluation de cet accord, on accordera cependant toute la considération voulue aux erreurs d'échantillonnage affectant aussi bien la valeur de référence que celle qu'on lui compare.

La fidélité ("degré d'accord mutuel entre des mesures individuelles effectuées dans des conditions prescrites semblables") sera dissociée en deux composantes, l'une de répétabilité, l'autre de reproductibilité. La répétabilité se rapporte à la fidélité de dosages effectués dans le même laboratoire tandis que la reproductibilité se rapporte au degré de dispersion trouvé entre différents laboratoires.

Les circuits d'analyses de comparaison de l'Association Technique de la Sidérurgie Française (ATS) permettent d'estimer la justesse, la répétabilité et la

reproductibilité d'analyses chimiques. Ces circuits sont effectués périodiquement pour vérifier le degré d'accord existant entre un grand nombre de laboratoires français et européens et aussi pour étendre leur compétence à l'analyse de matériaux nouveaux ou inhabituels [3].

Ces circuits d'analyses ont la structure suivante : il est demandé à chaque laboratoire participant de fournir deux résultats pour chacun d'un certain nombre d'éléments précisés par lettre circulaire accompagnant l'échantillon expédié (il n'est cependant pas obligatoire de fournir des résultats pour tous les éléments indiqués, un laboratoire pouvant ne pas être intéressé par le dosage d'un certain élément ou se trouver dans l'incapacité de l'effectuer). La quantité de matériau à analyser fournie à chaque laboratoire est minimale de façon à rendre difficile l'exécution d'un plus grand nombre de dosages qu'il n'est demandé.

Les matériaux distribués pour analyse peuvent être de véritables MR ou de futurs MR mais, pour des raisons économiques, ce sont souvent des sous-produits de la fabrication des MR consistant en des fractions granulométriques plus fines ou plus grossières que celles utilisées pour les MR. Ces fractions granulométriques sont toutefois étroitement bornées et soumises au même processus d'homogénéisation que les véritables MR.

Le nombre de laboratoires participant à ces circuits dépend de la difficulté et de l'intérêt de l'analyse mais il est rarement inférieur à 20 et habituellement de l'ordre de 50 à 60.

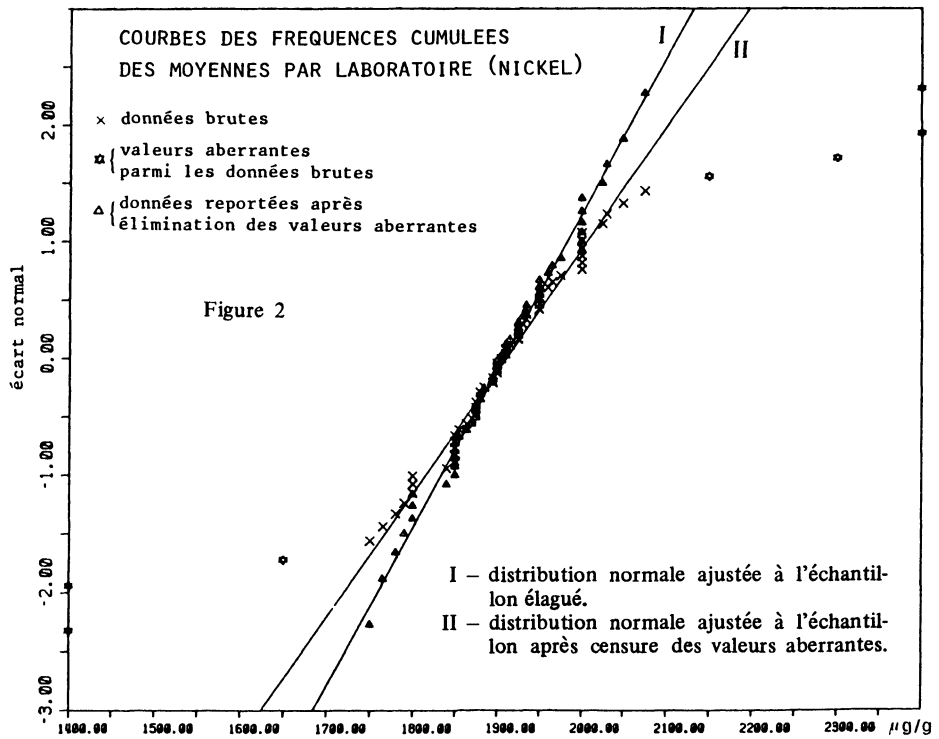
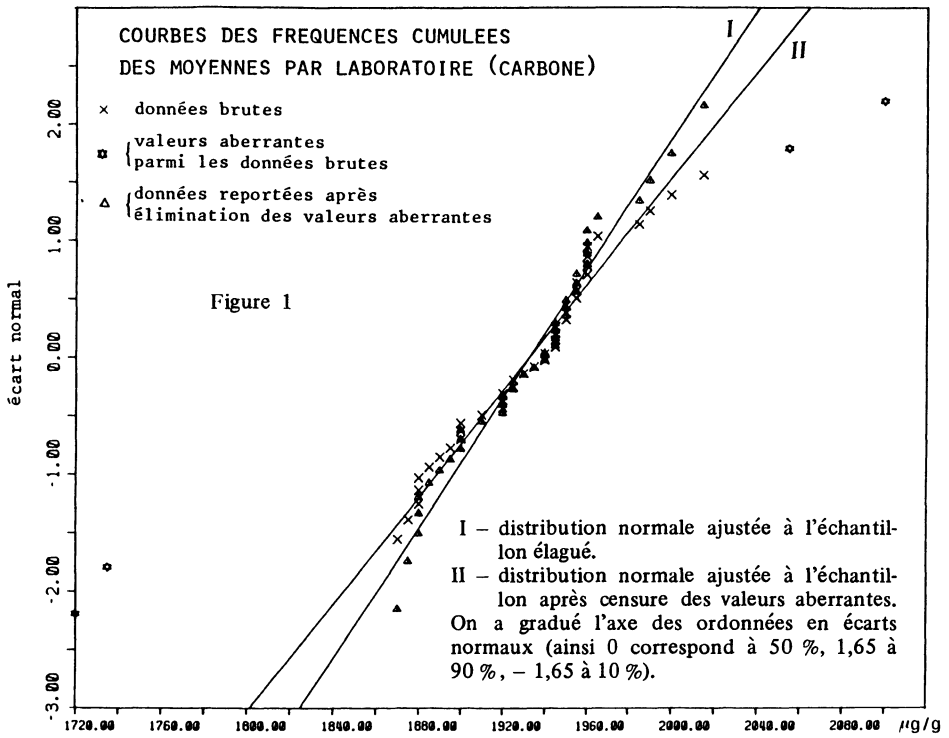
Depuis 1970, les résultats de ces essais ont été constamment soumis à une analyse statistique systématique effectuée sur ordinateur.

