

Recommandations pour des déterminations fiables des incertitudes

Louis-Jean Hollebecq
 Directeur scientifique et technique

Sommaire

1	Introduction	4
2	Symboles et abréviations	5
3	Fondements techniques.....	6
3.1	Fondements techniques de l'ISO 13528 pour l'évaluation des incertitudes.....	6
3.2	Résultats des évaluations réalisées chez CompaLab au cours des dernières années	7
3.3	Bases pour la détermination des incertitudes qui sont communes à la métrologie et aux essais.....	9
3.4	Bases pour la détermination des incertitudes qui sont spécifiques aux essais.....	11
3.5	Classification des méthodes pour déterminer les incertitudes.....	15
3.6	Conclusions relatives aux bases de détermination des incertitudes dans les laboratoires	16
4	Choisir correctement une méthode de détermination de U	16
4.1	Recommandations pour choisir une méthode de détermination de U	16
4.2	Utilisation prévue des incertitudes déterminées	17
4.3	Définition du domaine d'application des incertitudes à déterminer	17
4.4	Informations dont dispose le laboratoire et qui peuvent être réutilisées pour la détermination des incertitudes	17
4.5	Niveau de difficulté pour déterminer correctement les incertitudes en fonction de la méthode d'essai.....	18
4.6	Choisir la méthode pour déterminer les contributions principales aux incertitudes.....	18
4.7	Nécessité d'une expérience partielle complémentaire selon la méthode A du GUM	19
4.8	Conclusions concernant le choix de la méthode de détermination des incertitudes	19
5	Introduction à l'étude des méthodes possibles pour déterminer les incertitudes	20
6	Utilisation des résultats de CIL.....	21
6.1	Introduction.....	21
6.2	CIL spécifiquement prévue pour déterminer les incertitudes.....	23
6.3	Utilisation des résultats d'une CIL qui n'a pas été spécifiquement prévue pour déterminer les incertitudes	24
6.3.1	Introduction.....	24
6.3.2	Calcul des incertitudes individuelles des participants	24
6.3.3	Utilisation de l'ET de reproductibilité comme estimation de l'incertitude	28
6.3.4	Utilisation des ET de répétabilité et de reproductibilité pour estimer l'importance relative du biais et de l'erreur aléatoire.....	28
7	Expérience selon la méthode A du GUM pour laquelle les objets soumis à essais sont des MR.....	29
7.1	Introduction.....	29
7.2	Les valeurs de référence doivent provenir d'une source externe.....	30

7.3	Organisation d'une expérience globale en laboratoire	30
7.4	Utilisation d'un traitement statistique des résultats des essais effectués pour la surveillance de la qualité	32
7.5	Calcul de U	33
7.6	Commentaires concernant B et e	34
7.7	Conclusions concernant les expériences menées selon la méthode A du GUM avec des MR comme éléments testés 34	
8	Expériences selon la méthode A du GUM pour lesquelles les objets soumis à essai ne sont pas des MR	35
8.1	Introduction	35
8.2	Expériences globales ou utilisation des résultats des programmes de surveillance interne de la qualité	35
8.3	Expériences partielles selon la méthode A du GUM [2]	35
8.3.1	Introduction	35
8.3.2	Expérience partielle selon la méthode A du GUM [2] pour déterminer le biais	35
8.3.3	Expérience partielle selon la méthode A du GUM [2] pour déterminer l'impact de la méthode d'essai	36
8.3.4	Expérience partielle selon la méthode A du GUM [2] pour déterminer l'impact du matériau soumis à l'essai	36
8.3.5	Expérience partielle selon la méthode A du GUM [2] pour déterminer l'impact des niveaux des résultats d'essai	36
8.3.6	Expérience partielle selon la méthode A du GUM [2] pour déterminer l'impact de la préparation des échantillons	37
8.3.7	Expérience partielle selon la méthode A du GUM [2] pour déterminer l'impact des conditions de fidélité	37
8.3.8	Mise à jour de U avec les résultats des expériences partielles selon la méthode A du GUM [2]	37
8.3.9	Conclusions concernant les expériences partielles selon la méthode A du GUM [2]	38
9	Études selon la méthode B du GUM	38
9.1	Introduction	38
9.2	Mise en œuvre de la méthode B du GUM	38
9.3	Conclusions sur la méthode B du GUM	41
10	Questions et outils applicables à plusieurs des méthodes décrites	42
10.1	Formats non numériques pour les résultats d'essai	42
10.1.1	Introduction	42
10.1.2	Résultats des tests exprimés en catégories	42
10.1.3	Résultats binaires	44
10.1.4	Résultats exprimés en pourcentages de catégories	45
10.1.5	Résultats exprimés sous forme de courbes	45
10.2	Changement de variables	46
10.3	Questions liées à la résolution et à l'arrondissement des résultats d'essai	47
10.4	Cas où plusieurs valeurs d'incertitudes sont disponibles pour plusieurs situations différentes	48
10.5	Questions générales concernant l'estimation	49
10.6	Questions concernant l'estimation d'une valeur moyenne	50
10.7	Questions relatives à l'estimation d'un écart type	50
10.8	Questions relatives à la combinaison des écarts types	51
10.9	Impact du choix de la classification d'une source d'incertitude comme biais ou erreur aléatoire	53
10.10	Résultats d'essais pour des conditions environnementales données (typiquement, pour des températures données) ou, plus généralement, en fonction de conditions extérieures	54
10.11	La méthode de Monte-Carlo	54
10.12	Analyse de la variance	57
11	Évaluation de la qualité des incertitudes déterminées par le laboratoire	57

11.1	Introduction.....	57
11.2	Utilisation de l'écart-type de reproductibilité.....	57
11.3	Utilisation d'un score ζ adapté.....	58
11.4	Résultats de l'évaluation des incertitudes.....	59
12	Conclusions.....	61
13	Références.....	62

Annexe:

Exemples de mise en œuvre de différentes méthodes de détermination des incertitudes.

Résumé :

Les résultats des CIL (comparaisons interlaboratoires) CompaLab montrent que les incertitudes sont largement sous-estimées par les participants. Une gradation des méthodes d'essai peut être établie, allant de méthodes principalement métrologiques à des méthodes dont les sources d'incertitude sont principalement qualitatives. Les incertitudes sont globalement bien déterminées pour les premières alors qu'elles sont globalement sous-estimées d'un facteur 10 ou plus pour les dernières. Cela provient probablement d'un choix prépondérant de la méthode B du GUM pour les déterminer, quelle que soit la méthode d'essai. Or, si la méthode B est efficace en métrologie, elle ne l'est pas en présence de sources d'incertitude qualitatives importantes. Le GUM manque également de recommandations sur certaines questions spécifiques aux essais. En outre, les résultats des CIL et de la surveillance de la qualité des laboratoires peuvent être réutilisés en méthode A du GUM, qui fournit de bien meilleures estimations des incertitudes et demande beaucoup moins de temps et d'argent que la méthode B. Lorsqu'une détermination précise des incertitudes est importante, des expériences collaboratives selon la méthode A (c'est-à-dire des CIL dédiées) devraient être organisées, dont les résultats peuvent ensuite être utilisés dans des programmes internes de surveillance de la qualité très efficaces. La détermination des incertitudes devrait toujours commencer par une clarification de leur utilisation prévue et une collecte des informations disponibles concernant la précision des essais. La méthode la plus appropriée pour déterminer les incertitudes dépend fortement de ces éléments et, dans la plupart des cas, la réponse n'est pas la méthode B du GUM.

1 Introduction

Depuis 2005 et confirmé en 2017, les laboratoires accrédités selon la norme ISO/CEI 17025 [1] doivent déterminer les incertitudes liées à leurs résultats d'essais. Pour ce faire, ils utilisent principalement les méthodes (en particulier la méthode B) du GUM [2] qui a été publié par le BIPM. Ces méthodes sont basées sur des principes détaillés dans son annexe E et résumés ci-après :

- ✚ Les incertitudes doivent être estimées de la manière la plus réaliste possible, contrairement aux anciennes habitudes qui consistaient à déterminer les incertitudes par excès pour s'assurer que la valeur de référence se trouve à l'intérieur de l'intervalle de confiance ;
- ✚ Aucune des erreurs n'est complètement aléatoire (c'est-à-dire avec une valeur moyenne = 0) ou complètement systématique (c'est-à-dire avec un écart type = 0). Pour cette raison, elles doivent toutes être traitées de la même manière.

Par conséquent, les incertitudes doivent toutes être traitées de la même manière :

- ✚ Les incertitudes doivent toutes être traitées comme des écarts types ;
- ✚ Lorsque le biais est connu, le résultat de l'essai doit en être corrigé ;
- ✚ Lorsque le biais est inconnu, il doit être traité comme s'il s'agissait d'une composante aléatoire de l'incertitude.

Ce document est très précieux car il est basé sur une très longue expérience de la détermination des incertitudes dans le domaine de la métrologie. Par conséquent, il s'agit d'un document très détaillé qui aborde de nombreuses questions difficiles soulevées lorsqu'un calcul d'incertitudes est nécessaire.

Cependant, comme il a été rédigé par des métrologues pour des métrologues, certaines caractéristiques majeures spécifiques aux essais en laboratoire ne sont pas bien traitées dans le GUM, en particulier le biais. D'autre part, la norme ISO 13528 [3] fournit des outils (scores ζ) pour évaluer l'incertitude déclarée par un laboratoire

indépendamment de son biais pendant l'EA. L'exécution de cette évaluation par CompaLab a montré que les laboratoires sous-estiment massivement leurs incertitudes.

Pour résoudre ce problème, le présent document :

- ✚ Fournit les résultats de l'évaluation des incertitudes trouvées lors des CIL de CompaLab au cours des dernières années ;
- ✚ Discute les bases de l'incertitude dans le contexte spécifique des essais en laboratoire ;
- ✚ Explore les différences de conditions entre la métrologie et les essais en laboratoire qui peuvent conduire à des différences dans les méthodes avec lesquelles les incertitudes doivent être déterminées ;
- ✚ Établir une liste des méthodes possibles pour déterminer les incertitudes d'essais et de leurs avantages et inconvénients ;
- ✚ Discuter de certaines questions pratiques susceptibles d'améliorer la qualité des déterminations des incertitudes ;
- ✚ Fournir des propositions pour que les laboratoires évaluent la qualité de la détermination de leurs incertitudes.

Le but de ce document n'est pas de reproduire ou d'expliquer le contenu des documents de référence, en particulier le GUM [2] et le VIM [4], mais de fournir des conseils sur le moment et la manière de les utiliser ou d'utiliser d'autres techniques pour parvenir à une détermination efficace des incertitudes.

2 Symboles et abréviations

Les symboles utilisés dans ce document sont listés dans le Tableau 1.

Tableau 1. Liste des symboles utilisés dans le présent document.

Symbole	Désignation et commentaires
B	Biais d'un résultat d'essai
c	Coefficient de pondération
CoV	Coefficient de variation, défini par $CoV = \mu/\sigma$
e	Erreur aléatoire sur un résultat d'essai
k	Coefficient d'élargissement d'une incertitude
i	Rang d'une entité ou d'une valeur
IC	Intervalle de confiance avec le coefficient d'élargissement k pris égal à 2
m	Estimation d'une valeur moyenne
Med	Valeur vraie d'une médiane
n	Nombre total d'éléments d'une série
N	Nombre total de valeurs aléatoires
r	Coefficient de corrélation
s	Estimation d'un écart-type
s_i	i ^e estimation d'une série ordonnée d'estimations d'écarts-types
s_L	Ecart-type interlaboratoires
s_r	Ecart-type de répétabilité

Symbole	Désignation et commentaires
s_R	Ecart-type de reproductibilité
u	Incertitude-type
$u_{X_{pt}}$	Incertitude-type sur la valeur assignée lors d'une CIL
u_{RM}	Incertitude-type sur la valeur assignée d'un matériau de référence
U	Incertitude élargie, avec coefficient d'élargissement "2"
\bar{x}	Estimation d'une moyenne d'une série de valeurs x
X_{pt} and X_{ILC}	Valeur assignée lors d'une CIL
X_{Lab}	Valeur moyenne des résultats d'un laboratoire
X_{RM}	Valeur assignée pour un matériau de référence
μ	Valeur vraie d'une valeur moyenne
σ	Valeur vraie d'un écart-type
χ^2_{n-1}	Valeur de la loi de distribution du Khi ² avec n-1 degrés de liberté
ζ	Score ζ pour l'évaluation des incertitudes comme décrit dans l'ISO 13528 [3]

Abréviations :

-  CIL : comparaison interlaboratoires ;
-  EA : essais d'aptitude ;
-  ET : écart-type ;
-  IC : intervalle de confiance ;
-  MR : matériau de référence ;
-  MRC : matériau de référence certifié.

3 Fondements techniques

3.1 Fondements techniques de l'ISO 13528 pour l'évaluation des incertitudes

L'ISO 13528 [3] utilise des scores ζ définis par l'Equation (1) pour évaluer les incertitudes déclarées par les laboratoires participant aux programmes d'EA.

$$\zeta = \frac{x_i - X_{pt}}{\sqrt{u_{X_{pt}}^2 + u_i^2}} \quad (1)$$

Où x_i est le résultat du participant i
 X_{pt} est la valeur assignée pour la CIL,
 $u_{X_{pt}}$ est l'incertitude sur la valeur assignée,
 et u_i est l'incertitude déclarée par le participant pour son résultat.

La base de cette Equation (1) est de comparer :

-  L'écart entre le résultat du participant et une valeur considérée comme une valeur de référence ;

✚ Et la combinaison de l'incertitude qu'il revendique sur ses résultats et sur la valeur de référence.

Si l'incertitude revendiquée par le laboratoire est trop faible, l'écart entre ses résultats et la valeur de référence devient incompatible avec cette incertitude revendiquée et le score ζ dépasse les limites qui sont habituellement fixées à 2 et 3, en référence aux limites habituelles utilisées pour les distributions normales.

Il convient de noter que ce score ζ ne peut détecter que des incertitudes manifestement sous-estimées. Lorsque l'incertitude est surestimée, le ζ -score correspondant tend vers 0, mais une situation où le ζ -score tend vers 0 peut aussi signifier que le résultat du participant est par hasard très proche de la valeur de référence. Chez CompaLab, nous avons alors décidé de signaler les situations où les incertitudes sont susceptibles d'être surestimées en les comparant à l'ET de reproductibilité : un avertissement est déclenché lorsque $u_i > 5.s_R$. Ceci est basé sur les bases fournies au § 11. Il faut cependant noter immédiatement que le coefficient 5 est tout à fait conventionnel et qu'il aurait été judicieux de choisir un coefficient plus faible.

3.2 Résultats des évaluations réalisées chez CompaLab au cours des dernières années

Des résultats typiques d'évaluation d'incertitudes sont présentés dans les Figure 1.a à c.

A l'abscisse de ces figures, les participants sont ordonnés par score ζ croissants. L'échelle des ordonnées est graduée en unités de s_R . Par exemple, sur la Figure 1.a, $s_R = 0,0062\%$ masse. Donc l'écart entre l'ordonnée +5 et l'ordonnée -5 est $0,216 - 0,154 = 0,062\%$ correspondant à $10.s_R$. La ligne « 0 » correspond à la valeur X_{pt} et les lignes en pointillés représentent les limites pour $u_{X_{pt}}$ de chaque côté de la ligne X_{pt} .

Les points représentent les résultats des participants : en vert pour ceux dont $\zeta < 2$, en orange pour ceux dont $2 < \zeta < 3$, et en rouge pour ceux dont $\zeta > 3$. En outre, les points jaunes entourés de vert représentent les participants pour lesquels $\zeta < 0,5$ et $u_i > 5.s_R$. L'incertitude déclarée par ces participants est probablement surestimée. Les segments verticaux représentent la valeur u_i de chacun des participants : plus les limites des segments s'éloignent de la ligne X_{pt} , plus le score ζ est élevé.

Il convient de rappeler que, dans les programmes d'EA CompaLab, les participants ne sont pas tenus de fournir des chiffres pour l'incertitude. Il s'ensuit que les chiffres fournis sont ceux que les participants avaient déterminés avant leur participation et ne constituent pas un exercice qu'ils ont effectué uniquement pour satisfaire aux exigences du programme d'EA.

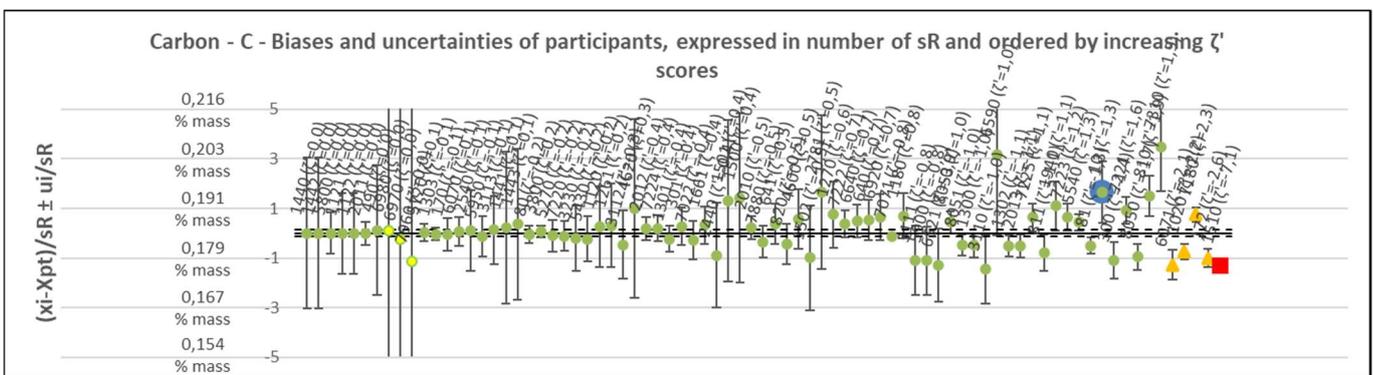


Figure 1.a : Résultats de 2023 d'évaluation des incertitudes pour la teneur en carbone d'un acier faiblement allié.

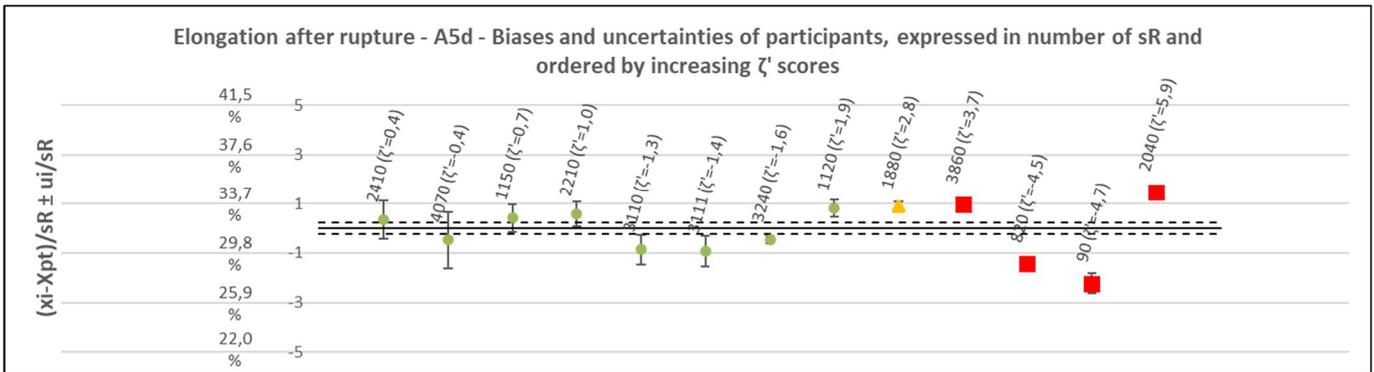


Figure 1.b : Résultats de 2023 d'évaluation des incertitudes pour A_{5d} (allongement à la rupture) pour l'essai de traction sur un acier au carbone.

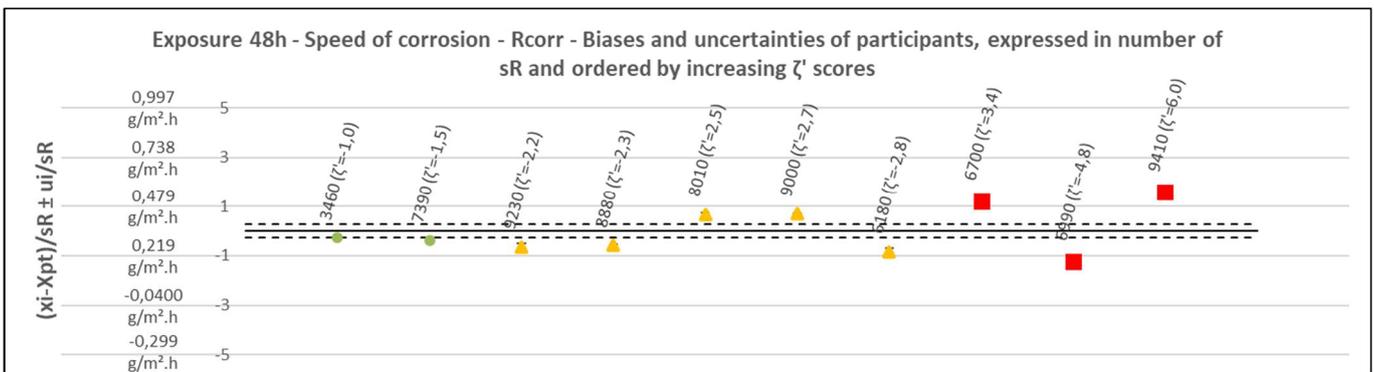


Figure 1.c : Résultats de 2023 d'évaluation des incertitudes pour la Vitesse de corrosion intergranulaire lors d'un essai Huey sur un acier inoxydable.

Les trois exemples de la Figure 1 représentent trois types de situations typiques :

- ✚ Cas 1, illustré par la Figure 1.a : résultats d'essais principalement régis par des questions métrologiques. Outre la teneur chimique des principaux éléments, il peut s'agir de résultats d'essais de dureté ou encore de propriétés mécaniques liées à des forces telles que la résistance à la traction. Dans ce cas, la plupart des longueurs des segments verticaux sont de l'ordre de s_R et quelques participants ne fournissent pas d'incertitudes pertinentes. En outre, les chiffres d'incertitude de mauvaise qualité sont équilibrés entre ceux qui sont sous-estimés et ceux qui sont surestimés ;
- ✚ Cas 2 : illustré par la Figure 1.b : résultats d'essais qui se situent entre le cas 1 et le cas 2, pour lesquels aussi bien les questions métrologiques que technologiques ont de l'importance. Dans ce cas, la plupart des longueurs des segments verticaux sont inférieures à s_R et une proportion anormale de participants a fourni des chiffres d'incertitude sous-estimés ;
- ✚ Cas 3 : illustré par la Figure 1.c : résultats d'essais principalement régis par des questions technologiques. Dans ce cas, la longueur des segments verticaux est toujours inférieure à l'écart entre les X_{pt} et les lignes pointillées, c'est-à-dire que tous les u_i sont inférieurs à $u_{X_{pt}}$ alors que nous avons toutes les raisons de penser le contraire. Presque tous les participants ont reçu une alerte et ceux qui n'en ont pas reçu est probablement le fait du hasard.

3.3 Bases pour la détermination des incertitudes qui sont communes à la métrologie et aux essais

La définition de l'incertitude est commune pour la métrologie et pour les essais. Dans les deux cas, les définitions du VIM [4] s'appliquent, dont certaines (en particulier la répétabilité et la reproductibilité) proviennent de la norme ISO 3534-2 [5] via la norme ISO 5725-1 [6]. L'incertitude est définie comme un paramètre non négatif qui fournit des informations sur la distance maximale susceptible d'exister entre une valeur mesurée et une valeur de référence supposée.

Un modèle d'évaluation de cette distance est fourni par la norme ISO 5725-1 [6] sous la forme de l'équation (2) suivante :

$$y = m + B + e \quad (2)$$

y is the measured quantity value (test result in the case of lab testing),

m is the reference value,

B is the laboratory component of bias in repeatability conditions,

e is the random error occurring in any measurement in repeatability conditions.

La norme ISO 5725-1 [6] introduit également les notions de :

- ✚ L'exactitude, définie comme la proximité de l'accord entre un résultat d'essai et la valeur réelle ;
- ✚ La justesse, définie comme l'étroitesse de l'accord entre un résultat d'essai attendu et la valeur réelle, généralement exprimée sous la forme d'un biais ;
- ✚ La fidélité, définie comme l'étroitesse de l'accord entre des résultats d'essai indépendants obtenus dans des conditions stipulées, généralement exprimée sous forme d'écart-type ;
- ✚ La répétabilité, définie comme la fidélité dans des conditions où des résultats d'essai indépendants ont été obtenus avec la même méthode, sur des éléments d'essai ou de mesure identiques, dans la même installation d'essai ou de mesure, par le même opérateur, en utilisant le même équipement d'essai, dans de courts intervalles de temps ;
- ✚ Reproductibilité, définie comme la fidélité dans des conditions où des résultats d'essai indépendants ont été obtenus avec la même méthode, sur des éléments d'essai ou de mesure identiques, dans des installations d'essai ou de mesure différentes, par des opérateurs différents, à l'aide d'équipements d'essai différents ;

Commentaires concernant *m* :

1. Pour être mise en œuvre, l'équation (2) nécessite la définition de la valeur *m*. Il est désormais largement admis qu'il n'existe jamais de "valeur réelle", au moins pour les résultats d'essais (mais aussi en métrologie). Dans presque tous les cas, un résultat d'essai est un chiffre "dépendant de la méthode".
Par exemple, la résistance à la traction d'un matériau n'a aucun sens si elle n'est pas liée à la méthode utilisée pour la déterminer.
Par conséquent, le résultat de l'essai est conventionnel et il n'existe pas de valeur absolue ou "vraie". Et même si une valeur absolue devait exister, il est pratiquement impossible de la définir avec une précision infinie.
Par exemple, pour déterminer le pourcentage massique de carbone dans une pièce d'acier, il serait théoriquement possible de compter tous les atomes de l'échantillon testé et de distinguer le nombre d'atomes de carbone parmi les autres, le nombre d'atomes qui font partie de la pièce d'acier est susceptible de changer constamment, par exemple sous l'effet de la corrosion. Les parties corrodées appartiennent-elles ou non à la pièce d'acier ? La réponse est conventionnelle, et cette convention influence le résultat de l'essai ;

2. En conséquence, la valeur vraie devrait toujours être définie comme un intervalle plutôt que comme un chiffre unique. De toute façon, même si une valeur vraie existait, elle ne pourrait jamais être déterminée avec exactitude, ce qui augmente la largeur de l'intervalle qui la définit ;
3. Cependant, le traitement de l'équation (2) avec m exprimé sous forme d'intervalle ne serait pas aisé. Sa valeur centrale supposée est alors utilisée et appelée "valeur de référence" plutôt que "valeur vraie" pour rappeler en permanence sa nature conventionnelle.

Commentaires concernant B et e :

1. Pour être mise en œuvre, l'équation (2) nécessite la définition de la valeur B . Cette définition est fournie par la norme ISO 3534 [4] et reproduite dans la norme ISO 5725 [6] et dans le VIM [5]. Il s'ensuit que le biais est lié aux conditions de fidélité dans lesquelles il est déterminé, c'est-à-dire à la méthode d'essai, à l'équipement, à l'opérateur et aux conditions d'essai environnementales. Par conséquent, si, par exemple, nous étudions le biais d'un laboratoire, nous devons considérer un biais moyen qui englobe toutes les méthodes mises en œuvre dans le laboratoire, tous les équipements utilisés dans le laboratoire, tous les opérateurs travaillant dans le laboratoire, toutes les conditions environnementales qui se produisent dans le laboratoire tout au long de l'année. De la même manière, si nous considérons le biais d'un opérateur, nous devons considérer un biais moyen qui englobe toutes les méthodes qu'il met en œuvre dans le laboratoire, tous les équipements qu'il utilise tout au long de l'année, toutes les conditions environnementales qui se produisent tout au long de l'année lorsqu'il est présent ;
2. Pour être mise en œuvre, l'équation (2) nécessite la définition de la valeur e . Il est évident que le terme $y - m$ de l'équation (2) est constant pour une valeur y donnée. Par conséquent, cette valeur e est complémentaire de B et évolue de manière opposée à lui lorsque l'on fait évoluer les conditions de fidélité. e est donc également liée aux conditions de fidélité dans lesquelles elle est déterminée.

Par exemple, lorsque l'effet de l'équipement est considéré comme une moyenne dans la détermination de B , la contrepartie aléatoire doit être incluse dans e , mais lorsqu'un seul équipement d'essai est considéré dans B , aucun effet aléatoire ne se produit en ce qui concerne l'équipement d'essai (le même équipement est toujours utilisé) et aucune contribution de sa contrepartie ne doit être incluse dans e .

Conséquences pour la détermination des incertitudes de mesure :

1. Déterminer correctement une valeur d'incertitude suppose de décider, avant de commencer la détermination, ce que cette incertitude doit couvrir ou, pour reprendre les termes de la norme ISO 5725-1, quelles conditions de fidélité elle doit prendre en compte. Il peut s'agir de l'incertitude sur les résultats d'essais produits par un laboratoire, ou en utilisant une méthode définie, ou par un opérateur spécifique utilisant un équipement d'essai spécifique. Dans le cadre de la norme ISO/CEI 17025, les exigences sont probablement liées à la première proposition (c'est-à-dire l'incertitude sur les résultats d'essai produits par le laboratoire accrédité). Dans la plupart des cas, il suffit de déterminer un chiffre d'incertitude applicable à tous les résultats d'essai produits par le laboratoire, mais nous pourrions imaginer des situations où des chiffres "individuels" liés aux résultats d'essai "individuels" correspondants seraient utiles lorsqu'une incertitude très étroite est requise par les utilisateurs des résultats d'essai ;
2. En fonction de la décision à prendre concernant les conditions de fidélité sur lesquelles portera la détermination de l'incertitude, certaines sources d'incertitude doivent être classées comme sources de biais et d'autres comme sources d'erreurs aléatoires. Par conséquent, elles seront exprimées respectivement en valeurs moyennes ou en écarts-types. Dans la pratique, lorsqu'une seule possibilité se présente pour une source d'écart (par exemple, lorsqu'un seul opérateur est qualifié pour appliquer la méthode d'essai), l'effet connexe sur les résultats de l'essai est manifestement un biais. Au contraire, lorsque plusieurs possibilités se

présentent (par exemple, lorsqu'un laboratoire dispose de nombreuses machines d'essai fabriquées et étalonnées à partir de sources différentes), l'effet sur les résultats des essais doit être considéré comme une erreur aléatoire. Dans la vie réelle des laboratoires, les situations se situent le plus souvent entre les deux. Typiquement, y compris lorsqu'elles sont nombreuses, toutes les machines d'essai d'un laboratoire sont généralement étalonnées par un même organisme d'étalonnage, ce qui induit un risque de biais lié aux normes d'étalonnage utilisées par cet organisme. Par conséquent, dans la plupart des cas, une valeur B et une valeur e doivent être prises en compte pour chaque source d'incertitude.

3.4 Bases pour la détermination des incertitudes qui sont spécifiques aux essais

Plusieurs questions d'importance majeure nous semblent très différentes en pratique pour la détermination des incertitudes en métrologie et dans les essais :

- ✚ Comment traiter les différents niveaux de mesurandes ;
- ✚ Les contributions des facteurs qualitatifs ;
- ✚ L'effet des matériaux ;
- ✚ L'homogénéité interne des échantillons ;
- ✚ La préparation des éprouvettes ;
- ✚ Le traitement du biais ;
- ✚ Le format des résultats d'essai.

En raison de ces différences, certaines dispositions incluses dans le GUM [2] ne sont pas adaptées au cas des incertitudes dans les essais en laboratoire et conduisent à une sous-estimation significative de celles-ci.

Comment gérer les différents niveaux de mesurande

En théorie, il n'y a pas de différence entre la métrologie et les essais à cet égard : dans les deux cas, il est nécessaire de définir l'intervalle pour lequel l'incertitude est déterminée et de déterminer comment l'incertitude varie à l'intérieur de cet intervalle. En conséquence, l'incertitude est exprimée sous la forme d'une constante ou d'un pourcentage de la valeur (valeur mesurée ou résultat de l'essai) ou d'une combinaison des deux, généralement sous la forme $U = a.V + b$.

En pratique, cette opération est généralement facile à réaliser en métrologie (une même opération d'étalonnage couvre habituellement une large gamme de valeurs mesurées, généralement dans un rapport de 1 à 100), mais demande beaucoup de travail, de temps et d'argent pour les essais, en particulier lorsque la méthode comprend de nombreuses sources qualitatives d'incertitude.

Parfois, l'éventail des résultats d'essai attendus est suffisamment étroit pour qu'une seule expérience puisse couvrir correctement la totalité des résultats d'essai produits par le laboratoire.

Exemple : Lorsque le laboratoire effectue des essais de libération d'une gamme limitée de produits fabriqués dans une usine. Dans ce cas, il arrive que l'éventail des résultats d'essai soit suffisamment limité pour permettre la détermination des incertitudes à l'aide d'un seul chiffre.

Lorsque la méthode B du GUM [2] peut être valablement utilisée, elle permet de couvrir facilement toute la gamme des niveaux de mesurandes rencontrés dans la pratique.

Exemple : La masse linéaire LM d'un acier pour béton est déterminée sur un tronçon dont la masse M et la longueur L sont mesurées. LM est alors donnée par l'équation $LM = M/L$. Dans ce cas, en conformité avec les affirmations de ce document, la méthode B du GUM peut être valablement utilisée et permet de couvrir facilement LM dans la gamme de 0,1 à 10 kg/m qui doit être prise en compte pour les aciers pour béton habituels.

Pour certaines méthodes d'essai, des RM ou CRM sont disponibles pour une large gamme de résultats d'essai. Dans ces cas, une expérience complète ou partielle de la méthode A du GUM [2] peut être envisagée pour déterminer l'impact de l'étendue des mesurandes sur les incertitudes.

Exemple : En chimie, la plupart des méthodes d'essai de routine utilisent des MRC, qui sont disponibles dans de nombreuses combinaisons de composition chimique des matériaux. L'utilisation de plusieurs MRC permet de couvrir toute la gamme des résultats d'essais produits par le laboratoire.

Dans la plupart des autres cas, de nombreuses expériences peuvent être nécessaires pour traiter cette question. Cela entraîne des quantités de travail, de temps et d'argent qu'il est économiquement impossible de supporter. Il est alors recommandé de considérer cette question à partir des connaissances techniques concernant la méthode d'essai et/ou de mettre en œuvre une expérience partielle de la méthode A du GUM [2] (voir § 8.3).

Contributions des facteurs qualitatifs :

En théorie, les facteurs qualitatifs qui affectent les résultats ont autant d'importance en métrologie que dans les essais en laboratoire. En métrologie, il s'agit de questions telles que la convection de l'air qui perturbe un étalon de masse ou les variations de l'accélération de la pesanteur. En essais, il peut s'agir par exemple de variations du pH d'une solution corrosive dues à la corrosion de l'objet testé. Théoriquement, les deux sont assez difficiles à gérer. Mais en métrologie, cela affecte généralement le 6^e du 9^e chiffre significatif, tandis que dans les essais en laboratoire, cela affecte généralement le 1^{er} au 3^e chiffre significatif. Par conséquent, dans la pratique, pour la métrologie, la question n'est importante que pour les instituts d'étalonnage de référence alors que, dans la plupart des cas, elle l'est pour tous les laboratoires d'essais en laboratoire. Pour y faire face, il faut disposer de ressources importantes qui peuvent être mises en œuvre dans les instituts d'étalonnage de référence, mais pas dans les laboratoires d'essais ordinaires. Lors des audits d'accréditation des laboratoires, nous avons souvent vu des laboratoires qui avaient bien identifié les principales sources d'incertitudes qualitatives, mais qui les avaient considérées comme mineures simplement parce que cela aurait représenté un travail énorme de les traiter.

Effet du matériau :

Un "effet qualitatif" très courant est l'effet du matériau. Pour de nombreuses méthodes d'essai, la nature du matériau soumis à l'essai influe sur la difficulté d'exécution et, par conséquent, sur la valeur de l'incertitude.

Exemple : Certains métaux sont plus sujets que d'autres au durcissement lors de la déformation à froid. Cet effet de durcissement à froid semble être l'une des sources les plus importantes d'incertitude au cours de l'essai et, par conséquent, les valeurs de U dépendent du type de matériau testé, en plus des autres sources habituelles.

Plusieurs possibilités peuvent être mises en œuvre pour étudier cette question :

- ✚ Utiliser les connaissances techniques concernant la méthode d'essai ;
- ✚ Utiliser les résultats de plusieurs CIL réalisées sur plusieurs niveaux de mesurande ;
- ✚ Inclure l'effet du matériau dans une expérience globale de la méthode A du GUM [2] ;
- ✚ Mettre en œuvre une expérience partielle de la méthode A du GUM [2] (voir § 8.3.4).

En fonction des ressources que le laboratoire est prêt à consacrer à la détermination des incertitudes, plusieurs options sont possibles :

- ✚ Choisir un seul niveau de difficulté lié au matériau, c'est-à-dire le niveau moyen. Cette option fournit des valeurs de U qui sont valables dans la plupart des cas, mais qui sont sous-estimées dans certains cas ;
- ✚ Choisir un seul niveau de difficulté lié au matériel, c'est-à-dire le plus difficile. Cette option fournit des valeurs U qui sont valables dans tous les cas, mais qui sont surestimées dans la plupart des cas ;

- ✚ Choisir de déterminer les incertitudes pour 2 types de difficultés : la moyenne et la plus difficile. Le laboratoire est alors en mesure de déclarer une valeur U plus adaptée au matériau testé ;
- ✚ Lister une série de matériaux typiques testés par le laboratoire et déterminer U pour chacun d'entre eux. Cette option fournit évidemment les meilleures estimations de U mais est également la plus exigeante en termes de ressources à y consacrer.

Homogénéité interne des échantillons :

Un autre "effet qualitatif" très courant est l'effet de l'homogénéité des matériaux. En métrologie, les activités d'étalonnage sont toujours effectuées directement sur des étalons sans qu'il soit nécessaire de les échantillonner ou de les préparer. Au contraire, les essais en laboratoire nécessitent des opérations d'échantillonnage et de préparation des éprouvettes d'essai. Comme la métrologie ne nécessite pas d'activités d'échantillonnage ou de préparation, l'impact de ces activités n'est pas du tout abordé dans le GUM [2] et, par conséquent, de nombreux laboratoires les ignorent ou ne reçoivent pas de conseils appropriés pour traiter cette question. En général, un laboratoire a tendance à considérer que l'homogénéité du matériau ne relève pas de sa responsabilité et prend cela comme une excuse pour éviter de traiter cette question difficile. Mais, en fait, le laboratoire reçoit un échantillon et en sélectionne une partie pour préparer une éprouvette d'essai : il est entièrement responsable de s'assurer que cette éprouvette représente fidèlement l'échantillon qu'il a reçu. Par conséquent, l'impact qui en résulte fait partie de l'incertitude qu'il génère.

Les étalons de calibration et les éprouvettes d'essai ont tous deux un ET d'homogénéité interne qui n'est pas égal à 0. Mais en métrologie, cet ET d'homogénéité interne est inclus dans l'incertitude de l'étalon de référence, alors qu'elle ne l'est pas dans l'échantillon d'essai. Dans de nombreux cas, les méthodes d'essai demandent d'effectuer une mesure quelque part sur l'échantillon d'essai et le résultat trouvé dépend de l'endroit où vous effectuez la mesure sur l'éprouvette.

Par exemple, l'allongement à la force maximale (A_{gt}) spécifié dans les normes ISO 6892-1 et ISO 15630-1 (essais de traction sur les aciers pour béton armé) est mesuré sur une section donnée de l'éprouvette. Les praticiens ont constaté que le résultat de l'essai dépend de la distance entre la rupture et la base de mesure, de la longueur de la base de mesure et de la longueur totale de l'éprouvette, tout ceci étant lié à des phénomènes métallurgiques se produisant à l'intérieur de l'éprouvette.

D'autre part, certaines méthodes d'essai peuvent être utilisées pour étudier les caractéristiques globales de l'échantillon ou les variations de celles-ci au sein des échantillons.

Par exemple, les essais de dureté peuvent être utilisés pour estimer les propriétés mécaniques d'une pièce de métal ainsi que les profils de dureté d'un revêtement durci de pièces mécaniques destinées à résister à l'usure mécanique.

Dans ces cas-là :

- ✚ Lorsque le résultat de l'essai est destiné à représenter la caractéristique globale de l'échantillon, l'effet lié à l'inhomogénéité de l'échantillon doit être inclus dans l'incertitude globale ;
- ✚ Lorsque le résultat de l'essai est destiné à fournir des informations sur les variations des caractéristiques au sein de l'échantillon, l'effet lié à l'inhomogénéité de l'échantillon ne doit pas être inclus dans l'incertitude globale.

De la même manière que précédemment, le GUM [2] ne fournit aucune indication à ce sujet, ce qui peut conduire à une sous-estimation des incertitudes.

Préparation des échantillons :

De même, la plupart des méthodes d'essai exigent une préparation des échantillons qui peut avoir un effet majeur sur les résultats de l'essai et, pour cette raison, est décrite en détail dans les documents de référence qui décrivent la méthode d'essai. Dans la plupart des cas, ces opérations de préparation sont effectuées par le laboratoire ou par un sous-traitant sous sa responsabilité et, par conséquent, doivent être incluses dans l'incertitude qu'il génère.

Par exemple, en chimie, la préparation des échantillons de test doit éviter toute forme de pollution. L'effet de la pollution inévitable dans le laboratoire doit être inclus dans l'incertitude.

Traitement du biais :

Les opérations métrologiques consistent toujours en une comparaison entre la valeur de l'élément contrôlé et une référence. En conséquence :

1. Dans la plupart des cas, le biais peut être évalué (typiquement sous forme de courbes d'étalonnage) et son effet est neutralisé en corrigeant le résultat de l'étalonnage à partir de ce biais estimé ;
2. Même lorsque la composante du biais est inconnue, elle est toujours incluse dans cette comparaison car elle est incluse dans l'incertitude de la référence et dans les résultats des comparaisons.

Au contraire, dans la plupart des cas, aucune référence n'est disponible pour les opérations de test, de sorte que l'erreur systématique, c'est-à-dire le biais, n'est pas du tout incluse dans une expérience visant à déterminer l'incertitude (typiquement la méthode A du GUM [2]). Il s'ensuit que :

1. Dans la plupart des cas, le biais est ignoré même s'il s'agit généralement de la contribution la plus importante à l'incertitude ;
2. Même lorsqu'il est connu (lorsqu'une référence telle qu'un MRC est disponible), il n'est généralement pas possible (sauf en chimie) de tracer quelque chose d'équivalent aux courbes d'étalonnage, et le biais est traité comme une contribution aléatoire au lieu d'une contribution systématique.

Dans les deux cas, l'effet résultant est de sous-estimer l'incertitude (voir § 10.9 concernant la conséquence de la conversion d'un écart systématique en un écart aléatoire).

Format des résultats d'essai :

Les résultats métrologiques sont toujours sous forme numérique, tandis que les résultats des tests peuvent se présenter sous d'autres formes :

- ✚ Une classification en catégories (par exemple le caractère sucré ou non et les arômes du vin), et si ces catégories peuvent être ordonnées ou non (le caractère sucré ou non peut être ordonné de sec à moelleux, alors que les arômes ne peuvent pas l'être) ;
- ✚ Des résultats binaires (par exemple réussite/échec ou présence/absence d'un agent polluant) qui peuvent être considérés comme une classification en deux catégories ;
- ✚ Des pourcentages des catégories (par exemple, classification du graphite dans la fonte par rapport à la norme ISO 945, où les résultats des essais sont fournis en % des catégories définies par les images standard affichées dans la norme) ;
- ✚ Des courbes (typiquement la spectrométrie infrarouge dans laquelle la courbe IR est comparée à une référence pour identifier le produit auquel appartient l'article testé).

Les incertitudes métrologiques étant toujours liées à des valeurs numériques, le GUM [2] ne fournit aucune indication sur ces types de résultats d'essais et les laboratoires d'essais fournissent généralement des informations très insuffisantes sur leurs incertitudes dans ces cas-là. Quelques propositions pour traiter cette question sont présentées au § 10.1 du présent document.

3.5 Classification des méthodes pour déterminer les incertitudes

Conformément à la définition 2.1 du VIM [4], la mesure peut être considérée comme un processus ayant des entrées et des sorties. Les sorties sont évidemment les résultats de la mesure (les résultats des essais dans le cas des essais en laboratoire) et les entrées sont toutes les caractéristiques, conditions et paramètres qui influencent les résultats de la mesure. L'incertitude est liée à l'instabilité du processus de mesure, elle-même liée à l'instabilité des paramètres d'entrée.

Il s'ensuit que nous pouvons distinguer deux manières différentes de déterminer les incertitudes :

1. Évaluation des résultats (c'est-à-dire des résultats de mesure).

Par exemple, le traitement des résultats d'une carte de contrôle utilisant un matériau de référence comme élément de test ;

2. Évaluation de l'impact des variations des paramètres d'entrée sur les sorties du processus (c'est-à-dire les résultats de mesure).