Dans un premier stade, les différences entre les deux résultats des paires retournées par les laboratoires sont traitées pour détecter les différences anormalement grandes. Un traceur connecté à l'ordinateur reproduit la courbe des fréquences cumulées de ces différences sur papier gaussien. Dans un second stade, les moyennes des paires sont alors soumises à un traitement similaire.

Dans les deux cas, les valeurs aberrantes sont détectées au moyen du test et des tables de GRUBBS et BECK [4]. Le nombre de valeurs aberrantes présentes dans l'échantillon statistique n'étant pas connu a priori, le test est répété sur les observations restantes de l'échantillon chaque fois qu'une ou plusieurs valeurs aberrantes ont été trouvées. Ceci conduit éventuellement par une succession de pas à un échantillon ne contenant plus de valeurs aberrantes au sens du test appliqué, échantillon que nous appelons "élagué".

Le tableau I donne une description de ce processus tel qu'il est appliqué aux moyennes des paires de résultats ; t est le nombre d'écart-types utilisé pour décider quelles moyennes sont aberrantes dans l'échantillon. Il est indiqué dans les tables de GRUBBS et BECK en fonction de l'effectif de l'échantillon. La ligne suivante du tableau indique le nombre de moyennes soumises au test à chaque pas du processus. La troisième et la quatrième lignes donnent respectivement les valeurs successives de la moyenne et de l'écart-type. L'égalité des valeurs figurant dans les quatrième et cinquième colonnes du tableau indique qu'aucune valeur aberrante de plus n'a été trouvée à la cinquième étape.

Le test étant utilisé au seuil de signification $1 - \alpha$, si chaque échantillon de moyennes de paires obtenu à partir d'un circuit d'analyses était issu d'une population de moyennes ne contenant pas de valeurs aberrantes alors, une proportion α de ces échantillons, c'est-à-dire des circuits d'analyses soumis au test, donnerait lieu, à la première étape, à un résultat significatif. On peut voir que



cette proportion reste égale à α lorsque l'on réitère le test, comme nous le pratiquons, au cas où l'on a détecté au moins une valeur aberrante en notant que ce processus itératif ne se déclenche pas si aucune valeur aberrante n'est détectée à la première exécution du test, ce qui, dans les conditions posées, aurait lieu dans une proportion $(1 - \alpha)$ des cas.

Nous utilisons, en pratique, $\alpha = 0.10$. L'expérience a montré qu'en fait, pour la famille des éléments chimiques soumis à analyse, quelques 70 à 100 % (selon l'élément) des groupes de données étaient trouvés significatifs à ce test.

La preuve de la présence de valeurs aberrantes peut être donnée à l'aide des diagrammes de fréquences cumulées sur papier gaussien établies directement pendant le traitement sur ordinateur. La figure 1 pour les moyennes de paires de déterminations du carbone et la figure 2 pour des données similaires relatives au nickel montrent comment les valeurs aberrantes ressortent dans ces diagrammes.

La moyenne et l'écart-type de l'échantillon "élagué" (qui apparaissent aux colonnes 4 et 5 du tableau I pour l'exemple se rapportant au carbone, fig. 1) définissent une distribution normale représentée par la droite I sur les figures.

Les points représentatifs, reportés à de nouvelles positions pour tenir compte des changements intervenus dans les valeurs des fréquences cumulées après suppression d'aberrants, sont raisonnablement en accord avec cette droite.

On pressent, cependant, que les valeurs aberrantes pourraient aussi être considérées comme telles au motif qu'elles ne s'alignent pas dans le prolongement de la partie centrale approximativement linéaire de la courbe de fréquences cumulées. De ce point de vue, l'écart-type devrait plutôt être caractérisé par la pente de la droite numérotée II.

Un estimateur de l'écart-type, selon ce concept, peut être obtenu au moyen d'une expression linéaire des valeurs ordonnées de l'échantillon x_1, x_2, \dots, x_n munies de coefficients appropriés a_1, a_2, \dots, a_n , c'est-à-dire une expression de la forme $a_1 x_1 + a_2 x_2 + \dots + a_n x_n$, les coefficients a_i attachés à des rangs symétriques par rapport à la médiane y étant de signes opposés et, en général, égaux en valeur absolue ou très voisins (estimateur linéaire de l'écart-type).

Un tel estimateur permet de ne tenir aucun compte de certaines données, notamment les valeurs extrêmes, en leur affectant des coefficients nuls, sans biaiser pour autant l'estimation obtenue. Autrement dit, cette méthode permet d'estimer simplement à partir d'un échantillon "censuré" (dans lequel les proportions de valeurs rejetées restent connues) l'écart-type de la population.

Pour les mêmes raisons, un estimateur linéaire basé sur les mêmes principes peut être utilisé pour estimer la médiane (cette fois avec des coefficients tous > 0).

Pour réaliser les estimations précédentes (et en particulier choisir les coefficients) on s'est servi des méthodes indiquées par GUPTA [5] et SARHAN et GREENBERG [6].

Un autre avantage des estimations ainsi obtenues est de pouvoir en calculer les précisions qui ne dépendent aucunement des valeurs des observations aberrantes rejetées. Au contraire, les propriétés statistiques de biais et de précision de la moyenne arithmétique et de l'écart-type de l'échantillon "élagué" tels que calculés au tableau I ne sont pas connues. Une possibilité, que nous utilisons dans la suite mais dont il faut interpréter les conséquences avec réserve est de considérer simplement l'échantillon élagué comme issu d'une distribution normale.

TABLEAU 1

Recherche des valeurs aberrantes dans les moyennes (de paires) d'un circuit d'analyses de comparaison (carbone)

Indice de boucle	1	2	3	4	5
t	2,719	2,710	2,692	2,682	2,682
n	44	43	41	40	40
Moyennes*	1913,0	1934,0	1935,7	1932,7	1932,7
Ecart-types	155,5	56,0	40,1	35,8	35,8

* Teneur en $\mu\text{g/g}$.

Pour chaque circuit d'analyses et chaque élément, nous avons donc déterminé les caractéristiques suivantes :

- la moyenne "brute", c'est-à-dire la moyenne arithmétique de toutes les valeurs y compris celles jugées aberrantes,
- la moyenne de l'échantillon "élagué",
- l'écart-type de l'échantillon élagué,
- l'estimation linéaire de la moyenne,
- l'estimation linéaire de l'écart-type.

De plus, pour se conformer à une pratique depuis longtemps établie, on a calculé la médiane de tous les résultats individuels.

3. RELATIONS ENTRE LES DISPERSIONS ET LA TENEUR MOYENNE SE DEGAGEANT DES CIRCUITS D'ANALYSES DE COMPARAISON

Quelques soixante-dix circuits d'analyses de comparaison ont été réalisés et dépouillés comme décrit précédemment. Chacun comportait le dosage de plusieurs éléments mais non toujours les mêmes de sorte que le nombre des circuits dans lesquels un élément *donné* a été dosé est toujours inférieur à 70. Ce nombre figure au tableau 5 sous la ligne des symboles chimiques d'éléments.