Par exemple, la méthode B du GUM [2].

Manifestement :

- ✚ La plupart des méthodes habituelles ne sont ni tout à fait du type 1 ni tout à fait du type 2 ;
- ✚ Les deux types présentent des avantages et des inconvénients.

Les méthodes de type 1 pur (évaluation des résultats) produisent évidemment de meilleures estimations des incertitudes que celles de type 2, car :

- ✚ Il n'y a pas de risque d'oublier une source d'incertitude inconnue ;
- ✚ Il n'est pas nécessaire d'étudier l'impact des variations des paramètres d'entrée sur les résultats de la mesure, ce qui est plus délicat qu'il n'y paraît à première vue ;
- ✚ Il est clair si le biais a été introduit ou non dans la détermination de l'incertitude.

Les méthodes de type 2 peuvent être mises en œuvre par n'importe quel laboratoire d'essai ou institut d'étalonnage, sans qu'il soit nécessaire de collaborer avec des pairs. Pour cette raison et parce qu'elles sont bien documentées, elles sont les plus utilisées. Cependant :

- ✚ Elles sont généralement plus compliquées à mettre en œuvre
- ✚ Elles demandent souvent une charge de travail beaucoup plus importante ;
- ✚ L'introduction du biais dans la détermination est plus compliquée (prise en compte dans les rapports d'étalonnage mais pas dans les expériences qui consistent en des répétitions au sein du laboratoire).

Par conséquent, elles sont plus susceptibles de conduire à des incertitudes sous-estimées.

Dans les deux cas, toutes les sources d'incertitude des résultats qui relèvent de la responsabilité du laboratoire doivent être prises en compte dans la détermination des incertitudes.

3.6 Conclusions relatives aux bases de détermination des incertitudes dans les laboratoires

L'évaluation des incertitudes réalisée lors des CIL organisées par CompaLab a montré que la plupart des laboratoires participants sous-estiment considérablement leurs incertitudes lorsque la méthode d'essai n'est pas purement "métrologique".

L'utilisation de la méthode B du GUM [2] est largement dominante parmi toutes les méthodes pouvant être utilisées pour déterminer les incertitudes, parce qu'elle est bien documentée (dans le GUM [2]) et parce qu'elle peut être mise en œuvre sans aucune collaboration avec d'autres laboratoires.

Mais l'utilisation de la méthode B du GUM [2] pour déterminer les incertitudes conduit à des incertitudes significativement sous-estimées car elle conduit à ignorer plusieurs des principales sources d'incertitude parce que leur impact est assez difficile à quantifier. En particulier, les sources suivantes sont généralement mal traitées :

- ✚ Biais parce qu'aucune référence externe ne peut être utilisée ;
- ✚ Préparation des échantillons ;
- ✚ Effet du matériau ;
- ✚ Hétérogénéité interne des éprouvettes ;
- ✚ Contributions qualitatives.

On verra plus loin (voir § 10.7 et § 10.8) que les principales sources d'incertitude sont celles qui en font l'estimation et qu'il importe donc au préalable soit :

- ✚ De les identifier et de les déterminer ;
- ✚ Soit d'utiliser une méthode qui intègre leur effet même si elles ne sont pas connues.

La façon la plus simple de mettre en œuvre cette 2ème option est d'utiliser une méthode qui se concentre sur les paramètres de sortie tels que décrits au § 3.5 chaque fois que cela est possible, c'est-à-dire d'éviter la méthode B du GUM [2].

En outre, nous verrons plus loin (voir § 4.6) que certaines méthodes alternatives demandent moins de travail (c'est-à-dire de temps et d'argent) que la méthode B du GUM [2].

4 Choisir correctement une méthode de détermination de U

4.1 Recommandations pour choisir une méthode de détermination de U

Le choix d'une méthode de détermination des incertitudes devrait suivre les étapes suivantes :

1. Déterminer le niveau des ressources qui devraient être investies dans la détermination des incertitudes, en fonction des motivations pour ce faire ;
2. Dresser une liste des informations disponibles pour le laboratoire qui peuvent être réutilisées pour déterminer les incertitudes ;
3. Estimer, en fonction de la méthode d'essais, la difficulté de déterminer correctement les incertitudes pour la méthode d'essai ;
4. Faire un choix pour la méthode de détermination principale parmi les possibilités décrites aux § 6 à 8 ;
5. En fonction des résultats aux questions 1 à 4 ci-dessus, décider si des déterminations complémentaires sont nécessaires ou non ;
6. Déterminer les incertitudes conformément à la méthode choisie (voir § 10) ;

7. Vérifier si les incertitudes déterminées sont cohérentes avec les informations disponibles par ailleurs (voir § 11).

Toutes ces questions sont abordées ci-après, aux § 4.2 à § 4.7.

4.2 Utilisation prévue des incertitudes déterminées

Les motivations possibles d'un laboratoire pour déterminer les incertitudes sont multiples :

- ✚ Satisfaire aux exigences de l'accréditation ;
- ✚ Satisfaire une demande du client du résultat de l'essai (par exemple lorsque le résultat d'essai est destiné à déterminer la conformité d'un produit) ;
- ✚ Aider à analyser les sources d'incertitudes afin d'améliorer la qualité des résultats d'essai fournis aux clients.

Dans le premier et le deuxième cas, une bonne stratégie pour le laboratoire devrait consister à obtenir un résultat acceptable avec le moins de travail, de temps et/ou d'argent possible. Dans le troisième cas, le laboratoire souhaite investir davantage dans le processus, ce qui peut influencer le choix des méthodes à utiliser pour les déterminer.

Le laboratoire peut également suivre différentes stratégies en fonction des situations, c'est-à-dire pour chaque type d'essai, selon qu'un grand nombre d'essais sont réalisés au cours de l'année ou non, selon que les clients sont susceptibles de s'intéresser aux incertitudes des résultats ou non, selon que l'analyse des sources d'incertitude est facile ou non.

Lorsque la détermination des incertitudes est une question critique pour le laboratoire, celui-ci peut utiliser plusieurs méthodes et confronter les résultats à la lumière des conclusions des § 6 à 8.

En fonction des réponses à ces questions, le laboratoire doit décider de son niveau d'investissement dans la détermination des incertitudes : minimum, moyen ou intensif.

4.3 Définition du domaine d'application des incertitudes à déterminer

Avant de commencer la détermination des incertitudes, le laboratoire doit définir le champ d'application des incertitudes à déterminer. En particulier, il convient d'examiner dans quelle mesure les incertitudes doivent ou non inclure les effets des éléments suivants (cf. § 3.4) :

- ✚ Les méthodes d'essai (c'est-à-dire si le laboratoire utilise plusieurs méthodes d'essai) ;
- ✚ L'homogénéité interne du matériau ;
- ✚ La préparation des échantillons.

Les utilisateurs des chiffres d'incertitude devraient être informés de ce qu'ils couvrent et de ce qu'ils ne couvrent pas.

4.4 Informations dont dispose le laboratoire et qui peuvent être réutilisées pour la détermination des incertitudes

Avant d'entamer tout travail, le laboratoire doit s'informer pour savoir si :

- ✚ Il existe des informations disponibles provenant des CIL auxquelles le laboratoire a participé ou de la documentation générale (généralement des normes) ;
- ✚ Des matériaux dont les valeurs centrales proviennent d'une source externe existent ou peuvent être produits ;
- ✚ Les résultats des programmes de surveillance de la qualité sont disponibles dans le laboratoire ;

- ✚ Si ce n'est pas le cas, le laboratoire envisage-t-il de participer à une CIL ou d'organiser un programme de surveillance de la qualité dans un avenir proche.

4.5 Niveau de difficulté pour déterminer correctement les incertitudes en fonction de la méthode d'essai

Le § 3.2 montre que le niveau de difficulté pour déterminer correctement les incertitudes dépend de la méthode d'essai, avec trois catégories principales :

1. Méthodes métrologiques ou quasi-métrologiques ;
2. Méthodes pour lesquelles il existe des sources d'incertitude à la fois quantitatives et qualitatives ;
3. Méthodes pour lesquelles les sources qualitatives d'incertitude sont les plus importantes.

Dans le cas 1, toutes les méthodes, y compris la méthode B du GUM [2], peuvent produire de bons résultats pour la détermination des incertitudes.

Dans le cas 2, la méthode GUM [2] méthode B n'est pas recommandée et, si elle est appliquée :

- ✚ Il faut veiller à ne pas négliger la contribution du biais et des principales sources d'incertitudes, qui sont très souvent celles qui sont qualitatives ;
- ✚ Une vérification approfondie de la validité des déterminations doit être effectuée conformément au § 11.

Dans le cas 3, l'utilisation de la méthode GUM [2] méthode B est une perte de temps et d'argent.

4.6 Choisir la méthode pour déterminer les contributions principales aux incertitudes

A ce stade, le laboratoire est prêt à choisir la méthode qui lui convient le mieux. Dans la plupart des cas, une expérience ou une étude partielle supplémentaire est nécessaire pour compléter la détermination parmi celles décrites aux § 6 à 8. Les critères de choix sont résumés dans le Tableau 2.

Tableau 2. Critères de choix d'une méthode adéquate de détermination des incertitudes.

Méthode de détermination de U	Importance du biais	Difficulté selon la méthode d'essais (cf. § 4.5)	Efficacité	Besoins en ressources	Commentaires
CIL spécifiquement prévue pour déterminer les incertitudes	Non	Toutes	++	++	Non nécessaire en raison de la faible importance du biais
	Oui	1			Utile que si le sujet des incertitudes est important
		2 et 3			Nécessite beaucoup de travail, de temps et d'argent mais très efficace
Calcul à partir des résultats de participation à une CIL	Les deux	Toutes	+	-	Peut nécessiter une expérience partielle selon la méthode A du GUM pour rendre exhaustif le traitement des sources d'incertitude
Utilisation de $u = s_R$	Les deux	Toutes	+	--	Adapté pour les laboratoires courants, dont les performances sont situés dans la moyenne

Méthode de détermination de U	Importance du biais	Difficulté selon la méthode d'essais (cf. § 4.5)	Efficacité	Besoins en ressources	Commentaires
Expérience exhaustive selon la méthode A du GUM par un réseau de laboratoires	Non	Toutes	++	+ à ++ selon le nombre de participants	Non nécessaire en raison de la faible importance du biais
	Oui	Toutes			Très efficace lorsque couplé à la production de matériaux de référence internes
Expérience exhaustive selon la méthode A du GUM en utilisant un MR ou un MRC	Non	Toutes	++	- to ++	Non nécessaire en raison de la faible importance du biais
	Oui	Toutes			Très efficace, et nécessite : <ul style="list-style-type: none"> - Peu de ressources lorsque des résultats d'essais de surveillance interne de la qualité sont disponibles au laboratoire ; - Une quantité moyenne de ressources lorsque la méthode d'essais n'est pas destructive ; - Beaucoup de ressources (coût des MR) lorsque la méthode est destructive.
Expérience exhaustive selon la méthode A du GUM en utilisant des échantillons internes	Non	Toutes	++	++	-
	Oui	Toutes	-		Nécessite une expérience partielle selon la méthode A du GUM pour déterminer le biais
Expérience exhaustive selon la méthode B	Non	Toutes	+	+	Peut nécessiter une expérience partielle selon la méthode A du GUM pour rendre exhaustif le traitement des sources d'incertitude
	Oui	1	+		La présence de sources d'incertitude qualitatives rendent cette option totalement inadaptée
		2	-		
		3	--		

4.7 Nécessité d'une expérience partielle complémentaire selon la méthode A du GUM

Certaines des méthodes énumérées ici assurent (si elles sont correctement mises en œuvre) d'englober toutes les sources d'incertitude possibles (CIL spécifiquement prévue pour déterminer les incertitudes, Expérience exhaustive selon la méthode A du GUM par un réseau de laboratoires, Expérience exhaustive selon la méthode A du GUM en utilisant un MR ou un MRC).

Dans d'autres cas, il convient de vérifier si toutes les sources d'incertitude possibles ont bien été prises en compte dans la détermination (typiquement, biais ou effet aléatoire du matériau, de la méthode, de l'équipement, du personnel, des conditions environnementales). Les cas typiques où expérience partielle selon la méthode A du GUM est nécessaire et la manière dont elle peut être organisée sont décrits au § 8.3.

4.8 Conclusions concernant le choix de la méthode de détermination des incertitudes

On constate que dans de nombreux cas, la méthode B du GUM [2] n'est pas la meilleure méthode pour déterminer les incertitudes. D'autres méthodes existent, qui demandent moins d'efforts pour de meilleurs résultats. Une série de situations typiques est décrite ci-après avec une meilleure proposition pour chacune d'entre elles :

- ✚ Lorsque des informations sur le s_R habituel sont disponibles et que le laboratoire ne souhaite pas investir beaucoup de temps et d'argent dans le processus, il peut adopter s_R comme une estimation valable de son incertitude.
- ✚ Lorsque le laboratoire a déjà participé à une CIL, il peut utiliser les résultats de cette CIL pour déterminer ses incertitudes conformément au § 6.3. Cela ne demande pas beaucoup de ressources de la part du laboratoire.
- ✚ Lorsqu'il dispose d'informations sur le rapport habituel s_i/s_R , le laboratoire peut savoir s'il doit s'efforcer de déterminer le biais ou l'erreur aléatoire, ou si les deux doivent être étudiés (voir § 6.3).
- ✚ Lorsque le biais est important (dans presque tous les cas) et que le laboratoire fait partie d'un réseau (par exemple les branches d'un grand organisme), il devrait envisager de partager des matériaux de référence internes en les testant dans tous les laboratoires (la valeur centrale de tous les laboratoires peut alors être considérée comme une source externe).
- ✚ Lorsque le biais est important (dans presque tous les cas) et que le laboratoire ne fait pas partie d'un réseau, il doit s'enquérir de l'existence de matériaux de référence internes ou de matériaux de référence certifiés, qui sont alors le seul moyen de déterminer le biais.
- ✚ Lorsque le laboratoire exécute un programme de surveillance de la qualité en utilisant des éléments dont la valeur centrale est connue d'une source externe, il devrait les utiliser pour déterminer ses incertitudes conformément au § 7.4. Cela n'exige pas beaucoup de ressources de la part du laboratoire.
- ✚ Lorsque le laboratoire envisage de participer à une CIL ou d'organiser un programme de surveillance de la qualité dans un avenir proche, il devrait envisager de l'organiser de manière à pouvoir utiliser les résultats correspondants pour déterminer ses incertitudes (voir § 7.4).
- ✚ Si les incertitudes sont une question critique pour le laboratoire, il devrait envisager l'organisation d'une CIL dédiée à leur détermination conformément au § 6.2 ou organiser une expérience exhaustive selon la méthode A du GUM en utilisant des éléments dont la valeur centrale est connue à partir d'une source externe.

Les § 5 à § 9 décrivent les méthodes qui peuvent être utilisées pour déterminer les incertitudes, la manière dont elles peuvent être mises en œuvre et leurs avantages respectifs.

5 Introduction à l'étude des méthodes possibles pour déterminer les incertitudes

Nous présentons ci-après une liste de méthodes susceptibles de produire des déterminations acceptables des incertitudes, avec les commentaires correspondants. Cette liste va au-delà des 2 méthodes GUM [2] que l'on voit toujours et partout, mais elle ne peut pas être considérée comme exhaustive. Faire appel à son imagination permet de trouver des idées originales qui peuvent être meilleures que les méthodes classiques, en particulier celles qui tirent parti de certaines situations particulières au laboratoire.

Par exemple, si le laboratoire fait partie d'un centre de recherche qui a une connaissance du produit qui n'est disponible nulle part ailleurs, cette connaissance devrait être utilisée pour déterminer les incertitudes.

En tout état de cause, toutes les méthodes de détermination des incertitudes sont des sortes d'expériences, qui peuvent prendre en compte toutes les sources d'incertitude ou seulement certaines d'entre elles, en fonction de ce qui est techniquement possible et économiquement pertinent pour chaque situation, eu égard aux énoncés du § 4.2. Plusieurs possibilités sont envisagées, comme suit :

- ✚ Utilisation des résultats d'une CIL spécifiquement conçue pour déterminer les incertitudes, qui peut être considérée comme une expérience exhaustive selon la méthode A du GUM par un réseau de laboratoires ;

- ✚ Utilisation des résultats disponibles d'une CIL, qui nécessite de vérifier si toutes les sources d'incertitude ont été prises en compte ;
- ✚ Organisation d'une expérience exhaustive selon la méthode A du GUM sur les matériaux de référence (c'est-à-dire dont la valeur de référence est connue) ;
- ✚ Utilisation des résultats des programmes de surveillance de la qualité intérieure, qui peuvent souvent être considérés comme une expérience exhaustive selon la méthode A du GUM sur des matériaux de référence ;
- ✚ Organisation d'une expérience exhaustive selon la méthode A du GUM sur des matériaux dont la valeur de référence est inconnue ;
- ✚ Étude utilisant la méthode GUM [2] B.

Quelle que soit la méthode, la détermination doit :

- ✚ Prendre en compte toutes les sources d'incertitude (méthode, matériau, homogénéité, niveaux du mesurande, préparation des éprouvettes, équipement, personnel et conditions environnementales) ;
- ✚ Éviter de les compter deux fois ;
- ✚ Éviter toute corrélation entre elles, ou au minima la contrôler, c'est-à-dire déterminer son impact sur le résultat final de la détermination. A noter que, compte-tenu des conclusions du § 10.8, l'impact d'une éventuelle corrélation n'est important que si les deux contributions concernées par la corrélation figurent parmi les plus importantes.

Ceci est particulièrement important et compliqué à assurer pour les études sur la méthode B du GUM [2], mais c'est également important pour les autres méthodes de détermination.

Lorsque plusieurs séries de résultats sont disponibles (*par exemple pour différentes méthodes ou pour différents types d'équipement d'essai ou différents opérateurs*), le laboratoire a deux possibilités :

1. Calculer autant de valeurs de U que de séries de données disponibles. Le laboratoire dispose alors d'informations plus détaillées, mais il doit alors gérer un plus grand nombre de données lorsque, par exemple, un client demande la valeur U associée à ses résultats d'essai ;
2. Consolider les valeurs U obtenues pour chaque série de résultats d'essais en une valeur U générale, valable pour tous les résultats d'essais produits par le laboratoire. Dans ce cas, il convient d'utiliser une valeur moyenne quadratique pondérée de U , comme indiqué au § 10.8.

6 Utilisation des résultats de CIL

6.1 Introduction

Les CIL peuvent être classées en 3 types principaux :

1. Les CIL dont l'objectif est de déterminer la répétabilité et la reproductibilité d'une méthode d'essai. Les textes de référence en la matière sont l'ISO 5725-2 [8] et l'ASTM E691 [9] ;
2. Les CIL dont l'objectif est d'évaluer la compétence des participants. Le texte de référence est l'ISO 13528 [3] ;
3. Les CIL dont l'objectif est de déterminer la valeur de référence d'un matériau de référence certifié (MRC). Le texte de référence est l'ISO 33405 (ex. Guide ISO 35) [10].

Même si leur processus est assez similaire (distribution d'échantillons aussi semblables que possible, réalisation des essais par les participants dans des conditions données, traitement statistique des résultats), les détails de la réalisation peuvent être complètement différents, compte tenu des différents objectifs poursuivis. En particulier :

- Les efforts de l'organisateur ne sont pas focalisés sur les mêmes questions. Dans le cas 1, l'attention est portée sur la détermination de la qualité des ET de répétabilité et de reproductibilité. Dans le cas 2, l'attention est portée sur les scores z . Dans le cas 3, l'attention est portée sur l'exactitude de la valeur centrale de référence. Cela peut entraîner des différences dans l'organisation et dans les résultats des CIL ;

Par exemple, les CIL consacrées à la détermination des caractéristiques des MRC ne produisent pas de s_r et s_R qui représentent la fidélité de la méthode d'essai mise en œuvre par un laboratoire moyen, parce que les participants à ces CIL sont généralement des laboratoires très compétents, dont les incertitudes sont inférieures à la moyenne.

- Les types de participants sont différents. Dans le cas 1, les participants sont des pairs de l'organisateur de la CIL, généralement issus du monde universitaire. Dans le cas 2, les participants sont des clients du fournisseur d'EA. Dans le cas 3, les participants sont des sous-traitants de l'organisateur de la CIL. Le niveau de compétence des participants et le contrôle que l'organisateur exerce sur eux sont différents dans les trois cas ;
- Les types de matériaux sont différents. Dans le cas 1, des MRC peuvent être utilisés pour assurer une meilleure homogénéité. Dans le cas 2, les éléments sont censés représenter la vie quotidienne des laboratoires participants, c'est-à-dire des produits industriels. Dans le cas 3, les échantillons soumis aux essais sont destinés à devenir des MRC qui ne sont pas encore évalués ;
- Les niveaux d'assurance qualité de l'organisateur sont susceptibles d'être différents. Dans le cas 1, aucune surveillance spécifique n'est requise, tandis que les cas 2 et 3 sont soumis à l'accréditation, c'est-à-dire à une surveillance extérieure (volontaire).

Cette liste n'est pas exhaustive, mais elle montre clairement que le niveau de confiance dans les informations provenant des CIL peut varier considérablement, en fonction des sources et des paramètres utilisés pour déterminer les incertitudes.

Même si nous n'avons jamais vu un tel exercice, il ferait sens d'organiser une CIL spécifique à la détermination des incertitudes, dans laquelle toutes les décisions relatives à son organisation (en particulier les options pour les conditions de fidélité telles que définies dans la norme ISO 5725-1 [6]) seraient axées sur cet objectif.

En tout état de cause, la norme ISO 21748 [11] fournit des orientations sur la manière de traiter les résultats des CIL afin de déterminer les valeurs d'incertitude. Cette norme repose sur les deux principes suivants

- L'écart-type de reproductibilité est une base valable pour la détermination de l'incertitude ;
- Les sources de dispersion non incluses dans la CIL doivent être identifiées et ajoutées s'il ne peut pas être démontré qu'elles sont négligeables.

Il identifie certaines sources d'incertitude qui, typiquement, peuvent ne pas être incluses dans la CIL :

- Un biais lié à la méthode d'essai ;
- L'effet du matériau (le matériau est-il différent de ceux habituellement rencontrés au laboratoire) ;
- Le niveau d'homogénéité (doit-il prendre en compte ou pas l'homogénéité interne de l'échantillon, cf. § 3.4) ;
- Le niveau du mesurande, s'il est significativement différent de ceux habituellement rencontrés au laboratoire ;
- La préparation des éprouvettes lorsqu'elles sont préparées par l'organisateur de la CIL.