Pour chaque élément, les données permettent d'étudier les relations entre, d'une part, la répétabilité ou la reproductibilité et d'autre part, la teneur moyenne.

Le tableau 2 donne un exemple des données ainsi rassemblées grâce aux circuits d'analyses. Pour des teneurs en chrome dans des aciers s'étendant de 152 $\mu\text{g/g}$ à 24,18 %, ce tableau donne les écarts-types des différences entre résultats appariés, ceux des moyennes de paires, les estimations linéaires des moyennes de circuit et les estimations linéaires des écart-types des moyennes de paires.

Le tableau 2 montre que ces divers écart-types sont évidemment fonction de la teneur moyenne en l'élément.

Dans l'espoir de caractériser la relation entre l'écart-type et la teneur, diverses fonctions mathématiques ont été ajustées aux données.

Il y a quelque temps, les écarts-types des différences s_d et ceux des moyennes s_m ont été représentés par des fonctions de type $A + Bx$ ou $A + Bx + Cx^2$, x

TABLEAU 2

Données de répétabilité et de reproductibilité pour le chrome

Teneur moyenne	Nbre de lab.	Nbre de paires restantes	Ecart-type des différences $\mu\text{g/g}$	Ecart-type des moyennes $\mu\text{g/g}$	Estimation linéaire de la moyenne $\mu\text{g/g}$	Estimation linéaire de l'écart-type $\mu\text{g/g}$
152	48	42	11.5	29.8	152.2	34.4
190	48	45	11.6	32.6	192.3	33.8
193	48	45	9.0	27.4	194.5	30.8
260	59	52	21.7	22.9	259.8	27.0
420	67	59	20.0	28.6	418.0	36.0
426	74	67	26.3	19.0	427.5	23.0
535	64	55	23.4	40.9	533.7	50.9
567	74	70	24.8	27.2	567.3	30.6
693	65	58	31.0	44.1	692.9	53.5
920	73	70	32.0	39.7	921.8	43.6
1290	74	70	46.3	47.6	1286.8	54.2
1293	64	58	43.3	81.0	1288.7	97.5
2096	74	67	41.6	77.9	2096.6	85.8
2629	60	57	86.8	184.5	2636.2	206.0
6599	71	68	129.3	185.2	6616.9	202.4
10696	74	68	163.2	323.0	10662.8	365.2
11747	84	80	176.5	206.1	11761.2	223.2
14507	61	53	161.8	183.2	14495.1	215.7
26597	53	46	174.9	342.4	26553.5	450.5
40179	78	69	345.7	916.2	40094.6	1049.9
47991	72	65	482.8	745.7	47958.2	888.2
48114	58	52	422.8	716.5	48084.3	805.5
50944	50	44	543.0	726.3	51040.4	871.8
88241	68	60	391.7	547.1	88205.9	699.9
103280	62	55	594.4	647.1	103182.6	773.2
103323	55	49	556.8	652.2	103221.6	783.6
103479	76	71	563.1	992.9	103470.0	1071.8
155613	55	51	760.0	1446.3	155541.8	1615.6
173982	58	50	804.0	819.4	173966.0	998.3
174338	53	51	1322.6	2166.8	174148.4	2330.8
174694	54	51	1140.6	2177.1	174623.1	2432.8
184663	47	44	1047.8	1290.1	184476.4	1460.2
241848	58	54	1337.6	2300.5	241971.0	2510.1

étant la teneur en l'élément. L'estimation linéaire de l'écart-type des moyennes, s_1 , a aussi été représenté de cette manière.

En fait, peu d'idées sur la forme mathématique de la relation entre l'écart-type et la teneur x sont justifiables a priori et l'empirisme paraît être de mise dans cette recherche.

Cependant, le domaine de variation de la teneur étant souvent très étendu en valeur relative (de quelques $\mu\text{g/g}$ à plus de $10^5 \mu\text{g/g}$ pour certains éléments) ainsi que, corrélativement, celui de l'écart-type (de quelques $\mu\text{g/g}$ à plus de $10^3 \mu\text{g/g}$), il est souhaitable, en vue de l'utilisation pratique de ces relations, que ce soit en valeur relative que, quelle que soit la teneur considérée, l'écart-type estimé ne diffère pas trop d'une valeur effectivement observée.

Un moyen commode d'imposer ce critère d'ajustement est d'utiliser pour variable dépendante le logarithme de l'écart-type. C'est pourquoi, dernièrement, nous avons étudié les ajustements de la forme :

$$\log \sigma = A + Bx$$

$$\log \sigma = A + B \log x$$

appliqués à s_d , s_m et s_1 ;

Toutes ces fonctions ont été ajustées en utilisant une méthode de moindres carrés pondérés car chaque écart-type est estimé avec une précision variable dépendant :

- 1) du nombre des résultats au moyen desquels il a été calculé,
- 2) du degré éventuel de censure de chaque queue de la distribution des résultats,
- 3) de la grandeur de cet écart-type.

Sur la base de ces informations, il est possible de calculer la précision de chaque estimation d'écart-type et de définir ainsi son poids dans la régression, égal à l'inverse de son carré.

Ces ajustements ont été effectués pour les 15 éléments analysés dans un certain nombre de circuits.

Parmi les laboratoires participant à ces circuits, certains appartiennent au réseau EURO-STANDARDS. Leurs résultats ont donné lieu à un tri et ont fait l'objet d'une analyse statistique séparée.

Les résultats présentés dans le tableau 3 sont limités aux ajustements de la forme $A + B \log_{10} x$ aux logarithmes des divers écarts-types parce qu'on a constaté que cette forme de relation était celle qui s'ajustait le mieux aux données. Le tableau 3 est divisé en trois parties :

- le tableau 3-1 pour les écarts-types des différences (s_d),
 3-2 pour ceux des moyennes (s_m) et, finalement,
 3-3 pour les estimations linéaires des écarts-types des moyennes (s_1).

TABLEAU 3-1

Ecart-type s_d des différences
 Tableau des coefficients A et B dans $\log s_d = A + B \log x^*$
 (r : coefficient de corrélation de l'ajustement)

Element	Tous laboratoires			Laboratoire de certification		
	A	B	r	A	B	r
Al	-0,048	0,5225	0,91	0,048	0,4848	0,86
As	-0,484	0,6553	0,93	-0,304	0,5613	0,82
C	-0,034	0,4819	0,94	-0,462	0,5760	0,94
C _i **	-0,336	0,5521	0,95	-0,534	0,5961	0,94
Cr	-0,307	0,6230	0,99	-0,464	0,6358	0,98
Cu	0,032	0,4723	0,89	-0,136	0,4797	0,81
Mn	0,167	0,4703	0,93	-0,121	0,5273	0,93
Mo	-0,246	0,6307	0,97	-0,345	0,6313	0,97
N	-0,278	0,4627	0,88	-0,394	0,5068	0,81
Nb	-0,481	0,7070	0,97	-0,617	0,7423	0,95
Ni	-0,340	0,6372	0,98	-0,414	0,6302	0,98
P	0,292	0,3266	0,81	0,111	0,3614	0,79
S	-0,122	0,4990	0,93	-0,163	0,4837	0,89
Si	0,229	0,4847	0,94	-0,009	0,5298	0,91
Sn	-0,202	0,5681	0,93	0,091	0,3921	0,83
V	-0,382	0,6630	0,98	-0,541	0,6760	0,95