Cela se produit lorsque l'organisateur de la CIL demande aux participants d'utiliser une méthode donnée ou lorsqu'il prépare lui-même les éprouvettes. Ces conditions dépendent normalement de l'objectif de la CIL et sont plus susceptibles de se produire dans les CIL de type 1 et 3 que dans celles de type 2.

Trois questions seront examinées ci-après :

- Ce que pourrait être une CIL spécifiquement prévue pour déterminer les incertitudes individuelles ;

- ✚ Comment les résultats d'une CIL peuvent-ils être utilisés par un participant pour déterminer ses incertitudes ?
- ✚ Comment les résultats globaux d'une CIL peuvent être utilisés comme valeurs de référence pour les incertitudes.

6.2 CIL spécifiquement prévue pour déterminer les incertitudes

Si une CIL spécifiquement prévue pour déterminer les incertitudes individuelles est organisée, elle doit inclure toutes les sources d'incertitude qui ont une incidence sur les résultats de l'essai, c'est-à-dire :

- ✚ Le biais éventuel lié à la méthode d'essai ;
- ✚ Les biais et erreurs aléatoires possibles liés aux matériaux ou à leur homogénéité, en fonction de l'utilisation prévue pour les résultats d'essais (cf. §3.4) ;
- ✚ Les biais et erreurs aléatoires liés aux niveaux de mesurande ;
- ✚ Les biais et erreurs aléatoires possibles liés à la préparation des spécimens d'essai ;
- ✚ Les biais et erreurs aléatoires liés à la réalisation des essais.

Pour ce faire, le programme de la CIL doit inclure :

- ✚ Un nombre suffisant de résultats d'essais provenant de toutes les méthodes pouvant être utilisées pour déterminer les résultats des essais ;
- ✚ La préparation des spécimens d'essai par les participants, de sorte que le biais et l'erreur aléatoire possibles soient inclus dans l'ET interlaboratoire et dans le biais global du laboratoire ;
- ✚ La réalisation des essais inclut des conditions de précision (cf. § 3.3) qui couvrent toutes les sources d'incertitude rencontrées dans la pratique au sein de chaque laboratoire, c'est-à-dire l'ensemble de ses équipements d'essai, l'ensemble de son personnel et toutes les conditions environnementales qui peuvent être rencontrées au cours de l'année.

La norme ISO 5725-3 [12] (qui définit des schémas alternatifs pour les essais de fidélité) peut être utilisée pour déterminer les plans de CIL adéquats pour atteindre ces objectifs.

En utilisant ces conditions de CIL, le traitement statistique des données peut déterminer :

- ✚ Le biais lié à chaque méthode d'essai. Lorsqu'une méthode d'essai est considérée comme la référence, la valeur de référence est calculée à partir des résultats correspondants. Dans le cas contraire, la valeur de référence doit être considérée comme la moyenne de tous les résultats d'essai ;
- ✚ La valeur réelle du biais lié à chaque laboratoire ;
- ✚ La valeur réelle de l'erreur aléatoire de chaque laboratoire.

A partir de ces résultats, chaque laboratoire peut déterminer la valeur de l'incertitude élargie U liée aux résultats d'essais qu'il produit en utilisant l'équation (3) :

$$U = |X_{Lab} - X_{ILC}| + k \cdot \sqrt{u_{X_{ILC}}^2 + s_{Lab}^2} \quad (3)$$

où U est l'incertitude élargie du résultat de l'essai,
 X_{Lab} est la valeur moyenne des résultats d'essai du laboratoire,
 X_{ILC} est la valeur de référence calculée à partir de l'ensemble des résultats d'essai de l'ILC,
 k est le coefficient d'élargissement,
 $u_{X_{ILC}}$ est l'incertitude sur la valeur X_{ILC} ,
 s est l'écart-type des résultats d'essai du laboratoire.

Un terme B_{met} représentant le biais de la méthode peut être ajouté s'il n'est pas inclus dans X_{Lab} .

Conclusion concernant l'organisation d'une CIL spécifiquement prévue pour déterminer les incertitudes :

Cette possibilité est probablement la méthode la plus puissante pour déterminer les incertitudes, car toutes les sources d'incertitudes sont incluses dans l'expérience. Cependant, elle nécessite toute une organisation et, par conséquent, est coûteuse en temps et en argent.

Cette possibilité peut être adaptée lorsque la détermination de valeurs précises pour les incertitudes est critique.

6.3 Utilisation des résultats d'une CIL qui n'a pas été spécifiquement prévue pour déterminer les incertitudes

6.3.1 Introduction

La norme ISO 21748 [11] fournit de nombreuses informations sur la manière d'utiliser les résultats d'une CIL pour déterminer les incertitudes. Elle indique que l'ET de reproductibilité constitue une bonne base pour l'estimation de l'incertitude type.

Lorsque l'impact de la méthode d'essai ou de la préparation des éprouvettes n'est pas inclus dans la dispersion estimée dans la CIL, il doit être ajouté. Dans les deux cas, cela peut être réalisé par une expérience partielle selon la méthode A du GUM [2], comme expliqué au § 8.3.

6.3.2 Calcul des incertitudes individuelles des participants

Comme le même modèle (c'est-à-dire l'équation (2) de [6]) est utilisé quel que soit le type de CIL, l'équation (3) peut également être utilisée pour calculer l'incertitude sur les résultats d'un participant individuel. Nous avons vu au § 6.2 une configuration idéale de CIL pour laquelle la totalité des sources d'incertitude est incluse dans les calculs. Dans la plupart des cas, les CIL habituelles présentent de légères divergences par rapport à cette situation :

- ✚ La représentativité des valeurs moyennes et des écarts types ;

Par exemple, une CIL organisée pour déterminer les caractéristiques d'un MRC est susceptible de fournir des valeurs s_r et s_R sous-estimées, car les laboratoires qui y ont participé sont probablement plus compétents que les laboratoires moyens ;

- ✚ La méthode d'essai, lorsque plusieurs méthodes d'essai sont possibles et qu'il a été demandé d'utiliser une méthode qui n'est pas la méthode de référence ;
- ✚ La préparation des éprouvettes, lorsqu'une telle préparation est nécessaire et que la CIL a été effectuée sur des éprouvettes préparées par l'organisateur ;
- ✚ Les conditions de fidélités (telles que définies dans [6]). Dans la plupart des cas, il est demandé aux participants d'effectuer les essais dans des conditions de répétabilité. Par conséquent, les impacts de l'équipement d'essai, des opérateurs et des conditions environnementales ne sont pas pris en compte dans l'estimation de l'erreur aléatoire.

Lorsque l'impact de la méthode d'essai ou de la préparation des échantillons n'est pas inclus dans la dispersion estimée dans la CIL, il doit être ajouté. Dans les deux cas, cela peut être réalisé par une expérience partielle selon la méthode A du GUM [2], comme proposé au § 8.3.

Lorsque de tels impacts doivent être ajoutés, l'équation (4) pour le calcul de U s'applique, comme suit :

$$U = |X_{Lab} - X_{ILC} + B_{Met} + B_{Prep} + B_{EPC}| + k \cdot \sqrt{u_{X_{ILC}}^2 + s_{Lab}^2 + s_{Prep}^2 + s_{EPC}^2} \quad (4)$$

où U est l'incertitude élargie du résultat de l'essai,
 X_{Lab} est la valeur moyenne des résultats d'essai du laboratoire,
 X_{ILC} est la valeur de référence calculée à partir de l'ensemble des résultats d'essai de la CIL,
 B_{Met} est le biais lié à la méthode,
 B_{Prep} est le biais lié à la préparation,
 B_{EPC} est le biais lié aux équipements d'essai, aux conditions personnelles et environnementales,
 k est le coefficient d'élargissement,
 $u_{X_{ILC}}$ est l'incertitude sur la valeur X_{ILC} ,
 s_{Lab} est l'écart-type des résultats d'essai du laboratoire,
 s_{Prep} est l'écart-type lié à la préparation,
 s_{EPC} est l'écart type lié aux conditions de fidélité dans lesquelles la CIL a été réalisée.

Lorsque s_{EPC} inclut un ET de répétabilité, s_{Lab} doit alors être omis afin d'éviter un double comptage de cet impact.

Ces calculs peuvent être effectués soit par le fournisseur de la CIL, soit par le participant à partir des informations fournies dans le rapport de l'ILC.

CompaLab calcule lui-même les incertitudes par cette méthode pour ses clients. Les CIL organisées par CompaLab sont du type 2 du § 6.1. Par conséquent, la politique de conception des CIL est la suivante :

1. Faire en sorte que toutes les opérations qui font partie de la compétence des participants soient effectuées par les participants ;
2. Placer les participants dans des conditions d'essais aussi proches que possible de leurs conditions habituelles.

Ensuite, dans le cadre des CIL organisées par CompaLab :

- ✚ Toutes les méthodes permettant de déterminer le résultat attendu sont autorisées. Les participants doivent utiliser leurs méthodes habituelles ;
- ✚ La préparation des éprouvettes doit être effectuée sous leur responsabilité.

Par conséquent, les deux questions principales abordées au § 6.1 sont satisfaites.

En outre, lorsque plusieurs méthodes sont possibles et qu'un nombre suffisant de participants fournissent des résultats d'essai pour une méthode d'essai donnée, un traitement statistique distinct est effectué pour cette méthode, ce qui permet de vérifier si :

- ✚ Cette méthode entraîne un biais significatif ;
- ✚ Tout éventuelle alerte est due à une mauvaise exécution de la méthode, ou à la méthode elle-même ou à une combinaison des deux.

Dans la pratique des CIL CompaLab, les comparaisons de méthodes s'appliquent généralement aux essais chimiques, à la détermination de l'épaisseur des revêtements, aux méthodes de contrôle de la vitesse de chargement dans les essais mécaniques et aux méthodes de détermination de l'allongement à la force maximale A_{gt} sur les aciers pour béton armé. Ces comparaisons sont effectuées à l'aide de la formule classique $|X_{Met} - X_{Overall}| / \sqrt{u_{X_{Met}}^2 + u_{X_{Overall}}^2}$, qui ne doit pas dépasser 2. Dans presque tous les cas, aucune différence significative ne peut être mise en évidence, à l'exception de l' A_{gt} pour lequel les résultats de l'ISO 6892-1 sont inférieurs à ceux de l'ISO 15630-1 dans un rapport moyen significatif de 0,9.

Toutefois, il est demandé aux participants d'effectuer les essais dans des conditions de répétabilité, afin de permettre l'évaluation de leur répétabilité et d'éviter que l'exécution de la CIL ne prenne trop de temps. Pour cette raison, un ET_{EPC} devrait être ajouté dans le calcul de U , mais les informations correspondantes ne sont pas disponibles pour le faire. Par conséquent, ce terme est ignoré et devrait être ajouté par les participants.

Malgré ce problème, les incertitudes calculées de cette manière sont assez bonnes, comme le montrent les Figure 2.a à 2.c.

Ces figures fournissent des comparaisons de u_d déclaré par les participants, u_c calculé par CompaLab et s_R (ET de reproductibilité) pour les CIL qui ont été utilisées dans la Figure 1. Elles sont construites comme suit :

- ✚ Les abscisses sont u_d , incertitudes standard déclarées par les participants ;
- ✚ Les ordonnées sont u_c , calculées avec l'équation (4) à partir des résultats de la CIL. Il convient de noter que l'équation (4) fournit des incertitudes U élargies et non des incertitudes-types u . Pour rendre les comparaisons possibles, nous avons considéré que k est presque toujours choisi égal à 2, et par conséquent, nous avons choisi de calculer $u_c = U_c/2$, c'est-à-dire d'inclure la moitié de la contribution du biais, même si cela n'est évidemment pas correct sur le plan théorique ;
- ✚ En conformité avec les propositions de la norme ISO 21748 [11], s_R est pris comme référence, c'est-à-dire qu'il est représenté par le point [100 %;100 %] ;
- ✚ Les échelles de l'axe sont logarithmiques en %, pour tenir compte des différences considérables qui se produisent généralement entre les différents types de détermination ;
- ✚ La légende est la suivante :

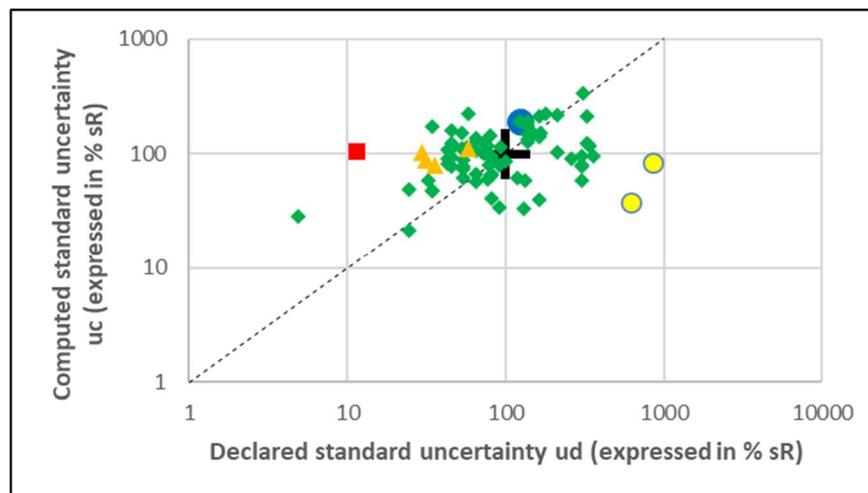
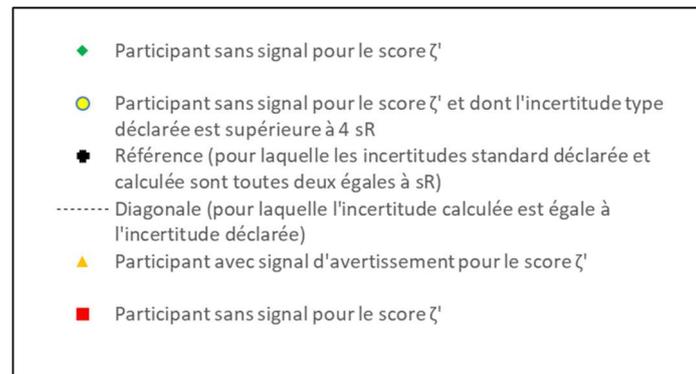


Figure 2.a : Résultats de 2023 pour les évaluations des incertitudes pour la teneur en carbone sur acier faiblement allié.

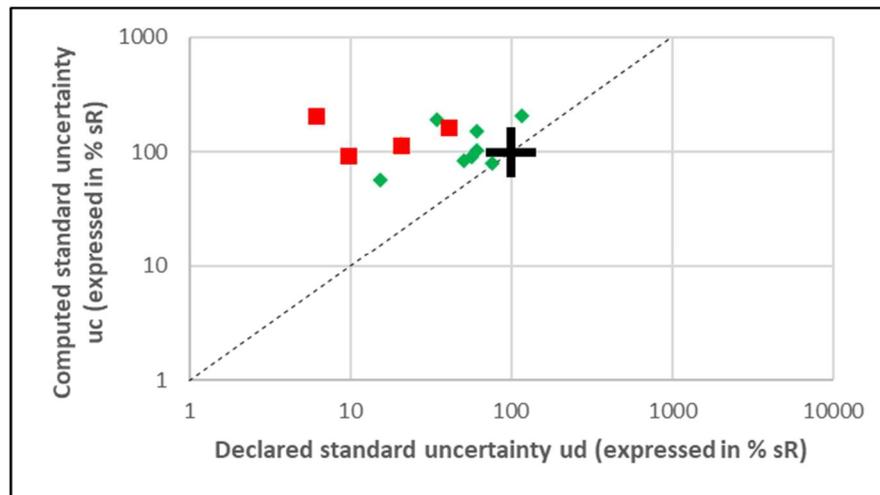


Figure 2.b : Résultats de 2023 pour les évaluations des incertitudes pour A_{sd} (allongement à rupture) pour l'essai de traction sur acier non allié.

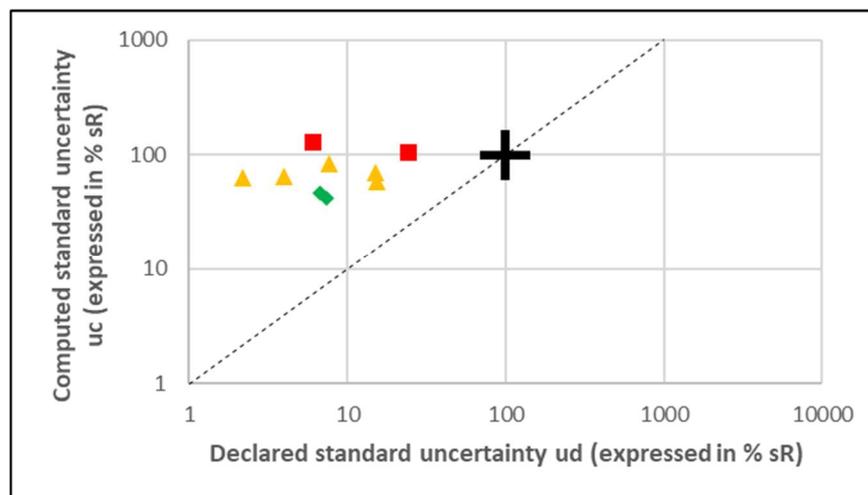


Figure 2.c : Résultats de 2023 pour les évaluations des incertitudes pour la vitesse de corrosion lors d'un essai Huey de corrosion intergranulaire sur un acier inoxydable.

Les Figure 2.a à c montrent que :

- ✚ Lorsque les incertitudes sont globalement bien déterminées (Figure 2.a, correspondant au cas 1 du § 3.2), u_c et u_d sont tous deux situés au voisinage de s_R . Cependant, la dispersion des valeurs de u_c est plus faible que celle des valeurs de u_d , ce qui nous laisse penser que u_c est plus précisément déterminé que u_d ;
- ✚ Lorsque les incertitudes sont globalement mal déterminées (Figure 2.b et c, correspondant aux cas 2 et 3 du § 3.2), les valeurs u_c se situent dans le voisinage de s_R alors que les valeurs u_d sont toutes inférieures à u_c et presque toutes significativement inférieures à s_R . Dans le cas 3, les rapports u_d/s_R se situent tous entre 3% et 30% ! De plus, la dispersion des valeurs u_c est également significativement plus faible que la dispersion des valeurs u_d , ce qui nous laisse penser que les valeurs u_c sont déterminées avec plus de précision que les valeurs u_d ;
- ✚ Les alertes provenant des scores ζ sont cohérentes avec les ratios u/s_R et u_d/u_c .

Conclusion concernant le calcul des incertitudes des participants individuels :

Chaque fois qu'il existe des CIL, l'utilisation de leurs résultats est une bonne solution pour calculer les incertitudes, car le plus gros du travail est effectué par des sources extérieures. Cependant, il convient de vérifier si certains termes sont manquants et, parfois, il faut recourir à une expérience partielle supplémentaire selon la méthode A du GUM [2] pour y parvenir. Par conséquent, elle peut parfois conduire à des sous-estimations lorsqu'il est techniquement ou économiquement compliqué de déterminer les termes manquants de l'équation (4).

6.3.3 Utilisation de l'ET de reproductibilité comme estimation de l'incertitude

Comme indiqué dans la norme ISO 21748 [11], l'ET s_R est une bonne base pour estimer les incertitudes, car les principales sources d'incertitude sont généralement incluses dans l'expérience et les termes éventuellement manquants sont bien identifiés (voir ci-dessus).

En outre, de nombreuses informations suffisamment fiables sont disponibles gratuitement à leur sujet, comme par exemple :

- ✚ Les normes, en particulier les normes ISO concernant la chimie et de nombreuses normes ASTM concernant les méthodes d'essai, comprennent des annexes qui contiennent des résultats d'expériences de fidélité par rapport à la norme ASTM E691 [9] ou ISO 5725-1 [6], (voir annexe) ;
- ✚ La littérature scientifique concernant les performances de nombreux types d'essais.

Toutefois, les valeurs u obtenues de cette manière sont pertinentes pour les laboratoires courants, c'est-à-dire :

- ✚ Ni les laboratoires très compétents, où des efforts importants sont faits pour réduire les incertitudes des résultats d'essai ;
- ✚ Ni les laboratoires peu compétents. La norme ISO 21748 [11] stipule qu'il convient de vérifier que le biais du laboratoire est maîtrisé (c'est-à-dire qu'il est conforme à ce qui est attendu lors d'une CIL).

Il est toujours possible de vérifier si le biais d'un laboratoire est maîtrisé en utilisant une expérience partielle selon la méthode A du GUM [2], comme expliqué au § 8.3.

Conclusion concernant l'utilisation de l'ET de reproductibilité comme estimation de l'incertitude :

Cette méthode très simple (il suffit de chercher des informations gratuites) produit une valeur de référence acceptable d'incertitudes pour les laboratoires normalement compétents, c'est-à-dire 90% d'entre eux.

6.3.4 Utilisation des ET de répétabilité et de reproductibilité pour estimer l'importance relative du biais et de l'erreur aléatoire

Dans la plupart des configurations de CIL :

- ✚ L'écart-type des valeurs moyennes des participants s_L est lié au biais moyen des participants ;
- ✚ L'écart-type des résultats des participants s_r est lié à la répétabilité ;
- ✚ Il existe une relation entre eux et l'écart-type de reproductibilité s_R , conformément à l'équation (5) de l'ISO 5725-2 [8], comme suit :

$$s_R = \sqrt{s_L^2 + s_r^2} \quad (5)$$

Par conséquent, le rapport s_r/s_L et le rapport s_r/s_R qui lui est mathématiquement lié fournissent des informations sur l'importance relative du biais et de l'erreur aléatoire pour une méthode d'essai donnée. De plus, comme au

§ 6.3.3, ces valeurs sont disponibles auprès de nombreuses sources. On peut distinguer trois situations types différentes :

- ✚ Lorsque $s_r/s_R < 0,2$. Dans ce cas, le biais est d'une importance majeure et les efforts doivent être concentrés sur sa détermination ;
- ✚ Lorsque $0,2 < s_r/s_R < 0,9$. Alors, le biais et l'erreur aléatoire peuvent être tous deux significatifs et doivent être pris en compte ;
- ✚ Lorsque $s_r/s_R > 0,9$. Dans ce cas, le biais n'a qu'une importance mineure et les efforts doivent être concentrés sur l'erreur aléatoire.

Cette dernière situation se produit lorsque la question de l'homogénéité interne du matériau soumis à l'essai revêt une importance majeure pour la détermination des incertitudes.

Lors des CIL CompaLab, le premier type de situation se produit généralement pour les essais chimiques, et le dernier pour la plupart des essais sur les aciers pour béton armé. Le deuxième type de situation se produit pour la plupart des autres méthodes d'essai.

Conclusion concernant le rapport s_r/s_R :

Des rapports s_r/s_R très faibles ou très élevés permettent au laboratoire d'éviter de consacrer du temps et de l'énergie à la détermination des contributions respectives de l'erreur aléatoire et du biais :

- ✚ Lorsque l'erreur aléatoire n'est pas significative par rapport au biais, une simple expérience partielle selon la méthode A du GUM [2] destinée à déterminer le biais (voir § 8.3.2) suffit pour déterminer correctement l'incertitude du laboratoire ;
- ✚ Lorsque le biais n'est pas significatif par rapport à l'erreur aléatoire, il n'est pas nécessaire de rechercher des sources externes pour les valeurs de référence (c.à.d. des MR ou une collaboration avec d'autres laboratoires).

7 Expérience selon la méthode A du GUM pour laquelle les objets soumis à essais sont des MR

7.1 Introduction

Les expériences selon la méthode A du GUM [2] consistent à réaliser un grand nombre d'essais pour lesquels les conditions de fidélité (telles que définies dans la norme ISO 5725-1 [6]) sont contrôlées. Ce contrôle des conditions de fidélité permet de déterminer l'erreur aléatoire. Si, de plus, il est effectué sur un MR (dont la valeur de référence est connue), le biais peut également être déterminé, et une détermination complète de l'incertitude élargie est possible.

Cette expérience est de type 1 selon le § 3.5, c'est-à-dire que l'évaluation est axée sur les résultats du processus d'essai.

De la même manière que précédemment, les problématiques suivantes doivent être prises en compte pour les essais de fidélité :

- ✚ Méthode ;
- ✚ Matériau et son homogénéité ;
- ✚ Niveaux de mesurande ;
- ✚ Préparation des éprouvettes ;

- ✚ Conditions d'essai (équipement d'essai, personnel, conditions environnementales).

Dans la pratique, ces expériences peuvent être réalisées :

- ✚ En organisant une expérience globale au sein du laboratoire, qui devrait couvrir toutes les sources internes d'incertitude ;
- ✚ En utilisant des données disponibles dans le laboratoire. Typiquement, les résultats d'un éventuel programme de surveillance de la qualité interne peuvent être utilisés.

7.2 Les valeurs de référence doivent provenir d'une source externe

Il est évident que le biais (qui est la déviation systématique du laboratoire) ne peut pas être estimé avec une expérience interne de fidélité (typiquement la méthode A ou B du GUM [2]). Par conséquent, il est très important que la valeur de référence X_{RM} provienne d'une source externe. Toutefois, il n'est pas nécessaire qu'il s'agisse d'un MRC (matériau de référence certifié), même si les MRC sont généralement les meilleures sources externes que l'on puisse trouver. Outre les MRC, les sources externes peuvent être les suivantes :

- ✚ Des objets qui ont été soumis à essais dans le cadre d'une CIL, dont le rapport fournit la valeur de référence centrale X_{RM} correspondante. Toutefois, cette option ne peut être utilisée que si la méthode d'essai n'est pas destructive ou si des objets soumis à essais sont encore disponibles après la CIL ;
- ✚ Les éléments dont la valeur de référence centrale X_{RM} peut être connue à partir d'une autre source, qui peut être interne ou externe ;

Par exemple, pour la détermination de la teneur en ciment du béton durci (qui est effectuée par une méthode d'analyse d'images), il est possible de produire en interne des échantillons dont la teneur en ciment est connue avec précision à partir de leurs paramètres de préparation.

- ✚ Des objets qui ont été préparés en interne à cette fin et testés par plusieurs laboratoires ;

Par exemple, une grande entreprise qui possède plusieurs filiales dans plusieurs endroits peut préparer un ensemble d'échantillons qui sont utilisés pour la surveillance de la qualité dans chaque endroit. Lorsque l'on dispose d'une grande quantité de résultats d'essais provenant de tous les lieux d'essai, l'ensemble des données peut être utilisé pour déterminer une valeur de référence.

Si la valeur de référence ne provient pas d'une source externe, l'estimation de U omet le biais, ce qui conduit à une sous-estimation significative de celui-ci (voir § 10.8).

7.3 Organisation d'une expérience globale en laboratoire

La première étape d'une telle expérience consiste à dresser la liste de toutes les sources d'incertitude rencontrées dans le laboratoire, puis à élaborer un plan d'expérience qui englobe toutes ces sources. Comme pour d'autres types d'expériences, le plan d'expérience doit représenter aussi fidèlement que possible les sources d'incertitude qui influencent les résultats des tests du laboratoire, ni plus, ni moins.

Par exemple, si un laboratoire prépare lui-même les éprouvettes, utilise 2 méthodes d'essai, 8 machines d'essai et 5 opérateurs, le plan d'expériences doit inclure :

- 2 méthodes d'essai ;
- 2 types de matériaux (moyen et difficile) ;
- 5 niveaux de mesurandes ;
- 3 sources de préparation des éprouvettes (ses propres installations et, dans la mesure du possible, 2 sources externes) ;
- 8 machines d'essai ;

- 5 opérateurs ;
- 2 conditions environnementales (c'est-à-dire réalisation des essais à 2 périodes différentes de l'année) ;
- 2 répétitions, si l'ET de répétabilité est nécessaire.

Soit un nombre total de $2 \times 2 \times 5 \times 3 \times 8 \times 5 \times 2 \times 2 = 9600$ essais.

Dans presque tous les cas, cela représente une quantité d'essais qu'il est techniquement ou économiquement impossible de prendre en charge. Il est toutefois possible de réduire considérablement ce nombre par les moyens suivants :

- ✚ Dans la plupart des cas, toutes les combinaisons ne s'appliquent pas dans la pratique. Le plan d'expérience ne doit représenter que la pratique réelle du laboratoire. En particulier, lorsque des corrélations apparaissent, elles doivent être prises en compte dans la conception de l'expérience ;

Par exemple, dans la plupart des cas, lorsqu'un laboratoire utilise 2 méthodes d'essai, 8 machines d'essai avec 5 opérateurs, chacun des opérateurs n'utilise que 2, 3 ou 4 machines, toutes les machines et tous les opérateurs ne sont pas utilisés pour produire des résultats avec les 2 méthodes.

- ✚ L'expérience peut être divisée en plusieurs expériences partielles, qui traitent séparément chacun des problèmes. Toutefois, les éventuelles interactions (favorables ou défavorables) entre les sources peuvent alors être omises ;

Par exemple, l'impact de la préparation peut être vérifié pour chacune des méthodes, mais avec une seule machine et un seul opérateur.

- ✚ Un plan d'expériences aléatoire peut être utilisé, dans lequel les contrôles sont répartis de manière à couvrir une sélection aléatoire de configurations parmi celles qui sont effectivement pratiquées dans le laboratoire (voir exemple en annexe) ;
- ✚ Ces expériences partielles ou plans d'expériences aléatoires peuvent inclure certains tests qui sont effectués sur MR et d'autres qui ne sont pas effectués sur MR.

Pour répondre à la 1^e question, il faut établir une liste des configurations réellement utilisées dans le laboratoire.

La deuxième option demande de s'assurer qu'il n'y a pas d'interaction entre les facteurs. Dans ce cas, le plan d'expérience doit en tenir compte.

Par exemple, l'impact de la préparation peut être différent selon les machines utilisées.

Lorsque cela est possible (c'est-à-dire lorsque l'on dispose de suffisamment de connaissances sur la question), il peut également être décidé de choisir une configuration moyenne, représentant l'effet moyen de la préparation.

L'utilisation de la troisième option doit permettre de s'assurer que le plan d'expérience aléatoire couvre bien les configurations réelles utilisées dans le laboratoire. En particulier, le nombre total d'essais ne doit pas être trop réduit et aucune source importante d'incertitude ne doit être omise dans le schéma d'essai.

Par exemple, si l'une des huit machines d'essai est utilisée pour produire 40 % des résultats des essais, le schéma aléatoire doit l'exclure de l'expérience.

Lorsqu'un grand nombre de configurations est effectivement utilisé, il est possible de les pondérer en fonction de la fréquence à laquelle elles sont utilisées, afin de produire le plan d'expérience aléatoire.

Un exemple de plan d'expérience complet est fourni en annexe.

Conclusions concernant l'organisation d'une expérience selon la méthode A du GUM pour laquelle les objets soumis à essais sont des matériaux de référence :

Lorsqu'elles sont bien conçues, ces expériences fournissent de très bonnes estimations des incertitudes, mais elles requièrent la disponibilité de matériaux de référence et sont coûteuses en temps et en argent.

7.4 Utilisation d'un traitement statistique des résultats des essais effectués pour la surveillance de la qualité

La norme ISO/IEC 17025 [1] (qui est la référence pour l'accréditation des laboratoires) exige des laboratoires qu'ils effectuent une surveillance de la qualité qui, entre autres possibilités, peut prendre la forme de cartes de contrôle. D'autre part, certaines normes décrivant les méthodes d'essai (par exemple ISO 6507-1 [7]) demandent également de vérifier périodiquement la machine d'essai à l'aide d'un MR. Cela produit un ensemble de données qui peuvent être utilisées pour déterminer les incertitudes. La Figure 3 donne un exemple de ces données.

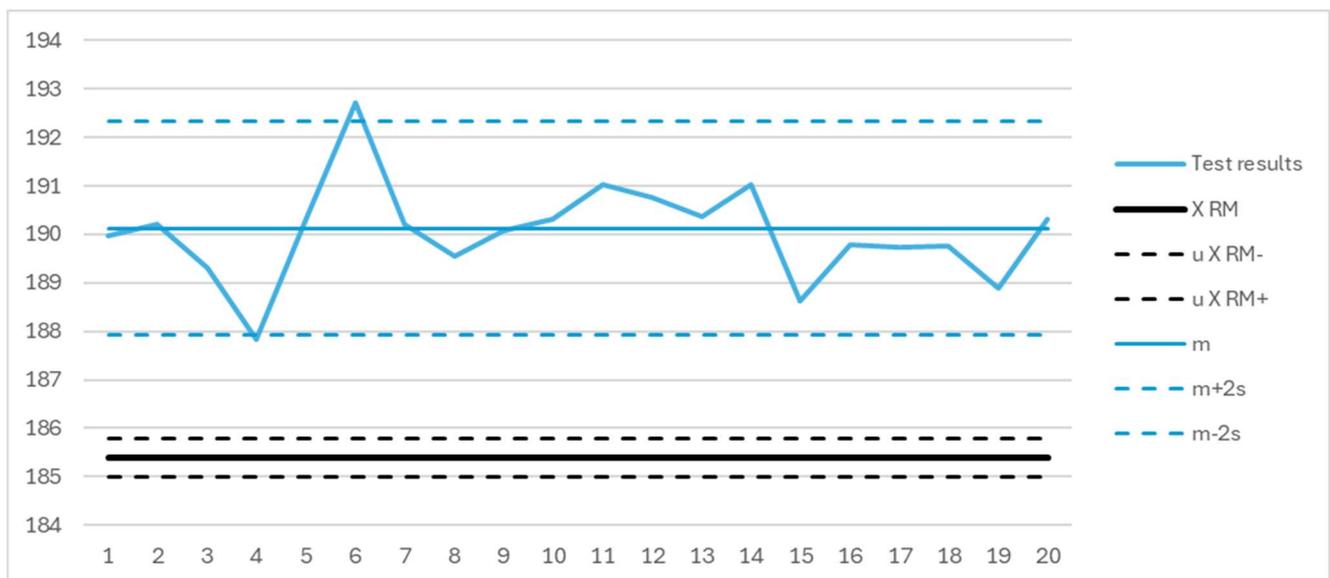


Figure 3 : Exemple de carte de contrôle dont les résultats peuvent être utilisés pour déterminer l'incertitude sur les résultats des tests.

Dans cet exemple, on peut constater que :

- ✚ $m - X_{CRM}$ où m est la valeur moyenne des résultats des tests sur le MR et X_{CRM} est la valeur assignée du MR, est un bon estimateur du biais B ;
- ✚ s est un bon estimateur de l'erreur aléatoire e .

Il convient également de noter que la carte de contrôle de la Figure 3 ne ressemble pas aux cartes de contrôle habituelles décrites dans la série ISO 7870. Les cartes de contrôle typiques sont normalement centrées sur la valeur de référence et les limites d'alerte sont normalement calculées et affichées sur la carte. Il est important de noter que l'objectif des cartes de contrôle habituelles est de contrôler la dérive et non le biais et la dispersion. En conséquence :

- ✚ La valeur centrale d'une carte de contrôle habituelle est déterminée en interne et représente un point "zéro" de l'état du processus. Si une dérive se produit, elle peut être considérée comme un écart par rapport à ce point "zéro" ;

- ✚ Une carte de contrôle habituelle ne peut pas être établie uniquement en utilisant des matériaux de référence externes, car ses limites d'IC ne représentent normalement pas l'écart acceptable par rapport à la valeur de référence (soit elles sont définitivement trop sévères, soit le MR est très mauvais).

Dans les petits laboratoires, les cartes de contrôle de la surveillance de la qualité des essais comprennent généralement des résultats d'essais provenant de différentes machines d'essai et de différents opérateurs, chacun d'entre eux étant contrôlé alternativement. Mais dans les grands laboratoires ou lorsque la qualité des résultats d'essai est très critique, plusieurs, voire de nombreuses cartes de contrôle peuvent être mises en œuvre, chacune d'entre elles étant consacrée à une machine spécifique ou à un opérateur spécifique. Évidemment, dans le premier cas, les paramètres B et e calculés sont ceux de l'ensemble du laboratoire, tandis que dans le second cas, les paramètres B et e calculés sont ceux de l'équipement d'essai ou de l'opérateur correspondant. Par conséquent, les U correspondants peuvent être différents d'une carte de contrôle à l'autre.

Conclusions concernant l'utilisation d'un traitement statistique des résultats des essais effectués pour la surveillance de la qualité :

Lorsque de telles données existent au sein du laboratoire, elles doivent absolument être utilisées pour déterminer les incertitudes car la méthode de base est très efficace et le temps et l'argent supplémentaires nécessaires pour y parvenir sont très faibles.

7.5 Calcul de U

L'incertitude élargie U peut alors être calculée à l'aide de l'équation (6), comme suit :

$$U = |m - X_{RM}| + k \cdot \sqrt{u_{RM}^2 + s^2} \quad (6)$$

où U est l'incertitude élargie du résultat de l'essai,
 m est la valeur moyenne de la série de résultats d'essai,
 X_{RM} est la valeur assignée du MR (matériau de référence)
 k est le coefficient d'élargissement,
 u_{RM} est l'incertitude sur la valeur X_{RM} du matériau de référence,
 s est l'écart-type de la série de résultats d'essai.

Dans l'exemple de la Figure 3, $U = |190,0 - 185,4| + 2 \cdot \sqrt{0,4^2 + 1,0^2} = 6,8$, ce qui englobe bien l'écart entre n'importe lequel des résultats du test et la valeur de référence.

On peut faire valoir que m et s sont des estimateurs qui peuvent être un peu éloignés des vraies valeurs qu'ils sont censés représenter. Pour y remédier, il est possible d'inclure dans le terme représentant e un élargissement en utilisant une valeur t provenant des tables de Student au lieu du coefficient d'élargissement k habituel. Ceci est cependant contradictoire avec le premier principe du GUM, qui stipule que les déterminations de u doivent être aussi réalistes que possible plutôt que de déterminer par excès.

Ceci a été confirmé par la méthode de Monte-Carlo (voir § 10.11), dont les résultats sont les suivants :

- ✚ $U = 6,819 \pm 0,006$;
- ✚ 95% des estimations de U appartiennent à l'intervalle [5,99;7,65].

7.6 Commentaires concernant B et e

Conformément aux conclusions du § 3.3, il est nécessaire de distinguer ce qui est inclus dans le terme B et ce qui est inclus dans le terme e au cours des essais de surveillance de la qualité.

Si les données des résultats d'essai proviennent d'une expérience organisée menée sur une courte période, le terme e peut être plus faible que lorsque les résultats d'une carte de contrôle sont utilisés, pour lesquels les résultats ont été obtenus sur une longue période.

Dans la plupart des cas, les MR sont fournis sous la forme d'éprouvettes. Ainsi, les résultats utilisés dans les deux situations (expérience organisée et cartes de contrôle) n'incluent pas les impacts de la préparation des éprouvettes. En conséquence :

- ✚ La contribution de la préparation des éprouvettes de MR apparaît comme un biais dans ce type d'expérience d'incertitude ;
- ✚ Ce biais est pris en compte dans la valeur u_{XRM} ;
- ✚ L'expérience n'inclut pas la contribution de la préparation des éprouvettes dans la vie quotidienne du laboratoire, et cette contribution doit être estimée séparément, par exemple par une expérience partielle selon la méthode A du GUM [2], cf. § 8.3.6 ;
- ✚ Les MR ont généralement une meilleure homogénéité interne que les échantillons d'essai quotidiens. L'impact correspondant sur l'incertitude pourrait alors être sous-estimé.

De la même manière, lorsque plusieurs méthodes peuvent être utilisées pour obtenir les résultats d'essai, il convient de vérifier si le MR est valable pour une seule méthode ou non et de décider en conséquence si une contribution de la méthode à l'incertitude est nécessaire ou non.

En tout état de cause, il est nécessaire, pour chaque situation, d'analyser soigneusement ce qui est inclus dans le terme B et ce qui est inclus dans le terme e .

7.7 Conclusions concernant les expériences menées selon la méthode A du GUM avec des MR comme éléments testés

Les expériences menées conformément à la méthode A du GUM [2] en utilisant des MR comme éléments testés constituent généralement un type d'expérience efficace, car elles n'oublient généralement aucune source majeure d'incertitude. Cependant, il faut toujours vérifier si les contributions (biais ou erreur aléatoire) de la préparation des éprouvettes, du matériau et de son homogénéité (biais et/ou erreur aléatoire), du niveau de mesurande et du biais éventuel de la méthode d'essai doivent être ajoutées ou non.

Lorsque les résultats d'un programme interne de surveillance de la qualité sont disponibles, ils devraient toujours être utilisés comme données pour ces expériences, car la détermination de U devient alors facile et peu coûteuse (une valeur moyenne et un écart type à calculer). Toutefois, dans de nombreux cas, il est difficile ou coûteux (lorsque la méthode d'essai est destructive) de trouver une source externe pour la valeur de référence X_{RM} .

Dans d'autres cas, il est toujours possible d'organiser une expérience globale pour laquelle l'organisateur peut contrôler tous les paramètres et, par conséquent, traiter toutes les sources d'incertitude, en particulier le biais. Cependant, il peut être techniquement ou économiquement difficile d'obtenir des MR en quantité suffisante et il peut être nécessaire de consacrer beaucoup de temps et d'argent à la réalisation des essais requis pour obtenir les résultats. En conclusion, l'organisation d'une telle expérience n'est pertinente que lorsqu'il est nécessaire pour fournir des valeurs précises de U .

8 Expériences selon la méthode A du GUM pour lesquelles les objets soumis à essai ne sont pas des MR

8.1 Introduction

Lorsqu'aucun MR n'est disponible, il est toujours possible d'organiser des expériences selon la méthode A du GUM [2], qui peuvent être :

- ✚ Des expériences globales ou l'utilisation des résultats de programmes de surveillance de la qualité interne ;
- ✚ Ou des expériences partielles, qui traitent d'une source spécifique de dispersion, en complément d'autres expériences.

8.2 Expériences globales ou utilisation des résultats des programmes de surveillance interne de la qualité

Les expériences globales réalisées selon la méthode A du GUM [2] sur des objets soumis à essai dont la valeur de référence n'est pas connue par une source externe ne permettent pas d'estimer le biais et, par conséquent, ignorent une source majeure d'incertitude dans la plupart des cas.

Par exemple, dans l'exemple de la Figure 3, $U = 6,8$ lorsque le biais est pris en compte et $U = 2,2$ lorsqu'il ne l'est pas.

Par conséquent, cette méthode n'est valable que si :

- ✚ Le biais du laboratoire peut être estimé par une expérience partielle séparée, cf. § 8.3.2 ;
- ✚ Ou lorsqu'il peut être démontré qu'aucun biais significatif ne s'applique à ce type de test, cf. § 6.3.4.

Dans ce cas, les recommandations du § 7 et du § 8.3.8 sont applicables.

Sinon, une telle expérience est en fait une expérience partielle telle que décrite au § 8.3.

8.3 Expériences partielles selon la méthode A du GUM [2]

8.3.1 Introduction

Les expériences partielles de la méthode A du GUM [2] peuvent être utilisées pour déterminer l'impact d'une ou de plusieurs sources d'incertitude identifiées, mais pas de toutes ensemble. En particulier, ces expériences partielles peuvent être utilisées pour déterminer les contributions qui sont difficiles ou impossibles à déterminer par d'autres moyens, telles que :

- ✚ Le biais ;
- ✚ La méthode d'essai ;
- ✚ L'effet du matériau ;
- ✚ L'étendue des mesurandes ;
- ✚ La préparation des éprouvettes ;
- ✚ Les conditions de fidélité telles que définies dans la norme ISO 5725-1 [6].

8.3.2 Expérience partielle selon la méthode A du GUM [2] pour déterminer le biais

Pour déterminer l'impact du biais, une expérience selon la méthode A du GUM [2] peut être organisée comme suit :

- ✚ Approvisionner une série de MR. Il n'est pas nécessaire qu'il s'agisse d'un MRC. D'autres possibilités énumérées au § 7.2 peuvent également être utilisées. Si aucun MR n'est disponible auprès de sources externes, il est

nécessaire de les préparer en interne et de les faire tester par un nombre suffisant de laboratoires indépendants les uns des autres (voir § 10.7 pour déterminer le nombre minimum à utiliser) ;

- ✚ Calculer la valeur moyenne de chaque laboratoire et la moyenne générale. Les valeurs d'ET peuvent également être déterminées à titre d'information ;
- ✚ Calculer le biais du laboratoire pour chacun des niveaux de performance.

8.3.3 Expérience partielle selon la méthode A du GUM [2] pour déterminer l'impact de la méthode d'essai

Pour déterminer l'impact de la méthode d'essai, une expérience selon la méthode A du GUM [2] peut être organisée comme suit :

- ✚ Préparer un nombre suffisant d'échantillons aussi semblables que possible ;
- ✚ Effectuer les essais selon chaque méthode d'essai sur un nombre égal d'échantillons ;
- ✚ Calculer la valeur de référence à partir des résultats obtenus avec la méthode de référence s'il y en a une, à partir de tous les résultats s'il n'y en a pas ;
- ✚ Calculer les valeurs globales de l'écart-type ;
- ✚ Calculer le biais et l'écart-type liés aux méthodes d'essai habituellement utilisées par le laboratoire.