* Relation exprimée entre logarithmes décimaux ; s_d et x en $\mu\text{g/g}$

** Carbone déterminé par des méthodes instrumentales

TABLEAU 3-2

Ecart-type s_m des moyennes
Tableau des coefficients A et B dans $\log s_m = A + B \log x^*$
(r : coefficient de corrélation de l'ajustement)

Element	Tous laboratoires			Laboratoires de certification		
	A	B	r	A	B	r
Al	0,558	0,4057	0,79	0,502	0,3953	0,75
As	-0,295	0,7322	0,96	0,049	0,5490	0,85
C	0,162	0,4787	0,91	-0,369	0,5807	0,93
C _i **	-0,189	0,5437	0,91	-0,453	0,6009	0,91
Cr	-0,129	0,6240	0,98	-0,472	0,6711	0,98
Cu	0,235	0,4547	0,83	-0,034	0,4743	0,78
Mn	0,03	0,5666	0,90	-0,213	0,5929	0,91
Mo	0,543	0,4960	0,88	0,081	0,5782	0,89
N	0,015	0,4941	0,80	-0,048	0,5191	0,88
Nb	-0,445	0,7889	0,92	-0,856	0,9014	0,93
Ni	-0,120	0,6312	0,97	-0,484	0,6698	0,96
P	0,390	0,3997	0,82	0,402	0,3288	0,71
S	-0,097	0,5887	0,91	-0,238	0,5994	0,86
Si	0,606	0,4421	0,89	0,380	0,4764	0,89
Sn	0,496	0,4097	0,64	0,680	0,2757	0,41
V	0,570	0,4564	0,83	0,216	0,5261	0,83

* Relation exprimée entre logarithmes décimaux ; s_m et x en $\mu\text{g/g}$

** Carbone déterminé par des méthodes instrumentales

TABLEAU 3-3

Estimation linéaire de l'écart-type des moyennes
Tableau des coefficients A et B dans $\log s_1 = A + B \log x^*$
(r : coefficient de corrélation de l'ajustement)

Element	Tous laboratoires			Laboratoires de certification		
	A	B	r	A	B	r
Al	0,576	0,4197	0,79	0,503	0,4125	0,77
As	-0,216	0,7197	0,95	-0,028	0,6024	0,86
C	0,219	0,4748	0,91	-0,254	0,5629	0,93
C _i **	-0,125	0,5398	0,91	-0,345	0,5839	0,92
Cr	-0,063	0,6229	0,98	-0,326	0,6502	0,98
Cu	0,301	0,4503	0,81	0,093	0,4618	0,81
Mn	0,084	0,5613	0,92	-0,078	0,5690	0,89
Mo	0,544	0,5156	0,90	0,139	0,5837	0,91
N	0,066	0,4945	0,83	-0,054	0,5366	0,88
Nb	-0,401	0,7986	0,93	-0,845	0,9246	0,95
Ni	-0,042	0,6285	0,98	-0,435	0,6750	0,97
P	0,439	0,3988	0,81	0,371	0,3616	0,77
S	-0,071	0,5970	0,92	-0,164	0,5884	0,86
Si	0,648	0,4445	0,90	0,407	0,4802	0,90
Sn	0,506	0,4331	0,71	0,641	0,3282	0,53
V	0,536	0,4865	0,85	0,370	0,4955	0,81

* Relation exprimée entre logarithmes décimaux ; s_1 et x en $\mu\text{g/g}$

** Carbone déterminé par des méthodes instrumentales.

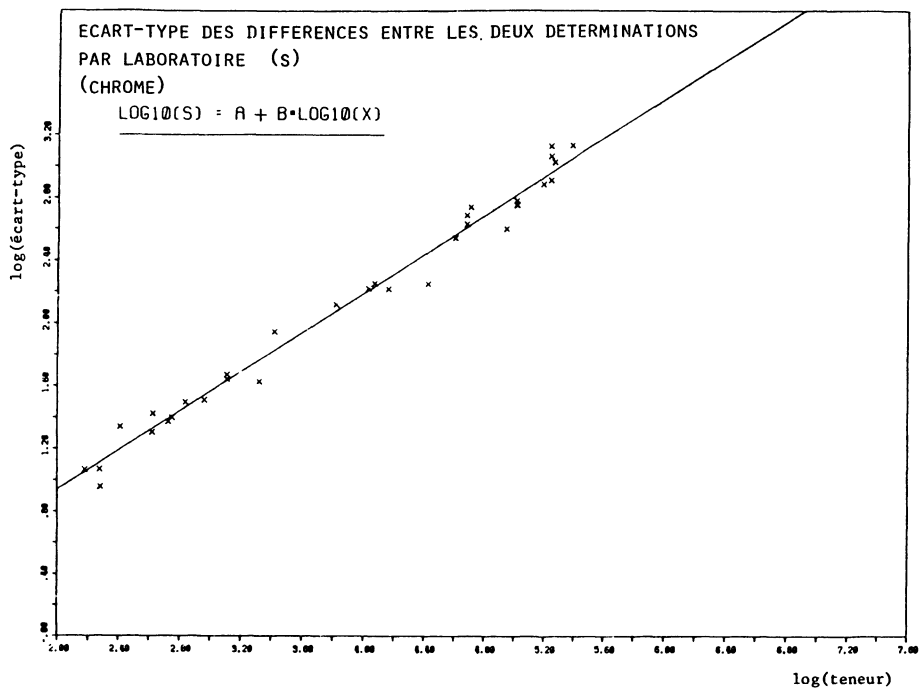


Figure 3

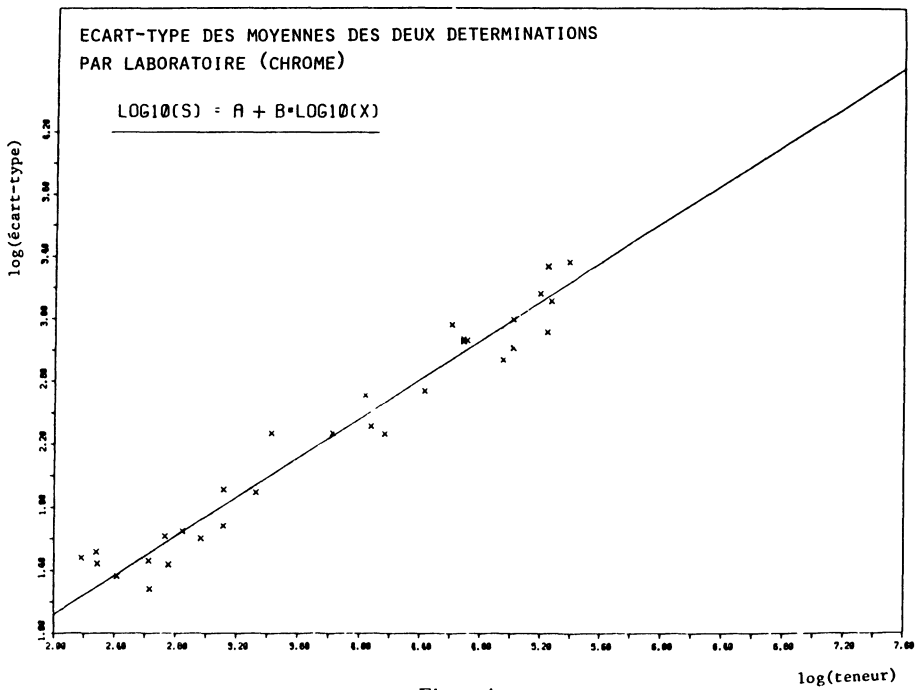


Figure 4

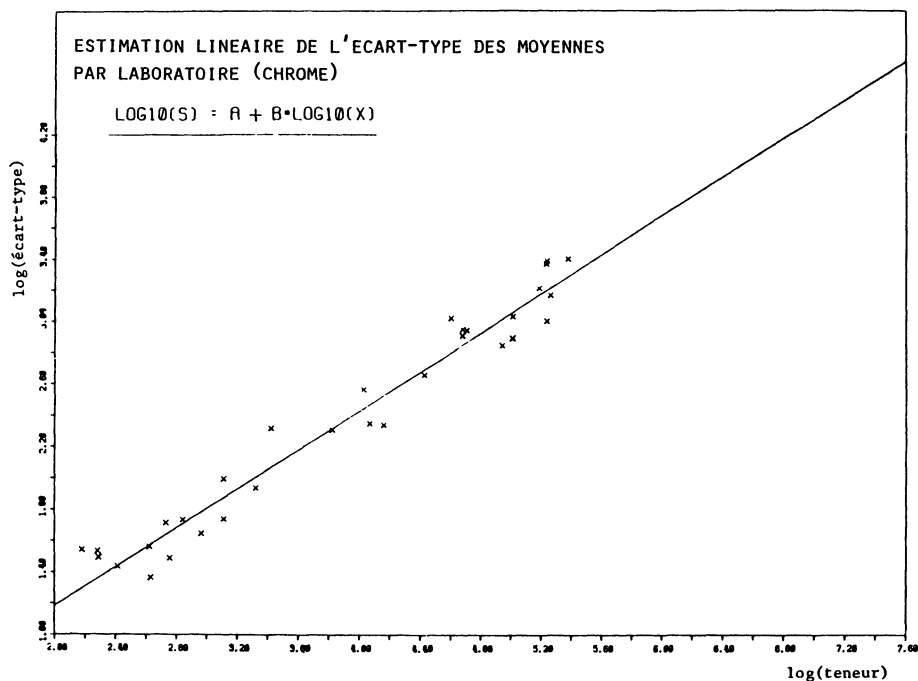


Figure 5

Le tableau donne les coefficients A et B et indique une mesure de la qualité de l'ajustement, à savoir le coefficient de corrélation.

La plupart des coefficients de corrélation sont élevés. Pour un élément donné et un écart-type donné (s_d , s_m ou s_1), B est souvent approximativement égal pour les deux groupes de laboratoires tandis que A est alors plus faible pour ceux qui participent à la certification de MR.

Les relations entre s_d , s_m ou s_1 , d'une part, et la teneur en l'élément, d'autre part, sont illustrées par les figures 3, 4 et 5 relatives au chrome.

4. COMPARAISON DES CARACTERISTIQUES DE TENDANCE CENTRALE DE LA POPULATION DES LABORATOIRES DE CERTIFICATION A CELLES DE LA POPULATION GLOBALE DES LABORATOIRES

Il a été indiqué plus haut que, pour chaque élément et chaque circuit, un certain nombre de caractéristiques de tendance centrale ont été calculées.

Dans une publication antérieure [3], on avait déjà noté que les caractéristiques de tendance centrale déduites des circuits d'analyses de comparaison étaient remarquablement en accord avec les valeurs certifiées lorsque des MR étaient employés dans ces circuits.

Les données de cette étude permettent d'approfondir cette question en examinant, pour chacune des diverses caractéristiques de tendance centrale, la différence entre les valeurs trouvées par l'ensemble des laboratoires et par le groupe des laboratoires de certification.

Il faut noter que, pour ces comparaisons, chaque caractéristique donnée a été obtenue de façon similaire pour les deux groupes de laboratoires — même type de données, même mode d'analyse statistique — et que, par conséquent, les valeurs de la caractéristique en question seront, pour les deux groupes, y compris celui des laboratoires de certification, affectées d'erreurs aléatoires d'amplitudes éventuellement différentes.

Les notations suivantes sont utilisées dans la suite :

- \tilde{M} : médiane des résultats individuels,
- M : moyenne brute,
- M_c : moyenne après "élagage",
- M_1 : estimation linéaire de la moyenne sur échantillon censuré.

Pour tester la signification de ces différences, il a fallu calculer l'écart-type de chaque caractéristique de tendance centrale, pour chaque groupe de laboratoires, sur des bases théoriques.

Comme on l'a déjà noté (cf. fin du § II), l'avantage de M_1 est qu'on peut en calculer la précision à partir de l'estimation linéaire de l'écart-type.

Par contre, la précision de la moyenne de l'échantillon élagué n'est pas connue. On a admis qu'elle se caractérise par une variance égale au carré de l'écart-type de l'échantillon élagué divisé par le nombre des moyennes conservées après élagage.

Il faudrait disposer d'un modèle mathématique de comportement des valeurs aberrantes pour calculer la variance de la moyenne brute mais les hypothèses relatives aux valeurs aberrantes sont toutes sujettes à controverse. On a alors attribué à la moyenne brute une variance s_1^2/n , s_1 étant l'estimation linéaire de l'écart-type et n le nombre initial de moyennes dans le groupe. C'est effectivement une estimation non biaisée de la variance de M lorsqu'il n'y a pas de moyennes aberrantes dans la population ; si c'est le contraire qui est vrai, la variance réelle de M est évidemment plus grande.

On a admis que la médiane de tous les résultats individuels d'un circuit d'analyses pour un élément donné avait même variance que la médiane des moyennes de paires, c'est-à-dire $\pi s_1^2/2n$ parce qu'elle est déterminée par la densité de la distribution de ces moyennes que l'on peut considérer comme égale à celle, en son point de symétrie, d'une distribution normale d'écart-type s_1 . Il faut noter que cette variance n'est pas affectée par les queues de la distribution donc par le nombre et le comportement des valeurs aberrantes.

La variance de la différence entre une caractéristique de tendance centrale donnée des laboratoires de certification et cette même caractéristique pour l'ensemble des laboratoires a été prise égale à la somme de leurs variances respectives. Ces deux termes sont approximativement indépendants, le premier groupe ne représentant qu'une minorité du second. De plus, ce mode de calcul majore la variance et est conservatif.