8.3.4 Expérience partielle selon la méthode A du GUM [2] pour déterminer l'impact du matériau soumis à l'essai

Pour déterminer l'impact du matériau soumis aux essais, une expérience selon méthode A du GUM [2] peut être organisée comme suit. En fonction des ressources que le laboratoire est prêt à consacrer à la détermination des incertitudes, plusieurs options sont possibles :

- ✚ Décider d'un ou plusieurs matériaux correspondant à différents niveaux de difficultés pour la réalisation des essais (cf. § 3.4) ;
- ✚ Préparer un nombre suffisant d'échantillons de chaque matériau, aussi similaires que possible au sein de chaque matériau ;
- ✚ Effectuer les essais dans un seul laboratoire sur toutes les éprouvettes ;
- ✚ Calculer les moyennes et les valeurs de l'écart-type pour chaque niveau de mesurande ;
- ✚ Attribuer des valeurs d'incertitude pour chacun d'entre eux.

8.3.5 Expérience partielle selon la méthode A du GUM [2] pour déterminer l'impact des niveaux des résultats d'essai

Pour déterminer l'impact des niveaux des résultats d'essai, une expérience selon la méthode A du GUM [2] peut être organisée comme suit :

- ✚ Préparer un nombre suffisant d'échantillons d'au moins 3 niveaux de mesurande, aussi similaires que possible à l'intérieur de chaque niveau de mesurande. Selon l'importance de l'impact, 3 à 10 niveaux doivent être testés. Ces niveaux doivent être choisis parmi les nombres préférentiels décrits dans la norme ISO 3 [13] (c'est-à-dire 1 - 1,25 - 1,6 - 2 - 2,5 - 3,15 - 4 - 5 - 6,3 - 8 - 10 - etc....) ;
- ✚ Effectuer les essais dans un même laboratoire sur toutes les éprouvettes ;
- ✚ Calculer les moyennes et les valeurs de l'écart-type pour chaque niveau de mesurande ;
- ✚ Tracer ces résultats en fonction des niveaux de mesurande et calculer une relation entre les deux (généralement sous la forme $U = a.V + b$).

8.3.6 Expérience partielle selon la méthode A du GUM [2] pour déterminer l'impact de la préparation des échantillons

Pour déterminer l'impact de la préparation des échantillons, une expérience selon la méthode A du GUM [2] peut être organisée comme suit :

- ✚ Préparer un nombre suffisant d'échantillons aussi similaires que possible ;
- ✚ Demander à plusieurs entités différentes de préparer un nombre égal d'éprouvettes ;
- ✚ Effectuer les essais dans un même laboratoire sur toutes les éprouvettes ;
- ✚ Calculer la moyenne générale et les valeurs de l'écart-type ;
- ✚ Calculer le biais et l'écart-type liés à la préparation pour les entités qui préparent habituellement les échantillons pour le laboratoire.

8.3.7 Expérience partielle selon la méthode A du GUM [2] pour déterminer l'impact des conditions de fidélité

Pour déterminer l'impact des conditions de précision (telles que définies dans la norme ISO 5725-1 [6]), une expérience selon la méthode A du GUM [2] peut être organisée comme suit :

- ✚ Préparer un nombre suffisant d'échantillons aussi similaires que possible ;
- ✚ Effectuer l'essai sur un nombre égal d'échantillons, en utilisant tous les équipements d'essai, les opérateurs sur une période suffisamment longue pour représenter la variété des conditions environnementales rencontrées dans le laboratoire tout au long de l'année ;
- ✚ Si la détermination de l'ET de répétabilité est nécessaire, chaque configuration doit être testée au moins 2 fois ;
- ✚ Calculer la valeur de l'écart-type de répétabilité et l'écart-type global selon l'équation (7).

$$s_r = \sqrt{(\sum_i s_i^2)/(n - 1)} \quad (7)$$

*où s_r est l'écart-type de répétabilité estimé,
 s_i sont les écarts types individuels calculés pour chaque configuration d'essai individuelle
 de l'équipement d'essai, de l'opérateur et des conditions environnementales,
 n est le nombre total de configurations d'essai.*

La proposition ci-dessus inclut toutes les contributions à l'erreur aléatoire au sein du laboratoire. Il est bien sûr possible de limiter l'expérience à certaines contributions seulement.

8.3.8 Mise à jour de U avec les résultats des expériences partielles selon la méthode A du GUM [2]

L'équation (4) peut être utilisée pour compléter le calcul de U . Lorsque les résultats de l'équation (4) sont connus de manière consolidée, l'ajout de termes supplémentaires provenant d'une expérience partielle selon la méthode A du GUM [2] ne doit pas être mis en œuvre de la même manière pour le biais et pour l'erreur aléatoire.

Biais :

La partie représentant le biais dans l'équation (4) est la suivante : $B = |\sum B_i|$. L'ajout d'un terme supplémentaire B_n à B peut l'augmenter ou le diminuer, selon que B_n est de même signe que $\sum B_i$ ou non. En conséquence :

- ✚ Lorsque $\sum B_i$ est connu, B mis à jour doit être calculé comme $B_{updated} = |\sum B_i + B_n|$;
- ✚ Lorsque seul B est connu, B mis à jour doit être calculé comme $B_{updated} = B + |B_n|$, ce qui peut être surestimé.

Erreur aléatoire :

La partie représentant l'erreur aléatoire dans l'équation (4) est la suivante : $e^2 = \sum e_i^2$. L'ajout d'un terme supplémentaire e_n à e revient à ajouter un terme e_n^2 à la somme $\sum e_i^2$. Par conséquent, dans tous les cas, $e_{updated}^2 = e^2 + e_n^2$.

8.3.9 Conclusions concernant les expériences partielles selon la méthode A du GUM [2]

Ces expériences partielles ne permettent pas d'étudier les interactions entre chacune des sources d'incertitude. Par conséquent, elles doivent être considérées comme complémentaires à une expérience globale lorsqu'il n'est pas techniquement ou économiquement possible de les inclure dans l'expérience principale.

Elles ne peuvent pas fournir une valeur fiable de U mais peuvent être très utiles pour compléter les calculs lorsque l'on dispose d'informations provenant d'ailleurs concernant certaines contributions mais pas toutes.

Dans de nombreux cas, on ne dispose pas de suffisamment d'informations externes et il est techniquement ou économiquement très difficile, voire impossible, d'organiser une expérience complète selon la méthode A du GUM [2]. Dans ces cas, elles sont vraiment utiles pour déterminer correctement U .

9 Études selon la méthode B du GUM**9.1 Introduction**

La méthode B du GUM [2] diffère significativement de toutes les méthodes exposées ci-dessus car elle n'est pas principalement basée sur des expériences mais sur une analyse des causes d'incertitude et une combinaison des effets de chacune de ces causes, selon l'équation (8), comme suit :

$$u = \sqrt{\sum_i u_i^2} \quad (8)$$

où u est l'incertitude globale estimée,

u_i sont les valeurs individuelles des effets de chaque cause d'incertitude.

Commentaires sur cette méthode :

- ✚ Il s'agit donc clairement d'une méthode de type 2 selon la classification du § 3.5, qui suppose de transformer l'incertitude sur les causes en incertitudes sur les effets ;
- ✚ On constate également que le biais ne peut pas être traité séparément par cette méthode, et qu'il doit être traité comme un effet aléatoire parmi d'autres ;
- ✚ Cette méthode ne demande pas nécessairement la réalisation d'expériences. De nombreuses informations (typiquement toutes les incertitudes liées aux mesures métrologiques) peuvent provenir de documents ou de toute autre information externe pertinente.

9.2 Mise en œuvre de la méthode B du GUM

Les opérations suivantes sont nécessaires pour la mettre en œuvre :

1. Dresser une liste précise de toutes les sources d'incertitude qui se produisent au cours du processus d'essai ;
2. Évaluer l'incertitude sur chacun de ces paramètres d'entrée ;

3. Évaluer l'effet individuel de ces paramètres sur le résultat final ;
4. Calculer la combinaison de tous les effets.

Dresser la liste des sources d'incertitude :

Cette liste doit être :

- ✚ Complète : si certaines sources d'incertitude sont omises, il manquera certains termes dans l'équation (8) et le résultat final u sera sous-estimé. De nombreuses méthodes peuvent être utilisées pour la rendre aussi complète que possible. En règle générale, l'analyse doit prendre en compte les sources liées à la méthode, au matériau, à l'homogénéité, aux niveaux des mesurandes, à la préparation des éprouvettes, à l'équipement, au personnel et aux conditions environnementales. Pour traiter ces éléments, il est souvent utile de vérifier la norme décrivant la méthode d'essai : il est évident que tous les paramètres qui doivent être maîtrisés sont ceux qui influencent le résultat de l'essai. Par conséquent, tout écart minime de ces paramètres est susceptible d'entraîner un écart minime dans le résultat de l'essai. La littérature scientifique peut également être utilisée pour rendre la liste aussi complète que possible. Toutefois, il peut être difficile d'atteindre l'exhaustivité car certaines sources peuvent rester inconnues ;
- ✚ Sans double comptage : si certaines sources d'incertitude sont comptées deux fois, l'équation (8) contiendra plusieurs fois le même effet et le résultat final sera surestimé. Cependant, la pratique montre qu'il peut être difficile de penser à certains doubles comptages cachés ;

Exemple 1 : lorsqu'une évaluation globale de l'effet de l'équipement d'essai est réalisée et que cette évaluation globale inclut le biais, l'incertitude liée à l'étalonnage de capteurs apparentés ne doit pas être considérée comme une autre source d'incertitude.

Exemple 2 : lorsqu'un rapport d'étalonnage d'un capteur est utilisé, il inclut la contribution de la résolution de celui-ci. Par conséquent, aucune contribution de la résolution ne doit être ajoutée.

Exemple 3 : lorsqu'une marque est appliquée manuellement sur des éprouvettes (pour mesurer l'allongement lors d'essais de traction) et que l'effet des opérateurs est déterminé par une expérience de la méthode A utilisant plusieurs éprouvettes, l'effet de cette opération manuelle ne doit pas être ajouté car il est déjà pris en compte dans l'expérience de la méthode A.

- ✚ Avec une corrélation entre eux maîtrisée. L'équation (8) n'est valable que si toutes les sources d'incertitude sont statistiquement indépendantes, c'est-à-dire si, dans la pratique, il n'existe aucune relation entre les termes. Vérifier cela statistiquement est un travail énorme que personne ne fait jamais. La façon la plus simple de le vérifier est de l'examiner d'un point de vue technique. Lorsque de telles corrélations sont susceptibles de se produire, la meilleure façon est d'évaluer les effets connexes ensemble dans une analyse globale ou par une expérience partielle selon la méthode A GUM [2] (cf. § 8.3).

Par exemple, lorsque le même équipement de test est toujours utilisé par les mêmes opérateurs, une corrélation est susceptible de se produire entre l'effet de l'équipement et l'effet de l'opérateur. Il est alors conseillé de considérer un effet couplé équipement-opérateur qui inclut les deux effets et leur corrélation.

Évaluer l'incertitude liée à chacun de ces paramètres d'entrée :

L'incertitude liée à chacun des paramètres d'entrée identifiés doit être estimée. Elle peut provenir :

- ✚ Des exigences de la méthode d'essai ;
- ✚ Des résultats d'expériences antérieures, y compris les résultats de participations antérieures du laboratoire à des CIL ;

- ✚ De l'expérience générale ou de la connaissance des propriétés des matériaux et des instruments utilisés ;
- ✚ Des spécifications du fabricant ;
- ✚ Des données des certificats d'étalonnage ou autres ;
- ✚ Des incertitudes attribuées aux valeurs de référence.

D'une manière générale, le laboratoire doit choisir entre des valeurs générales (généralement des exigences de la méthode d'essai) et des valeurs qui représentent la vie réelle du laboratoire.

Par exemple, si la température est une source d'incertitude, on peut choisir l'intervalle de tolérance de la norme [15°C;25°C] ou l'intervalle réel d'exécution de l'essai au sein du laboratoire [18°C;23°C]. La première option est plus facile, la seconde demande plus de travail mais est meilleure pour le laboratoire.

Evaluation de l'effet de chaque source d'incertitude sur le résultat final :

La transformation de la dispersion sur un paramètre d'entrée doit être transformée en une dispersion sur résultat final. En d'autres termes, il est nécessaire de déterminer dans quelle mesure le résultat final est modifié par les petits changements qui se produisent effectivement pour le paramètre d'entrée. Le rapport correspondant est parfois appelé coefficient de sensibilité. Les méthodes suivantes permettent d'atteindre cet objectif :

- ✚ Lorsque le paramètre d'entrée fait partie des variables d'une équation qui produit le résultat final, la dérivation partielle de cette équation par rapport au paramètre d'entrée. Le GUM [2] fournit la dérivée des fonctions habituelles (typiquement, addition, multiplication, polynômes, racines carrées). Ces fonctions dérivées peuvent ensuite être utilisées avec l'intervalle pertinent de dispersion des paramètres d'entrée. Ces intervalles de dispersion peuvent être ceux de la norme ou ceux qui se produisent réellement dans le laboratoire (qui peuvent être plus étroits). Lorsque cette opération est compliquée (parce que l'équation initiale est compliquée), ces coefficients de sensibilité peuvent être déterminés par la méthode de Monte-Carlo ;

Exemple 1 : La masse linéique LM d'un acier pour béton armé est déterminée sur un tronçon dont la masse M et la longueur L sont mesurées. LM est alors donnée par l'équation $LM = M/L$. Dans ce cas, il est facile de trouver que $u_M = \Delta M/M$ et $u_L = \Delta L/L$, où ΔM et ΔL représentent la dispersion rencontrée en laboratoire autour de la "vraie" valeur de M et L respectivement.

Exemple 2 : l'ISO TR 15263 fournit les résultats de la dérivation de certaines formules complexes concernant l'essai de traction des métaux.

- ✚ Lorsque le paramètre d'entrée est une variable numérique qui ne fait pas partie d'une équation produisant le résultat final, il est nécessaire de déterminer l'effet d'une variation contrôlée de ce paramètre sur le résultat final de l'essai. Cela peut être réalisé par une expérience interne du laboratoire selon la méthode A du GUM [2] ou par une source externe ;

Par exemple, dans de nombreuses méthodes d'essai, la température de réalisation des essais doit être contrôlée. Dans ces cas, il est nécessaire de déterminer dans quelle mesure le résultat final de l'essai est modifié par les variations de température réellement rencontrées dans le laboratoire.

- ✚ Lorsque le paramètre d'entrée est une variable qualitative, il est nécessaire de déterminer l'effet d'une variation contrôlée de cette variable sur le résultat final de l'essai. Cela peut être réalisé par une expérience partielle selon la méthode A du GUM [2], par l'expérience interne du laboratoire ou par une source externe ;

Exemple 1 : La force de cisaillement d'un joint soudé d'un treillis soudé est déterminée par un essai de traction pour lequel l'éprouvette est maintenue dans un "porte-échantillon". La norme de référence ISO 15630-2 prévoit plusieurs exigences qualitatives pour ce porte-échantillon, comme "empêcher le fil transversal de tourner". Il est alors nécessaire de déterminer

dans quelle mesure la conception du porte-échantillon utilisé dans le laboratoire produit des écarts dans les forces de cisaillement déterminées par celui-ci.

Exemple 2 : L'effet du matériau est une variable quasi-qualitative, qui est assez compliquée à traiter lorsqu'aucune expérience partielle selon la méthode A du GUM [2] n'est réalisée.

- ☞ Certaines méthodes d'essai produisent une courbe reliant 2 paramètres ou plus. Le résultat de l'essai peut être la courbe elle-même ou les paramètres calculés à partir de celle-ci. Dans ces cas, il est nécessaire de déterminer l'effet d'une variation contrôlée de cette variable sur la courbe ou sur les paramètres qui sont calculés à partir de celle-ci ;

Exemple 1 : courbes de transmittance ou de réflectance dans l'infrarouge, utilisées pour identifier un produit par comparaison avec des courbes de référence de produits connus.

Exemple 2 : essai de traction des métaux conformément à la norme ISO 6892-1, dont les résultats bruts sont exprimés sous la forme d'une courbe force-allongement. Cette norme définit plusieurs paramètres et la manière de les déterminer à partir de la courbe. La détermination du $R_{p0,2}$ nécessite un traitement de la courbe qui n'est ni entièrement quantitatif ni entièrement qualitatif.

Lorsque l'impact est déterminé comme un intervalle global, cet intervalle global doit être divisé par 4 (correspondant à un IC95% pour les distributions gaussiennes) pour être cohérent avec u qui est une incertitude-type.

Par exemple, si la température est une source d'incertitude et que l'intervalle de tolérance sur celle-ci est fixé à $[15^{\circ}\text{C};25^{\circ}\text{C}]$, l'impact d'une variation de $(25-10)/4 = 2,5^{\circ}\text{C}$ doit être pris en compte.

Dans d'autres cas, d'autres types de distributions peuvent s'appliquer. Le GUM [2] fournit les équations permettant de calculer u pour une sélection de types de distributions rencontrés dans la pratique. La méthode de Monte-Carlo peut être utilisée pour déterminer les valeurs de u pour les types de distribution pour lesquels aucune information n'est disponible.

Par exemple, les valeurs de u liées à la distribution de la résolution de l'équipement d'essai sont généralement carrées. Dans ce cas, $u = a/\sqrt{3}$, où a est l'intervalle entre deux graduations d'un instrument de mesure.

Si le travail a été effectué correctement, tous les u_i doivent être exprimés dans l'unité du résultat de l'essai ou en pourcentage de celui-ci. Si cette condition n'est pas remplie, la détermination de U risque d'être erronée.

Calcul de la combinaison de tous les effets :

Ce calcul est réalisé à l'aide de l'équation (8).

9.3 Conclusions sur la méthode B du GUM

Étant donné qu'elle ne nécessite aucune expérience, environ 90% des déterminations d'incertitudes sont effectuées à l'aide de cette méthode. Elle a également des raisons fondamentales de conduire à des incertitudes sous-estimées, et c'est pourquoi la plupart des incertitudes déclarées par les laboratoires sont sous-estimées. Ces raisons sont détaillées ci-après :

1. Le biais est considéré comme l'une des causes aléatoires de l'incertitude. Cela conduit à une sous-estimation de U , cf. § 10 ;
2. Les contributions inconnues ne sont pas prises en compte ;

3. Cette méthode est utilisée parce qu'elle évite l'expérimentation. Par conséquent, même lorsqu'une expérience serait absolument nécessaire (c'est-à-dire presque à chaque fois qu'un facteur qualitatif intervient), ce n'est pas le cas dans la pratique. Au cours de notre expérience d'auditeur d'accréditation, nous avons constaté à maintes reprises que les paramètres d'entrée qualitatifs sont systématiquement considérés comme ayant une influence négligeable, simplement parce qu'il est très compliqué de les prendre en compte sérieusement. Par conséquent, les principales contributions sont souvent omises, ce qui conduit à une sous-estimation, voir § 10.

10 Questions et outils applicables à plusieurs des méthodes décrites

10.1 Formats non numériques pour les résultats d'essai

10.1.1 Introduction

Les résultats des essais peuvent être exprimés dans des formats non numériques, comme par exemple :

- ✚ Une classification en catégories (par exemple la souplesse et les arômes d'un vin), et ces catégories peuvent être ordonnées ou non (la souplesse peut être ordonnée de « sec » à « liquoreux », alors que les arômes ne peuvent pas l'être) ;
- ✚ Des résultats binaires (par exemple réussite/échec ou présence/absence d'un agent polluant) qui peuvent être considérés comme une classification en deux catégories ;
- ✚ Des pourcentages des catégories (par exemple, classification du graphite dans la fonte selon la norme ISO 945, où les résultats d'essais sont fournis en % de catégories définies par des images de référence fournies par la norme) ;
- ✚ Des courbes (typiquement la spectrométrie infrarouge dans laquelle la courbe IR est comparée à une référence pour identifier le produit auquel appartient le produit testé).

Tous les outils et méthodes habituellement décrits pour déterminer les incertitudes ne peuvent pas être appliqués à ces formats. Deux possibilités peuvent être mises en œuvre pour surmonter cette difficulté :

1. Transformer les résultats non numériques en résultats numériques ;
2. Utiliser des statistiques qui peuvent être appliquées à des informations non numériques (c'est-à-dire des statistiques non paramétriques).

Cette question est, dans un autre contexte, également traitée de façon extensive dans l'ISO 33406 .

10.1.2 Résultats des tests exprimés en catégories

Statistiques appliquées aux valeurs des catégories :

Il est incorrect de calculer les valeurs moyennes et les écarts types des valeurs des catégories, même lorsqu'elles sont ordonnées, parce que les valeurs des catégories ne sont pas de véritables "valeurs numériques" mais des "noms" qui peuvent être transformés en n'importe quel autre type d'étiquette (lettres, symboles, ...) sans que cela n'entraîne une perte d'information.

Toutefois, certaines méthodes d'essai définissent les catégories de telle sorte qu'elles sont effectivement des "valeurs numériques".

Exemple : Les normes ISO 643 (détermination de la taille des grains des métaux) et ISO 4967 (détermination des inclusions non métalliques dans l'acier) définissent les catégories (respectivement la taille des grains et la taille des inclusions non

métalliques) comme des gammes de tailles régulièrement distribuées, suivant une équation de loi de puissance. Il s'agit donc d'une sorte de "valeur numérique" arrondie de manière non décimale.

La première étape consiste donc à vérifier comment le document de référence de la méthode d'essai définit les catégories. Si elles sont définies comme des plages de valeurs numériques régulièrement réparties, les techniques numériques habituelles peuvent être appliquées.

Utilisation des valeurs numériques sous-jacentes :

Dans certains cas, les résultats d'essai exprimés sous forme de catégories sont basés sur des valeurs numériques sous-jacentes qui ne peuvent pas être considérées comme régulièrement réparties mais qui peuvent être déterminées par d'autres méthodes d'essai. Les méthodes statistiques habituelles ne peuvent alors pas être appliquées à leur classement.

Exemple : Niveau de pollution par un agent polluant. Dans ce cas, une autre méthode d'essai peut fournir des valeurs numériques de la présence (par exemple en g/cm³) de cet agent polluant.

Dans ce cas, il est possible de tracer des courbes de probabilités pour chaque résultat d'essai en fonction de la valeur numérique sous-jacente (voir Figure 4) et d'en déduire des incertitudes sur les résultats d'essai exprimés sous forme de catégories dans le même format (c'est-à-dire des probabilités d'être dans une catégorie donnée). Cependant, cela demande de déterminer ces courbes de probabilités, ce qui représente une grande quantité de travail.