Deux caractéristiques comparées peuvent avoir ou non la même espérance mathématique. Si leurs espérances mathématiques sont égales, le rapport du carré de la différence de ces caractéristiques à la variance de cette différence — variance

considérée comme connue – est une variable dite de χ^2 avec un degré de liberté. Ceci permet de tester l'égalité des espérances mathématiques des deux caractéristiques (les valeurs critiques du χ^2 sont donnés dans les tables statistiques).

Compte tenu du nombre des circuits d'analyses de comparaison étudiés et, dans chacun d'eux, du nombre d'éléments dosés, on se trouve en présence de près de 500 niveaux de teneur pour 15 éléments en tout. Or, en répétant cinq cent fois sur des données indépendantes l'exécution d'un test statistique au niveau α de signification, il faut s'attendre à commettre 500α erreurs de jugement de première espèce soit, avec $\alpha = 0.05$ par exemple, environ 25 erreurs. Pour éviter ceci, la valeur du χ^2 été fixée de telle façon qu'elle ne soit dépassée par hasard qu'une fois en moyenne en cinq cents épreuves ($\alpha = 0.002$). Il est ainsi peu probable de commettre plus de quelques erreurs de première espèce.

La valeur critique de χ^2_1 pour $\alpha = 0.002$ est 9.55. Pour simplifier, nous l'avons arrondie à 10, les valeurs à traiter étant elles-mêmes souvent arrondies.

Le tableau 4 montre, pour le vanadium choisi à titre d'exemple, les résultats de cette analyse statistique. Les valeurs des quatre caractéristiques de tendance centrale dans 16 circuits d'analyses pour l'ensemble des laboratoires et pour le groupe des laboratoires de certification sont indiquées ainsi que les valeurs de χ^2 pour chaque différence.

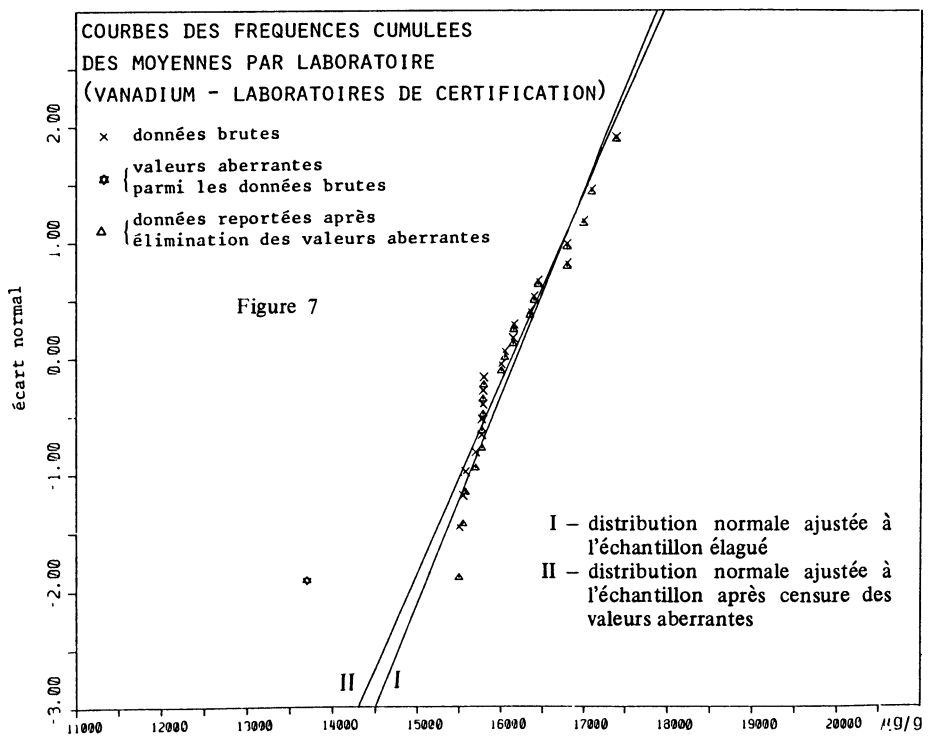
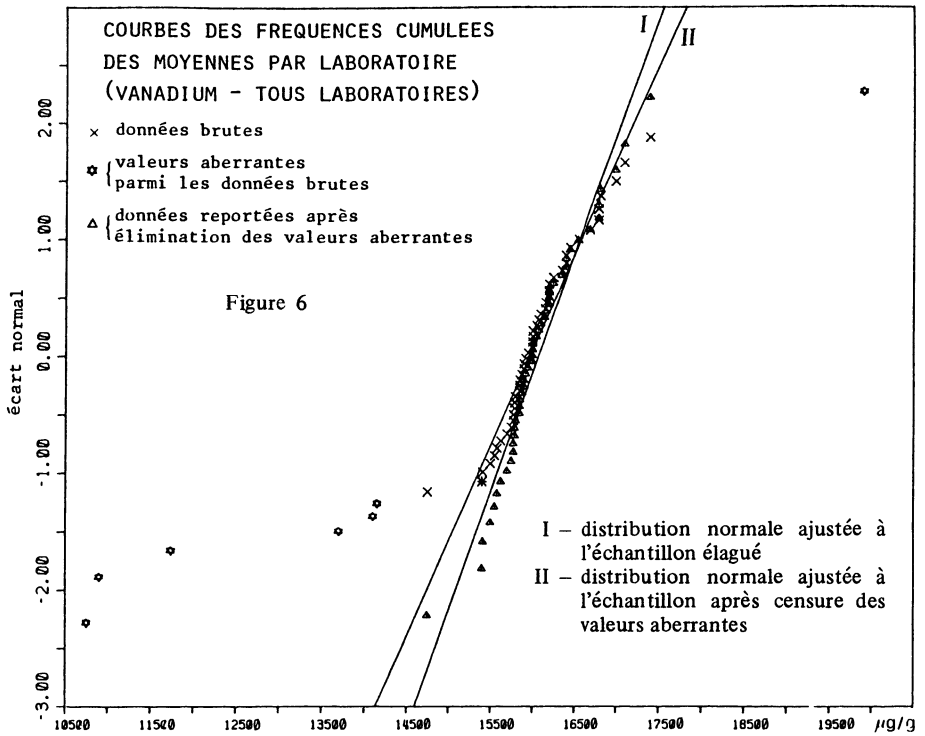
Neuf valeurs significatives (≥ 10) sont indiquées en italiques dans ce tableau. Il est remarquable que huit d'entre elles sont trouvées dans les seules deux premières lignes du tableau. On peut affirmer en toute sécurité que, pour ces deux matériaux, l'ensemble des laboratoires n'était pas en accord avec le groupe des laboratoires de certification. Il n'est pas inutile de noter que, dans ces deux cas, les teneurs en vanadium sont inférieures à 100 $\mu\text{g/g}$.

La seule autre différence significative se trouve à l'avant dernière ligne du tableau, dans la colonne des moyennes brutes.

TABLEAU 4

Comparaison des caractéristiques de tendance centrale pour tous les laboratoires (TL) et pour les laboratoires de certification (LC) dans les circuits d'analyses pour le Vanadium ($\mu\text{g/g}$)

\tilde{M}			M			M_c			M_1		
TL	LC	χ^2	TL	LC	χ^2	TL	LC	χ^2	TL	LC	χ^2
74.	22.	55.0	76.	47.	26.9	71.	18.	131.9	76.5	29.0	70.4
100.	32.	42.4	186.	76.	174.1	99.	59.	39.7	122.5	76.0	34.1
420.	409.	0.5	424.	416.	0.4	421.	411.	0.8	424.3	414.6	0.6
430.	426.	0.1	422.	412.	0.6	430.	427.	0.1	427.9	425.7	0.0
540.	531.	0.3	531.	520.	0.6	540.	539.	0.0	537.9	536.9	0.0
529.	529.	0.0	552.	532.	2.1	548.	531.	1.9	548.4	535.8	0.7
1260.	1259.	0.0	1231.	1219.	0.3	1260.	1263.	0.0	1252.9	1259.3	0.1
1260.	1265.	0.0	1238.	1276.	3.2	1266.	1252.	0.5	1259.5	1267.5	0.1
2720.	2705.	0.1	2745.	2748.	0.0	2707.	2694.	0.2	2722.1	2709.2	0.1
2720.	2710.	0.1	2750.	2751.	0.0	2752.	2711.	1.8	2753.4	2729.2	0.4
2750.	2740.	0.1	2779.	2716.	4.6	2754.	2717.	1.8	2759.8	2716.0	2.0
2895.	2891.	0.0	2839.	2885.	2.0	2892.	2885.	0.1	2888.6	2884.8	0.0
6050.	6040.	0.0	5939.	6018.	3.0	6006.	6037.	0.6	5989.8	6022.8	0.4
6050.	6050.	0.0	6058.	6087.	0.3	6066.	6018.	1.1	6057.8	6047.2	0.0
15900.	16000.	1.0	15754.	16074.	16.2	16088.	16187.	1.8	15973.1	16131.4	3.1
19930.	19900.	0.1	20067.	20068.	0.0	20029.	19876.	3.3	20001.3	19925.9	0.5
Q = 99.60			Q = 234.29			Q = 185.43			Q = 112.60		



Ceci n'a rien de surprenant car c'est dans cette colonne que l'on a le plus de chances de trouver des différences significatives du fait que les valeurs aberrantes ont été conservées dans le calcul des deux valeurs de M . Cette vue est confirmée par les figures 6 et 7 montrant les courbes de fréquences cumulées des moyennes de paires dans ce circuit d'analyses pour l'ensemble des laboratoires (Fig. 6) et pour les laboratoires de certification (Fig. 7). On peut noter que ces deux groupes de données montrent des différences évidentes dans les nombres de valeurs aberrantes tandis que les caractéristiques de tendance centrale autres que la moyenne brute ne donnent pas lieu à différence significative (tableau 4).

Des tableaux semblables au tableau 4 ont été établis pour tous les éléments.

Pour tenter d'y voir plus clair dans les relations entre différences significatives dans les différentes colonnes du tableau 4 et des tableaux similaires relatifs aux éléments autres que le vanadium, on a considéré tous les cas pouvant se produire pour une ligne déterminée de ces tableaux correspondant à un circuit d'analyses donné :

a) Une seule différence est significative. Elle intéresse soit \tilde{M} , soit M , soit M_c , soit M_1 .

b) Deux différences sont significatives dans la même ligne. Elles sont relatives à (\tilde{M}, M) , (\tilde{M}, M_c) , (\tilde{M}, M_1) , (M, M_c) , (M, M_1) ou à (M_c, M_1) .

c) Trois différences sont simultanément significatives : (\tilde{M}, M, M_c) , (\tilde{M}, M, M_1) , (\tilde{M}, M_c, M_1) ou (M, M_c, M_1) .

Les quatre différences sont significatives : (\tilde{M}, M, M_c, M_1) .

En tout, 16 cas sont possibles (2^4). Ils sont énumérés dans les lignes successives du tableau 5, 1^{ère} colonne.

Les fréquences de ces différents cas parmi les 498 circuits x éléments sont récapitulées dans la dernière colonne du tableau 5. Ces fréquences sont décomposées, élément par élément, dans les différentes colonnes (sous chaque symbole d'élément figure le nombre de fois où cet élément a donné lieu à un circuit d'analyses et, sur les lignes suivantes, les nombres de configurations des différences significatives observées pour cet élément).

Le tableau 5 montre que, dans 435 cas sur 498, aucune des quatre différences n'est significative au seuil choisi.

Le cas le plus fréquent après celui-ci et qui se produit 48 fois est la constatation d'une différence significative entre les moyennes brutes M des deux groupes de laboratoires, sans qu'aucune autre caractéristique de tendance centrale ne donne lieu à différence significative. On peut noter, d'ailleurs, que la conjonction d'une différence significative entre les moyennes brutes et entre les valeurs d'une seule autre caractéristique est rare (une fois avec M_c et une fois avec M_1 seulement), les cas où trois ou plus de trois différences sont significatives pouvant être considérés comme indiquant plus probablement un écart réel entre les deux groupes de laboratoires. Les différences entre les moyennes brutes ont donc un comportement indépendant des autres différences, ce qui peut s'expliquer par le caractère erratique des moyennes brutes sous l'influence des valeurs aberrantes.

Une autre manière de mettre en évidence les écarts considérables entre les moyennes brutes est d'additionner les χ^2 relatifs à toutes les différences d'un type donné, élément par élément. On obtient une quantité Q pour chaque type de caractéristique de tendance centrale, indiquée dans les quatre dernières lignes du tableau 5. La somme des valeurs de variables de χ^2 indépendantes étant elle-même une variable de χ^2 dont le nombre de degrés de liberté est égal à la somme des

TABLEAU 5

Différences significative entre les caractéristiques de tendance centrale de tous les laboratoires et les laboratoires de certification

Différences (X) significatives			Element														Totaux de lignes		
			AL	As	N	Cl	C	Cr	Cu	Sn	Mn	Mo	Ni	Nb	P	Si		S	V
\tilde{M}	M	M _c	31	11	20	65	67	33	25	13	40	22	18	12	37	45	43	16	498
			23	11	16	64	63	31	21	11	32	15	13	12	34	41	35	13	435
X																			0
	X		7		4	1	1	2	2	2	8	4	5		3	3	5	1	48
		X															1		1
												1							1
X	X								1										0
X		X																	1
	X	X							1										1
																			0
	X	X																	1
																			0
	X	X																	1
	X	X																	0
																			0
X	X	X										1							5
X	X	X																	2
																			2
		\tilde{M}	20.8	0.5	5.3	28.4	33.6	20.	27.9	2.41	18.1	15.6	3.3	3.7	23.6	19.6	41.7	99.6	362.2
		M	262	6.4	107	139	541	119	939	93.5	221.	242.	161.9	10.6	256.9	143.7	292	234.	3829.
		M _c	79	3.4	7.3	47.9	131	35.2	79.5	7.1	34.5	128.	13.	15.6	38.4	59.7	107.8	185.4	972.80
		M ₁	140	3.2	7.2	33.2	102.	21.	42.1	4.6	33	127	12.5	7.2	28.4	61.5	81.4	112.6	816.90