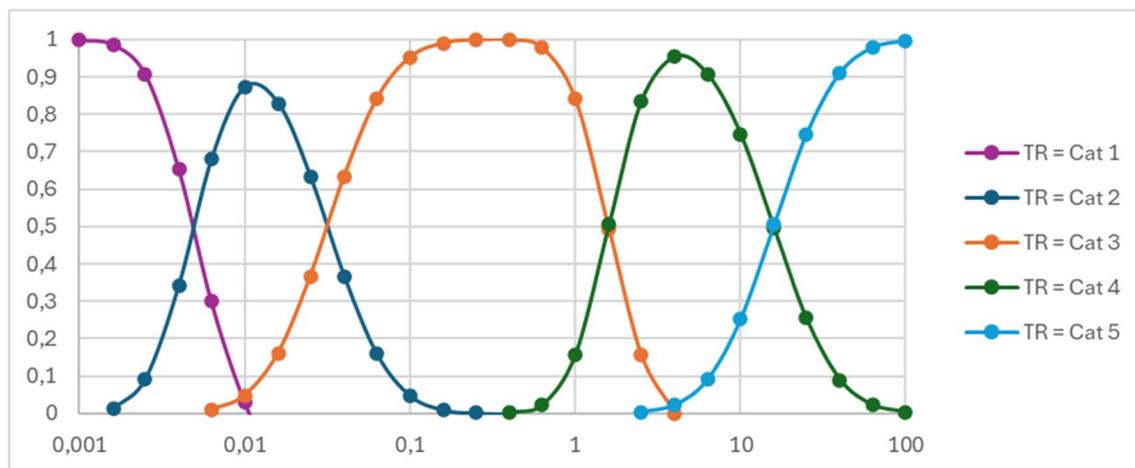


Figure 4 : Probabilité pour un échantillon d'être classé dans une catégorie parmi 5 catégories ordonnées, en fonction de la valeur numérique sous-jacente V.

Exemple : si $V = 10$, 75 % des résultats de test (TR) sont de catégorie 4 et 25 % de catégorie 5.

Utilisation d'indices basés sur les rangs des catégories :

La norme ISO 13528 [3] fournit des informations sur la manière d'utiliser les indices de Gower pour transformer un ensemble de résultats exprimés sous forme de catégories ordonnées en un ensemble de résultats exprimés sous forme de valeurs numériques. Ces transformations sont basées sur des équations telles que l'équation (9) (d'autres équations de forme similaire peuvent être utilisées), comme suit :

$$NTR = \frac{i - \bar{i}}{N} \quad (9)$$

Où NTR est le résultat du test transformé en format numérique,
i est le rang de la catégorie,

\bar{i} est le rang moyen des résultats du test, généralement calculé comme le rang du résultat médian du test,
 N est le nombre de catégories dans lesquelles les résultats du test sont répartis.

Exemple : Si un ensemble de résultats de test est [1,5 - 2 - 2 - 2,5 - 2,5 - 2,5 - 3 - 4] parmi des catégories qui peuvent être [0 - 0,5 - 1 - 1,5 - 2 - 2,5 - 3 - 3,5 - 4 - 4,5 - 5], soit 11 possibilités différentes, alors $\bar{i} = 6$ (rang de 2,5 parmi les résultats possibles du test), $N = 11$, NTR sont respectivement $NTR = (4 - 6)/11 = -0,182$ pour $TR = 1,5$, $NTR = (5 - 6)/11 = -0,091$ pour $TR = 2$, $NTR = 0$ pour $TR = 2,5$, $NTR = 0,091$ pour $TR = 3$ et $NTR = 0,273$ pour $TR = 4$.

Par construction, l'équation (9) produit des résultats NTR appartenant à l'intervalle $]-1;+1[$. La norme ISO 13528 [3] indique que des statistiques paramétriques peuvent être appliqués à ces types de valeurs. Il convient de noter que les limites calculées pour définir les incertitudes se situent généralement entre deux catégories. Elles doivent alors être retransformées en catégories exprimées en pourcentage de catégories, comme le montre l'exemple ci-dessous.

Exemple : Si l'ensemble des résultats d'essai [1,5 - 2 - 2 - 2,5 - 2,5 - 2,5 - 3 - 4] est le résultat d'une expérience partielle de la méthode A du GUM telle qu'exposée au § 8.3, la valeur moyenne est alors 0, l'écart type est 0,137. Pour revenir aux catégories, la valeur moyenne est $11 \times 0 + 2,5 = 2,5$, l'écart-type est $0,137 \times 11 = 1,51$, soit 151% d'une catégorie, la limite inférieure de l'IC95% est la catégorie de rang 6 moins $2 \times 151\%$ d'une catégorie, soit 2% de la catégorie de rang 2 et 98% de la catégorie de rang 3, c'est-à-dire pas plus de 2% de résultats "0,5" et aucun résultat "0". De même, la limite supérieure de l'IC95% est pas plus de 2% des résultats "4,5" et aucun résultat "5".

Lorsque les catégories ne sont pas ordonnées, la norme ISO 13528 [3] propose de les classer de la plus fréquente à la moins fréquente et de calculer les indices de Gower comme décrit ci-dessus.

Exemple : Si un ensemble de résultats de tests est [A - L - L - F - F - F - K] parmi des catégories qui peuvent être [A - B - D - K - L - M], c'est-à-dire 6 possibilités différentes, alors la catégorie F correspond au rang 1, la catégorie L au rang 2, les catégories A et K au rang 3 et les autres catégories au rang 4. La valeur moyenne est la catégorie "L" de rang 2.

Il est alors possible d'appliquer l'équation (9) à ces résultats.

Exemple : Si un ensemble de résultats de test [A - L - L - F - F - F - K] est le résultat d'une expérience partielle selon la méthode A du GUM telle qu'exposée au § 8.3, la valeur moyenne est alors de -0,024, soit 86% des résultats "L" et 14% des résultats "F", et l'écart-type est de 0,154, soit $0,154 \times 6 = 93\%$ d'une catégorie. De la même manière que précédemment, l'IC95% contient les catégories "F" et "L", pas plus de 86% des résultats "A" et "K" ensemble et aucun résultat "B" ou "M".

Il convient de noter que ces indices de Gower sont des calculs sur les rangs des catégories et que d'autres méthodes non paramétriques également basées sur les rangs peuvent être utilisées. Un grand nombre de normes ISO traitent de ces méthodes statistiques non paramétriques.

Statistiques s'appliquant aux proportions :

Un grand nombre de méthodes statistiques ont été développées pour calculer les intervalles de confiance des proportions, en particulier pour les sondages. Ces statistiques peuvent rarement être utilisées car elles requièrent normalement une grande quantité de résultats de tests. Mais lorsqu'une grande quantité de résultats est disponible, ces méthodes fournissent de meilleurs résultats que les indices de Gower. Des informations sur ces méthodes sont disponibles dans les normes ISO (en particulier ISO 16269-6 [15]).

10.1.3 Résultats binaires

Les résultats binaires peuvent être considérés comme un cas particulier de catégories avec un nombre de catégories $N = 2$. Les propositions du § 10.1.2 peuvent alors être appliquées.

Une autre façon de traiter les résultats binaires consiste à utiliser la loi statistique binomiale pour déterminer les probabilités associées à chaque résultat de test. Une grande quantité de littérature scientifique et de normes dérivées de la loi binomiale existe sur le problème des décisions erronées (risques d'acceptation erronée et risques de rejet erroné).

De la même manière que ci-dessus, lorsqu'un grand nombre de résultats d'essais est disponible, les méthodes traitant des IC sur les proportions sont recommandées.

10.1.4 Résultats exprimés en pourcentages de catégories

Dans ces méthodes, les résultats sont exprimés en pourcentage des catégories.

Exemple : Le résultat du test est exprimé comme suit : "A" : 25%, "B" : 40%, "C" : 35%.

Dans ce cas, *TR* doit être traduit en deux chiffres :

- ✚ L'un exprimant la valeur centrale ;
- ✚ L'autre exprimant la dispersion habituelle à laquelle on peut s'attendre pour *TR*.

Exemple : Si le résultat du test est exprimé comme suit : "A" : 25%, "B" : 40%, "C" : 35 %, à partir des catégories ordonnées (A<B<C), la valeur centrale peut être calculée comme la médiane (c'est-à-dire "B") ou comme une valeur moyenne utilisant les indices de Gower (c'est-à-dire 93 % "B" et 7 % "C"). La dispersion peut être calculée à partir de l'écart-type (0,301 ou 30 % d'une catégorie).

Les techniques habituelles concernant les valeurs moyennes et les écarts types peuvent ensuite être utilisées pour calculer l'IC.

10.1.5 Résultats exprimés sous forme de courbes

Dans la plupart de ces cas, les documents de référence qui décrivent les méthodes d'essai fournissent des exigences détaillées sur la manière de déterminer des valeurs numériques à partir des courbes.

Exemple : Les essais de traction consistent essentiellement à tracer le diagramme force - allongement enregistré au cours de l'essai. La norme ISO 6892-1 (essais de traction sur métaux) définit ensuite comment les paramètres typiques (limite d'élasticité, résistance à la rupture, allongement à la force maximale, allongement à la rupture, ...) doivent être déterminés à partir de ce diagramme.

Mais il arrive que la courbe elle-même représente un résultat d'essai. Dans ce cas, afin de déterminer les incertitudes associées, le laboratoire peut :

- ✚ Définir les paramètres numériques qui caractérisent le diagramme ;

Exemple : Sur les diagrammes de spectrométrie IR utilisés pour l'identification des peintures ou des plastiques, déterminer le nombre d'onde des 10 pics principaux).

- ✚ Définir les paramètres non numériques qui caractérisent le diagramme ;

Exemple : Sur les diagrammes de spectrométrie IR utilisés pour l'identification des peintures ou des plastiques, déterminer l'intensité des 10 pics principaux, qualifiés de "Important", "Moyen", "Faible". Il est également possible de les classer du plus intense au moins intense.

- ✚ Lorsque le diagramme est destiné à être comparé à un diagramme de référence, considérer qu'il s'agit d'un résultat de test binaire.

Exemple : Pour les diagrammes de spectrométrie IR utilisés pour l'identification des peintures ou des plastiques, déterminer le risque de décision erronée comme décrit au § 10.1.3.

10.2 Changement de variables

Certaines des équations utilisées pour déterminer les incertitudes (généralement l'équation (8)) ne requièrent pas la normalité pour la distribution des résultats des essais, alors que d'autres si (pour la détermination d'un IC, que ce soit en utilisant un coefficient d'élargissement, que ce soit lors de la détermination d'un IC sur un ET comme au § 10.7). En outre, la détermination d'un ET peut être considérablement affectée en cas d'asymétrie importante.

Dans tous les cas, les déterminations sont plus fiables lorsque les distributions des données sont au moins symétriques. Les situations typiques où ce n'est pas le cas sont lorsque la dispersion des résultats est importante et :

- ✚ Une limite physique ou technique existent pour les résultats des tests ;
- ✚ Les résultats des tests sont une proportion, ce qui implique deux limites techniques, à savoir 0% et 100%.

Cas où une limite technique est présente :

Dans de nombreux cas, les résultats des essais ne peuvent pas être négatifs pour des raisons techniques. Parfois, d'autres limites s'appliquent (par exemple, dans les essais de traction, d'après leurs définitions, le rapport R_m/R_e ne peut être inférieur à 1 et A_{gt} ne peut être inférieur à R_m/E). Ces limites peuvent en fait être des minima ou des maxima. Dans ce cas :

- ✚ Lorsque la probabilité que certains résultats d'essais soient proches de la limite est faible, la distribution des résultats d'essais n'est pas modifiée de manière significative, et l'on peut supposer que les distributions des résultats d'essais sont gaussiennes ;
- ✚ Dans le cas contraire, une asymétrie est susceptible de se produire, ce qui peut affecter de manière significative les déterminations des incertitudes. Lorsque la limite inférieure des résultats d'essai est 0, ce problème devrait être pris en compte lorsque le CoV (coefficient de variation) est inférieur à 0,15.

La transformation des résultats d'essais TR en $tr = \log(TR)$ est un moyen très efficace de résoudre un problème d'asymétrie. Les calculs sont ensuite effectués sur les valeurs tr et les résultats doivent ensuite être retransformés dans l'échelle initiale. Ceci peut être réalisé en utilisant la rétro-transformation $TR = 10^{tr}$ (lorsque des logarithmes de base 10 sont utilisés). Cela permet d'éviter les situations absurdes où les limites inférieures de l'IC sont des valeurs impossibles.

Exemple d'une distribution asymétrique dont la valeur moyenne est $\mu = 100$ et l'écart type est $\sigma = 40$ (CoV = 0,4). Après transformation logarithmique, la valeur moyenne est $\mu = 2$ et l'écart type est $\sigma = 0,2$ (estimation meilleure car moins affectée par les valeurs aberrantes de la partie supérieure de la distribution). L'IC95% est alors $[10^{2-2,0,2}; 10^{2+2,0,2}]$, soit $[40; 250]$ à comparer à $[20; 180]$ sans transformation.

Cas où les limites de 0 % et de 100 % sont présentes :

Le même problème peut se poser dans certains cas de proportions pour lesquelles la dispersion est importante ou lorsque l'incertitude dépend du niveau du résultat du test.

Dans ces cas, les transformations utilisant les équations $tr = \log(TR/(1 - TR))$ (ou $tr = \log(TR/(100 - TR))$ si TR est exprimé en pourcentage) ou $tr = (TR - 0,5)/\sqrt{TR \cdot (1 - TR)}$ (ou $tr = (TR - 50)/\sqrt{TR \cdot (100 - TR)}$ si TR est exprimé en pourcentage) peut être efficace pour :

- ✚ Transformer des distributions asymétriques en distributions symétriques ;
- ✚ Rendre u indépendant du niveau du mesurande.

Exemple : La détermination de la granulométrie d'un ciment nécessite de déterminer la courbe des pourcentages masses des particules en fonction de leur taille. En pratique, ceci est réalisé pour une série de tailles comprises entre 1,25 μm et 100 μm dont la répartition suit les recommandations de la norme ISO 3 [13]. L'utilisation d'une des transformations supérieures proposées permet de :

- Rendre u indépendant de la taille des particules ;
- Augmenter le nombre de résultats d'essais disponibles pour déterminer les incertitudes ;
- Déterminer des incertitudes fiables pour les extrémités de la distribution, c'est-à-dire pour les tailles de particules inférieures et supérieures.

10.3 Questions liées à la résolution et à l'arrondissement des résultats d'essai

Il est rappelé que la résolution est l'écart minimal qui peut se produire entre deux mesures ou résultats d'essai. Dans la plupart des cas, elle a pour effet, tout comme l'arrondissement, de réduire l'estimation de l'écart type correspondant.

Exemple : Si un ensemble de 7 estimations d'un résultat d'essai dont la valeur réelle est 3,141592653578789... et dont l'écart-type réel de l'estimation est 0,01, l'écart-type des estimations arrondi à 1 ($r = 1$), 0,1 ($r = 0,1$), 0,01 ($r = 0,01$) et 0,001 ($r = 0,001$) le plus proche se répartit comme montré dans la Figure 5.

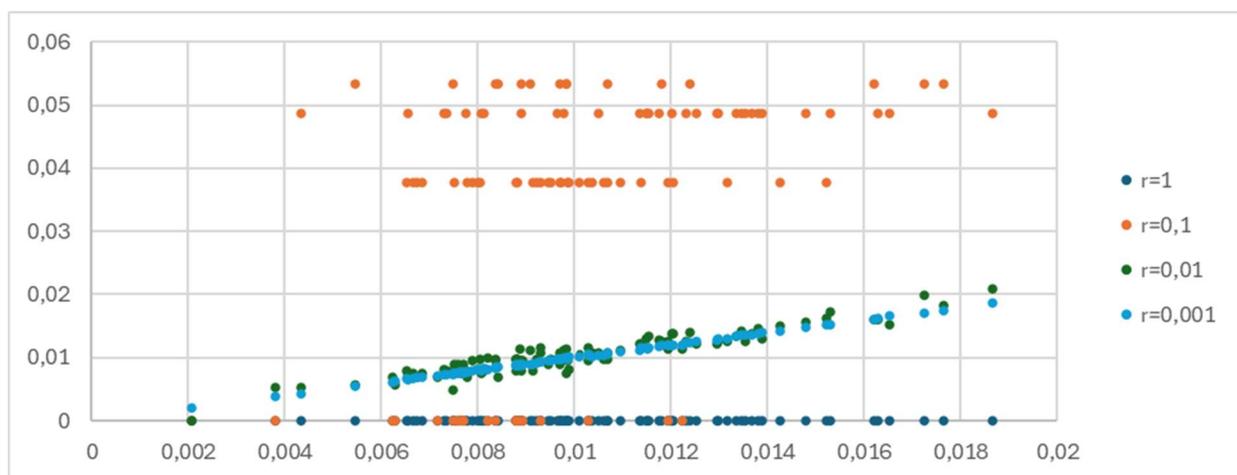


Figure 5 : Distributions des rapports s/s_{true} (rapports de s calculés à partir des valeurs arrondies sur s calculés à partir des valeurs non arrondies)

Cet exemple montre que :

- Lorsque $r = 1$ (100 fois l'écart-type réel de l'estimation), l'écart-type de l'estimation est toujours réduit à 0 ;
- Lorsque $r = 0,1$ (10 fois l'écart-type réel de l'estimation), les écarts-types des estimations se répartissent sur un très petit nombre d'occurrences (c'est-à-dire, dans ce cas, 0 - 3,8 - 4,9 - 5,3 fois l'écart-type réel), c'est-à-dire une très mauvaise qualité d'estimation ;
- Lorsque $r = 0,01$ (1 fois l'écart-type réel de l'estimation), l'IC95% de l'écart-type de l'estimation est de $]0,8.s;1,34.s[$, c'est-à-dire une estimation acceptable mais significativement affectée ;
- Lorsque $r = 0,001$ (0,1 fois l'écart-type réel de l'estimation), l'IC95% de l'écart-type de l'estimation est $]0,97.s;1,03.s[$, c'est-à-dire que l'estimation de l'écart-type n'est pas affectée par l'arrondi ou la résolution.

Cela confirme la règle habituelle selon laquelle la résolution et l'arrondi doivent être au moins égal à 1/10 de l'écart-type réel. Ceci est également cohérent avec les conclusions du § 10.8.

Il convient de noter que lorsque r est du même ordre que l'écart type de l'estimation, le comportement des rapports s/s_{true} dépend du nombre d'estimations incluses dans l'ensemble et de l'endroit où la valeur non arrondie est située dans le pas d'arrondissement.

Dans l'exemple ci-dessus, lorsque $r = 0,01$ (1 fois l'écart-type réel de l'estimation), les écarts-types des estimations se répartissent sur 4 valeurs différentes, correspondant aux cas où :

1. Toutes les estimations sont arrondies à la même valeur ;
2. Une estimation est arrondie d'un côté du pas et six de l'autre côté ;
3. Deux estimations sont arrondies d'un côté du pas et cinq de l'autre côté ;
4. Trois estimations sont arrondies d'un côté du pas et quatre de l'autre côté.

c'est-à-dire $n/2+1$ ou $(n+1)/2$ occurrences, selon que n (nombre d'estimations incluses dans l'ensemble) est un nombre pair ou impair.

En outre, les valeurs réelles de s/s_{true} (c'est-à-dire, dans cet exemple, 0 - 3,8 - 4,9 - 5,3 fois l'écart type réel), dépendent de l'endroit où la valeur non arrondie est située dans le pas (c'est-à-dire 0,41592653578789... fois la valeur de l'échelon dans l'exemple). Si la valeur réelle avait été proche d'une valeur arrondie, les valeurs réelles de s/s_{true} auraient été proches de 0.

Il convient aussi de noter que le niveau d'arrondi ou de résolution peut être masqué par un calcul ultérieur. Par conséquent, il convient d'être attentif à ces éventualités.

Exemple : Pour la détermination des résistances lors des essais de traction sur les métaux conformément à la norme ISO 6892-1, la résolution du résultat de l'essai est principalement régie par la résolution du capteur de force. Si la plage du capteur est de 628 kN et que le nombre d'incrémentes est de 24575, la force mesurée est arrondie à la valeur la plus proche de 25,5544 N, ce qui correspond à la résolution. D'autre part, la section est mesurée avec des dimensions arrondies au 0,005 mm le plus proche, ce qui génère également une résolution cachée pour la mesure de la section (par exemple 0,03142 mm²). Le résultat de l'essai (division des deux mesurandes) est alors également marqué par une résolution, qui peut être irrégulière.

Lorsqu'un tel rapport n'est pas atteint, la résolution ou l'arrondi doit être amélioré. Cela peut être réalisé par :

- ✚ L'augmentation de la valeur de la base de mesure ;
- ✚ L'amélioration de la précision du dispositif de mesure.

Pendant, dans certains cas, pour des raisons techniques, il n'est pas possible d'augmenter la résolution. Dans ce cas, une contribution de la résolution doit être ajoutée à l'incertitude. Le GUM [2] fournit les conseils nécessaires pour y parvenir.

Dans la plupart des cas, il s'agit d'ajouter une contribution $u=a/\sqrt{3}$, où a est l'incrément. Ceci est valable lorsque la valeur réelle est distribuée aléatoirement entre les 2 limites de l'incrément (i.e. distribution carrée).

10.4 Cas où plusieurs valeurs d'incertitudes sont disponibles pour plusieurs situations différentes

Il peut arriver que plusieurs estimations d'incertitudes provenant de différentes sources soient disponibles dans le laboratoire, chacune d'entre elles étant liée à différents ensembles de conditions (matériel, équipement d'essai, opérateurs, ...).

Par exemple, cette situation peut se produire lorsque le laboratoire réalise plusieurs cartes de contrôle, chacune d'entre elles étant liée à une machine d'essai.

Dans ce cas, le laboratoire a deux possibilités :

- ✚ Calculer autant de valeurs de U que le nombre d'ensembles de données disponibles. Le laboratoire dispose alors d'informations plus détaillées, mais il doit gérer un plus grand nombre de données lorsque, par exemple, un client demande la valeur U associée à ses résultats d'essai ;
- ✚ Consolider les valeurs U obtenues pour chaque série de résultats d'essais en une valeur U générale, valable pour tous les résultats d'essais produits par le laboratoire. Dans ce cas, il convient d'utiliser une valeur moyenne quadratique pondérée de U , comme indiqué dans l'équation (10).

$$U = \left| \frac{c_1 \cdot B_1 + \dots + c_i \cdot B_i + \dots + c_n \cdot B_n}{c_1 + \dots + c_i + \dots + c_n} \right| + k \cdot \sqrt{\frac{c_1 \cdot u_1^2 + \dots + c_i \cdot u_i^2 + \dots + c_n \cdot u_n^2}{c_1 + \dots + c_i + \dots + c_n}} \quad (10)$$

où B_i est le biais (c'est-à-dire $m_i - X_{RM}$) du $i^{\text{ème}}$ ensemble de données
 u_i est l'incertitude type de la $i^{\text{ème}}$ série de données,
 c_i est un coefficient de pondération pour le $i^{\text{ème}}$ ensemble de données,
 n est le nombre d'ensembles de données utilisés pour déterminer u_i ,
 k est le coefficient d'élargissement.