TABLEAU 6
Nombres de moyennes de paires aberrantes

Elément	⇒	Al	As	N	C	Cr	Cu	Mn	Mo	Ni	Nb	P	Si	Sn	S	V	Tous éléments
Laboratoires de certification	Nombre d'aberrants	39	14	19	76	63	60	53	42	24	21	47	54	25	49	25	611
	Nombre de résultats	643	188	622	1488	793	594	960	484	410	253	865	1139	257	988	378	10062
	%	6.1	7.5	3.0	5.1	7.9	10.1	5.5	8.7	5.8	8.3	5.4	4.7	9.7	5.0	6.6	6.07%
Tous laboratoires	Nombre d'aberrants	94	17	58	147	133	92	171	118	83	27	116	149	32	125	59	1421
	Nombre de résultats	1391	356	813	3135	2079	1236	1139	1236	1139	456	2382	2975	441	2742	995	24372
	%	6.8	4.8	7.1	4.7	6.4	6.0	6.2	9.6	7.3	5.9	4.9	5.0	7.3	4.6	6.4	5.83%

degrès de liberté des variables qui la composent, on peut tester l'hypothèse que les composantes sont bien des variables de χ^2 . Chaque différence systématique apporte en effet une contribution élémentaire plus élevée en moyenne que χ_1^2 et leur cumul dans une somme, si elles sont nombreuses, donne à Q une valeur également plus élevée que la valeur théorique.

On constate que la valeur de Q pour la caractéristique M est, en général, beaucoup plus élevée que pour les trois autres caractéristiques, le seul élément faisant exception étant le niobium. L'accord entre les moyennes brutes est donc beaucoup moins bon qu'entre médianes, moyennes d'échantillons "élagués" ou estimations linéaires des moyennes.

5. DISCUSSION

Par notre pratique nous avons relevé que, selon l'élément dosé, 70 à 100 % des circuits d'analyses révèlent des résultats aberrants. Dans le groupe des laboratoires de certification, ces proportions vont de 65 à 92 %.

Le tableau 6 rassemble ces résultats par élément et par groupe de laboratoires. Ils doivent être interprétés en notant que le caractère aberrant d'un résultat se juge toujours par rapport à l'écart-type et que celui-ci est plus faible pour le groupe des laboratoires de certification. Exprimés en unités physiques (par exemple en $\mu\text{g/g}$), les résultats aberrants de ce groupe le sont moins que pour l'ensemble des laboratoires tout en étant aussi fréquents.

Si l'on met à part les 435 cas dans lesquels aucune caractéristique ne met en évidence de différence significative entre groupes de laboratoires et les 48 cas où une différence significative n'apparaît qu'entre moyennes brutes il reste 15 cas à examiner.

Il est apparu que, dans 9 de ces cas, les différences de caractéristiques qui ne sont pas significatives ($\chi^2 < 10$) sont telles, en fait, que leur χ^2 est assez proche de 10, sont aussi toutes de même signe et approximativement égales. Dans ces différents cas, il semble sûr de conclure qu'un certain décalage existe entre les distributions des moyennes des deux groupes de laboratoires.

Deux autres cas, moins significatifs, semblent pouvoir donner lieu à la même explication.

Enfin, dans les quatre cas restants, les différents critères conduisent à des conclusions contradictoires. Un examen plus approfondi de ces 4 cas montre que les courbes des fréquences cumulées des résultats sont, soit pour l'ensemble des laboratoires, soit pour le groupe des laboratoires de certification, soit pour ces deux groupes à la fois fortement dissymétriques ou de forme encore plus anormale et que, dans ces conditions, il est à peine légitime d'essayer d'estimer une tendance centrale.

6. CONCLUSIONS

Les circuits d'analyses de comparaison apportent de précieuses indications sur la justesse des analyses et permettent d'estimer la répétabilité et la reproductibilité des analyses.

L'étude statistique de leurs résultats révèle l'existence de valeurs aberrantes qu'il convient d'identifier et de traiter ensuite de manière appropriée avant d'estimer des caractéristiques de tendance centrale et de dispersion.

En se basant sur les caractéristiques de répétabilité et de reproductibilité ainsi obtenues, on peut établir d'étroites relations avec la teneur en l'élément. On constate que les laboratoires de certification se caractérisent par des écarts-types de répétabilité et de reproductibilité plus faibles mais qu'en prenant ceux-ci comme référence les valeurs aberrantes sont, pour ces laboratoires, à peu près aussi fréquentes que dans les résultats de l'ensemble des laboratoires lorsqu'on se réfère, cette fois aux dispersions correspondantes observées dans cet ensemble.

La comparaison des médianes, des moyennes brutes, des moyennes après élagage et des estimations linéaires des moyennes des deux groupes de laboratoires indique un accord de toutes ces caractéristiques dans 87,35 % des cas. Dans 9,65 % des cas, seules les moyennes brutes diffèrent entre elles, du fait des perturbations causées par les valeurs aberrantes.

Dans 2 % des cas, environ, les différentes caractéristiques employées concordent entre elles pour indiquer une différence entre les valeurs centrales de l'ensemble des laboratoires et des laboratoires de certification.

Enfin, dans 1 % des cas, les différentes caractéristiques donnent des résultats contradictoires qui paraissent liés à des écarts considérables des distributions des résultats à la normalité.

REFERENCES BIBLIOGRAPHIQUES

- [1] Guide for use of terms in reporting data in analytical chemistry. – *Anal. Chem.*, Vol. 45, n° 14, déc. 1973.
- [2] Standard definitions of terms relating to statistical methods. – *ASTM Standard E 456-72*.
- [3] M.F. ANCEY, F. BASTENAIRE, R. BOULIN, T. CHPILIOFF. – Les circuits d'analyses de comparaison de l'ATS. Aciers et ferro-alliages, applications aux matériaux de référence. *Rapport IRSID RI 711 et circulaire d'informations techniques CIT*.
- [4] F.E. GRUBBS, G. BECK. – Extension of sample sizes and percentage points for significance tests of outlying observations. *Technometrics*, 14, n° 14, 847-52, nov. 1972.
- [5] A.K. GUPTA. – Estimation of the mean and standard deviation of a normal population from a censored sample. *Biometrika*, Vol. 39, 1952, pp. 88-95.
- [6] A.E. SARHAN, B.G. GREENBERG. – Estimation of location and scale parameters by order statistics from singly and doubly censored samples. *The annals of mathematical statistics*, Vol. 29, n° 1, March 1958.