Il convient de noter que :

- ✚ Le signe de chaque B_i doit être conservé dans les calculs intermédiaires et la valeur absolue doit être appliquée à la somme de toutes les valeurs positives et négatives de B_i ;
- ✚ Les coefficients c_i sont destinés à rendre la valeur U globale plus représentative de la quantité totale d'incertitudes qui s'applique à la production globale de résultats d'essais du laboratoire. Ils doivent donc être choisis en fonction de la proportion de résultats d'essais provenant de chacune des situations représentées par chacune des incertitudes standard individuelles u_i . Il est possible de gérer les coefficients c_i de sorte que leur somme soit égale à 1 : cela simplifie l'équation (10) et montre plus clairement quelles sont les situations qui contribuent le plus aux incertitudes.

10.5 Questions générales concernant l'estimation

Les règles d'estimation s'appliquent lorsque la valeur d'un paramètre est calculée à partir d'une sélection d'occurrences de ce paramètre.

Exemple : lorsque la valeur moyenne d'une population est calculée à partir de 10 valeurs, appartenant toutes à cette population.

C'est le plus souvent le cas dans la détermination des incertitudes, mais pas toujours. Lorsque la totalité des occurrences possibles est incluse dans l'expérience, le paramètre considéré n'est alors plus estimé, mais calculé. En particulier, la formule pour déterminer les écarts types n'est pas la même dans les deux cas : l'équation (11) lorsque toutes les occurrences sont prises en compte dans le calcul et l'équation (12) lorsqu'une sélection d'occurrences est prise en compte dans le calcul.

$$\sigma = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}} \quad (11)$$

$$s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n - 1}} \quad (12)$$

où σ est l'écart-type à calculer,
 s est l'estimation d'un écart type σ à calculer,
 x_i sont les valeurs individuelles de l'ensemble utilisé pour calculer l'écart-type,
 \bar{x} est la valeur moyenne des x_i ,
 et n est le nombre de valeurs utilisées pour calculer la valeur moyenne.

Exemple 1 : Lorsqu'une expérience partielle selon la méthode GUM A est organisée pour déterminer la contribution des opérateurs et que cette expérience inclut tous les opérateurs du laboratoire, et que chaque valeur moyenne des opérateurs est supposée être connue avec précision, l'équation (11) s'applique et la valeur de l'écart type est connue exactement. Par conséquent, les recommandations du § 10.7 ne s'appliquent pas à ce cas.

Exemple 2 : Dans cette même expérience, chaque opérateur pourrait potentiellement répéter les tests un nombre infini de fois. L'ensemble des valeurs utilisées pour calculer l'écart-type de répétabilité est alors une sélection de toutes les réalisations possibles. Par conséquent, l'équation (12) s'applique et la valeur de l'écart type n'est pas connue avec exactitude. Les recommandations du § 10.7 peuvent être utilisées pour déterminer l'IC95% de la vraie valeur σ .

10.6 Questions concernant l'estimation d'une valeur moyenne

La loi de distribution des estimations des valeurs moyennes est décrite dans l'équation (13) (origine : ISO 2854 [16]).

$$m \approx N\left(\mu, \frac{\sigma}{\sqrt{n}}\right) \quad (13)$$

où m est l'estimation d'une valeur moyenne,
 μ est la valeur moyenne à estimer,
 σ est l'écart-type de la population correspondante,
 et n est le nombre de valeurs utilisées pour calculer la valeur moyenne.

Lorsque μ et σ sont tous deux inconnus, les tables de Student peuvent être utilisées pour prendre en compte l'incertitude liée aux estimations de ces paramètres.

10.7 Questions relatives à l'estimation d'un écart type

La loi de distribution des estimations des variances est décrite dans l'équation (14) (origine : ISO 2854 [16]).

$$(n-1) \frac{s^2}{\sigma^2} \approx \chi_{n-1}^2 \quad (14)$$

où s est l'estimation d'un écart-type,
 σ est l'écart-type à estimer,
 et n est le nombre de valeurs utilisées pour calculer l'écart-type.

La loi de distribution des estimations des écarts types peut alors être déduite par transformation algébrique de l'équation (14) comme indiqué dans l'équation (15).

$$s = \sigma \sqrt{\frac{\chi_{n-1}^2(P)}{n-1}} \quad (15)$$

où s est l'estimation d'un écart-type,
 σ est l'écart-type à estimer,

*P est la probabilité cumulée théorique correspondante,
et n est le nombre de valeurs utilisées pour calculer l'écart-type.*

Dans la pratique, s est connu (déterminé à partir des expériences) et l'information intéressante est la valeur de σ . Nous pouvons alors calculer les limites d'un IC95% en calculant les rapports σ/s pour $P = 0,025$ et $P = 0,975$. Ces limites sont présentées dans le Tableau 3 pour des valeurs croissantes de $\nu = n-1$ (nombre de degrés de liberté).

Tableau 3. Limites inférieures et supérieures de l'intervalle IC95% des rapports σ/s en fonction de $\nu = n-1$.

2	3	4	5	6	8	10	13	16	20
0,521	0,566	0,599	0,624	0,644	0,675	0,699	0,725	0,745	0,765
6,285	3,729	2,874	2,453	2,202	1,916	1,755	1,611	1,522	1,444
25	32	40	50	63	80	100	125	160	200
0,784	0,804	0,821	0,837	0,852	0,866	0,879	0,89	0,901	0,911
1,38	1,323	1,28	1,243	1,211	1,183	1,161	1,141	1,123	1,109

On constate qu'un grand nombre de résultats d'essais est nécessaire pour obtenir une estimation acceptable de σ à partir des valeurs disponibles de s .

Lorsque l'écart-type est estimé à partir d'une série de répétitions (par exemple dans l'équation (7)), le même tableau s'applique en remplaçant $n-1$ par $n.(r-1)$, où n est le nombre de séries et r le nombre de répétitions.

Par exemple, dans une expérience telle que décrite au § 7.3, si 50 configurations sont testées avec une répétition pour chacune, l'estimation de la répétabilité calculée avec l'équation (7) a obtenu un IC95% comme indiqué dans le Tableau 3 avec $\nu = 50.(2-1) = 50$, c'est-à-dire [0,837;1,243].

Cette augmentation des valeurs ν doit être prise en compte lors de la conception de ces expériences.

10.8 Questions relatives à la combinaison des écarts types

Dans de nombreux cas, une combinaison d'écarts types est appliquée (voir les équations (3), (4), (5), (6), et (8)). Il convient de noter que ces équations de combinaison d'écart-type sont valables même lorsque les distributions correspondantes ne sont pas gaussiennes.

Il faut cependant noter que les écarts types inclus dans la formule doivent être indépendants. Si ce n'est pas le cas, un terme de corrélation entre eux doit être introduit.

D'un point de vue mathématique, l'existence d'une corrélation est prouvée lorsque r (coefficient de corrélation) est significativement différent de 0, mais la réciproque n'est pas vraie (c'est-à-dire qu'une corrélation peut être présente même si r n'est pas significativement différent de 0).

Cette dernière situation est notamment susceptible de se produire lorsque le nombre de couples introduits dans la corrélation est insuffisant ou lorsque la corrélation n'est pas linéaire (par exemple, une hypothétique corrélation en $x^2+y^2=1$ produirait des coefficients r toujours proches de 0).

Dans la pratique, lorsqu'une corrélation a des raisons techniques d'être présente, il est généralement plus facile de déterminer une valeur u globale qui englobe les deux effets et leur corrélation.

Par exemple, si $u_{total} = \sqrt{1^2 + 0,5^2 + 0,2^2 + 0,1^2 + 0,05^2}$, mais que u_1 et u_2 sont en fait positivement corrélés ($r = 1$), l'équation doit être transformée en $u_{total} = \sqrt{(1 + 0,5)^2 + 0,2^2 + 0,1^2 + 0,05^2}$, et si u_1 et u_2 sont en fait négativement

corrélés ($r = -1$), l'équation doit être transformée en $u_{total} = \sqrt{(1 - 0,5)^2 + 0,2^2 + 0,1^2 + 0,05^2}$. Les résultats u_{total} correspondants sont alors respectivement 0,55 et 1,517 au lieu de 1,141.

Dans la plupart des cas, les corrélations ne sont jamais complètes et des équations adéquates en fonction de r doivent être utilisées. Lorsque seules deux variables sont corrélées avec un coefficient de corrélation r , la variance globale de ces deux variables est $\sigma^2 = \sigma_1^2 + 2 \cdot r \cdot \sigma_1 \cdot \sigma_2 + \sigma_2^2$.

Les principales contributions introduites dans la formule déterminent en grande partie le résultat final, comme le montre l'exemple ci-dessous, où les contributions sont classées de la plus importante à la moins importante.

$u_{total} = \sqrt{1^2 + 0,5^2 + 0,2^2 + 0,1^2 + 0,05^2} = 1,14127$, mais si l'on omet la contribution la moins importante,

$u_{total} = \sqrt{1^2 + 0,5^2 + 0,2^2 + 0,1^2} = 1,14018$,

$u_{total} = \sqrt{1^2 + 0,5^2 + 0,2^2} = 1,13578$,

$u_{total} = \sqrt{1^2 + 0,5^2} = 1,11803$,

$u_{total} = \sqrt{1^2} = 1$

Nous pouvons constater que seuls les deux premiers termes sont importants.

Cependant, en accord avec le § 10.7 ci-dessus, l'incertitude sur les estimations de u est généralement comprise entre 1 et 4 (lorsqu'elle est calculée à partir de 6 résultats d'essais ou moins), et l'importance réelle mutuelle des deux premiers termes est incertaine. Une expérience de Monte-Carlo a été réalisée pour résoudre ce problème.

Cette expérience a été menée sur un exemple où $u_{total} = \sqrt{1^2 + 0,5^2 + 0,2^2 + 0,1^2 + 0,05^2}$, chacun de ces u_i étant déterminé comme des écarts-types calculés à partir de n résultats. Les rapports σ/s résultants en fonction de n (nombre de répétitions utilisées pour calculer s) et N_t (nombre de termes pris dans u_{total} ci-dessus) sont affichés dans la Figure 6. Les courbes centrales sont des moyennes, les courbes supérieures et inférieures sont des limites d'IC95%. Les résultats détaillés sont disponibles au § 10.11.

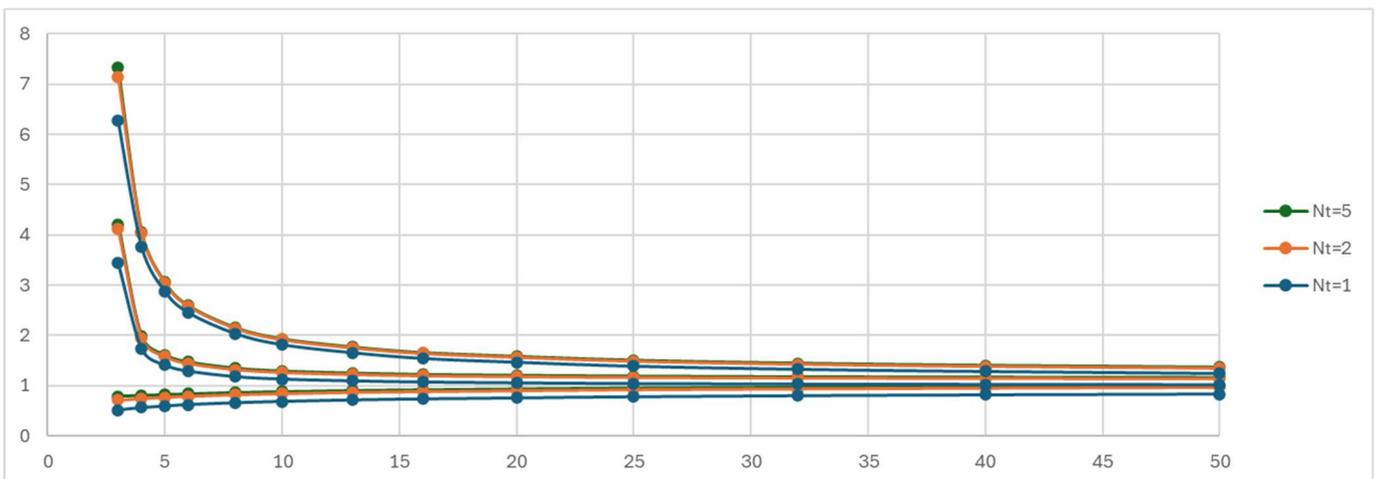


Figure 6 : Rapports σ/s en fonction de n (nombre de répétitions utilisées pour calculer s) et N_t (nombre de termes pris dans u_{total} ci-dessus), Les courbes centrales sont des moyennes, les courbes supérieures et inférieures sont des limites d'IC95%.

Il faut noter que, dans la réalité, les effets d'arrondi ont tendance à réduire la valeur estimée de s , c'est-à-dire à augmenter les rapports σ/s , ce qui amplifie les phénomènes observés ici.

Cette expérience permet de constater que :

- ✚ Seuls les 2 premiers termes sont importants (les courbes $N_t = 2$ et $N_t = 5$ coïncident presque) ;
- ✚ Les distances entre les courbes moyennes et les courbes limites IC95% sont bien plus grandes que les distances entre les courbes $N_t = 1$ et $N_t = 2$. Cependant, des résultats non fournis montrent que le centile 97,5% du rapport du 2ème terme au 1er terme σ_2/σ_1 est supérieur à 1 lorsque $n < 10$. Il convient alors de déterminer avec précision tous les termes qui représentent au moins 50% du terme principal (ici les 2 termes principaux) ;
- ✚ En raison de la forte asymétrie des distributions des rapports σ/s , les valeurs moyennes de ceux-ci sont significativement supérieures à 1, même pour de grandes valeurs de n . Cela semble être une autre cause majeure de sous-estimation des valeurs de u .

Par exemple, lorsque la détermination de u_{total} à partir de 5 termes, chacun d'entre eux provenant de calculs d'ET à partir de 3 résultats d'essais, conduit à $u = 1,413$, la vraie valeur moyenne de celle-ci est en fait de 4,21 et les limites d'IC95% pour celle-ci sont de 0,80 et 7,93.

Par conséquent, lorsque la détermination de U nécessite un grand nombre d'expériences, une procédure en 2 étapes peut s'avérer utile, comme suit :

1. Réaliser une étude préliminaire des contributions de chaque source d'incertitude;
2. Réaliser une étude détaillée des contributions des sources qui représentent plus de 50% de la source principale.

10.9 Impact du choix de la classification d'une source d'incertitude comme biais ou erreur aléatoire

Impact sur u et U de la classification d'un biais en tant qu'erreur aléatoire :

Lorsqu'un biais est déterminé et calculé comme tel, la valeur U est calculée à l'aide de l'équation (3), (4) ou (6). Lorsqu'il est considéré comme une erreur aléatoire, la valeur U est calculée à l'aide de l'équation (8). Lorsque $e \ll B$, cela entraîne une différence significative dans le calcul de U (le rapport tend vers la valeur de k), comme illustré ci-dessous (exemple avec $B = 1$, $e = 0,1$ et $k = 2$) :

- ✚ Option avec B considérée comme un biais : $U = 1 + 2 \times 0,1 = 1,2$;
- ✚ Option avec B incluse dans l'erreur aléatoire : $U = 2 \cdot \sqrt{1^2 + 0,1^2} = 2,01$;

Dans les autres cas, des différences importantes peuvent apparaître pour u , mais pas pour U , comme on le voit ci-après (exemple avec $B = 1$, $e = 1$ et $k = 2$) :

- ✚ Option avec B considéré comme un biais : $u = 1$ et $U = 1 + 2 \times 1 = 3$;
- ✚ Option avec B inclus dans l'erreur aléatoire : $u = \sqrt{1^2 + 1^2} = 1,4$ et $U = 2 \times 1,4 = 2,8$.

Impact sur l'IC95% de la classification d'un biais en tant qu'erreur aléatoire :

Comme nous l'avons vu précédemment, le GUM [2] propose d'utiliser un coefficient d'élargissement k pour déterminer un IC, en supposant que la distribution globale des résultats d'essais est gaussienne, ce qui est justifié par le théorème de la limite centrale des statistiques, et souvent correct d'un point de vue pratique. k est choisi en fonction de l'IC souhaité (par exemple, $k = 2$ pour l'IC95%).

Cependant, ceci ne s'applique pas correctement lorsque $e \ll B$. Dans ces cas :

- ✚ La valeur de U est mal déterminée si le biais n'est pas déterminé séparément (voir ci-dessus) ;

- ✚ L'effet de l'erreur aléatoire n'est délétère que d'un côté de sa distribution (voir Figure 3). La valeur k devrait alors être choisie pour des valeurs P unilatérales.

Il s'ensuit que :

- ✚ Lorsque $e \ll B$, k doit être choisi égal à 1,64 (IC95% unilatéral) de sorte que, dans le premier exemple, le vrai $U = 1 + 1,64 \times 0,1 = 1,16$ au lieu de 1,2 ;
- ✚ Dans les autres cas, le choix de $k = 2$ est correct, de sorte que les calculs ci-dessus ne sont pas modifiés.

Conclusions :

Lorsque $e \ll B$, l'inclusion du biais B dans le terme aléatoire u conduit à une surestimation de U par le facteur k d'élargissement. Dans les autres cas, l'inclusion du biais B dans le terme aléatoire u conduit à sous-estimer u , mais cette sous-estimation est plus ou moins compensée par l'application du coefficient d'élargissement lors du calcul de l'incertitude élargie U .

10.10 Résultats d'essais pour des conditions environnementales données (typiquement, pour des températures données) ou, plus généralement, en fonction de conditions extérieures

Il arrive fréquemment que des résultats d'essais soient demandés pour des conditions environnementales qui ne sont pas les conditions ambiantes. Dans ce cas, le laboratoire a le choix entre deux options :

- ✚ Inclure la contribution de l'écart par rapport aux conditions environnementales requises dans la détermination globale de u ;
- ✚ Fournir deux valeurs distinctes de u , l'une pour l'écart par rapport aux conditions environnementales demandées et l'autre pour l'incertitude sur le résultat de l'essai lui-même, sans y inclure l'impact de l'écart aux conditions environnementales requises.

La deuxième option est plus facile, car elle ne demande pas de déterminer l'impact de l'écart des conditions environnementales sur le résultat de l'essai (ce qui peut parfois être difficile à faire correctement). En fonction de l'utilisation prévue des résultats d'essais, la première ou la deuxième option peut être la meilleure.

Exemple : Les aciers au carbone peuvent devenir fragiles à une certaine température appelée température de transition, qui est déterminée par une série d'essais de résilience Charpy selon la norme ISO 148-1 en fonction de la température.

Lorsque les résultats des essais sont destinés à établir cette courbe de transition, il est préférable d'exprimer les incertitudes séparément. Les incertitudes peuvent alors être représentées sous forme d'ellipses sur les courbes et leurs valeurs sont plus fiables.

Lorsque les résultats des essais sont utilisés pour déclarer la conformité à une exigence (sous la forme $KV_2 > 30J$ à $-50^\circ C$), il est nécessaire d'inclure la contribution de l'incertitude sur la température d'essai dans l'incertitude globale afin d'assurer la conformité à l'exigence.

En conclusion, dans de tels cas, le laboratoire devrait choisir son option en fonction de l'utilisation prévue des incertitudes déclarées.

10.11 La méthode de Monte-Carlo

Les méthodes de Monte-Carlo constituent une vaste catégorie d'algorithmes qui utilisent des réalisations numériques aléatoires d'un modèle donné. Elles sont souvent utilisées pour résoudre des problèmes mathématiques ou physiques difficiles ou impossibles à résoudre par d'autres méthodes. Pour un aperçu de

l'histoire et des applications des méthodes de Monte-Carlo, voir par exemple[17]. De même, le document JCGM 101 [18] du BIPM fournit des recommandations concernant l'application de la méthode de Monte-Carlo à l'expression des incertitudes.

Des calculs difficiles sont nécessaires pour résoudre plusieurs des questions soulevées dans ce document. Afin de simplifier ces calculs, nous avons utilisé la méthode de Monte-Carlo. Dans le cadre de cette étude, la détermination des centiles, des médianes ou des valeurs moyennes des distributions nécessite de résoudre des intégrales difficiles et de trouver les zéros d'équations polynomiales. Pour éviter cela, de grandes séries de réalisations aléatoires sont créées, ce qui permet de calculer ces centiles, médianes ou valeurs moyennes.

Cependant, l'utilisation des méthodes de Monte-Carlo nécessite l'utilisation d'un modèle qui représente raisonnablement bien les situations que l'on souhaite traiter. Pour ce faire, une modélisation appropriée est nécessaire. Dans le cadre de cette étude, ces modèles sont fournis par les équations proposées dans ce document, en particulier les équations fondamentales (2) et (3).

L'utilisation des méthodes de Monte-Carlo nécessite également l'utilisation de valeurs d'entrée aléatoires. Lorsque plusieurs valeurs aléatoires sont nécessaires pour produire un résultat de Monte-Carlo et que des corrélations entre elles s'appliquent dans le phénomène à représenter, ces corrélations doivent être incorporées dans les valeurs d'entrée des calculs. Cela peut être un peu difficile à faire correctement.

Pour garantir la validité des conclusions, les séries aléatoires doivent être suffisamment nombreuses, en fonction de nombreux facteurs. Une solution pour contrôler cela est de diviser ces séries en sous-groupes. Cela nous permet de calculer la répétabilité des paramètres que nous déterminons. Cet écart-type de répétabilité est ensuite utilisé pour déterminer un IC pour chacune des déterminations, avec un coefficient d'élargissement égal à 2.

Exemple : Au § 10.8, une étude d'impact sur la combinaison de l'estimation de chacun des u_i a été réalisée. Cette étude peut difficilement être réalisée à l'aide de calculs mathématiques (elle aurait été totalement impossible si le nombre n de tests utilisés avait été différent pour chaque u_i , ce qui est la situation habituelle). Pour résoudre ce problème, la méthode de Monte-Carlo a été utilisée comme suit :

- Pour chaque terme u_i de l'équation $u_{total} = \sqrt{1^2 + 0,5^2 + 0,2^2 + 0,1^2 + 0,05^2}$, une valeur σ potentielle aléatoire a été calculée à l'aide de l'équation (15) ;
- Chaque valeur de u_{total} , from $u_{total} = \sqrt{1^2}$ to $u_{total} = \sqrt{1^2 + 0,5^2 + 0,2^2 + 0,1^2 + 0,05^2}$ a été calculée ;
- Cette opération a été effectuée 200000 fois en 20 sous-séries de 10000, pour chaque valeur de n ;
- Pour chacune des sous-séries, la valeur moyenne quadratique QM et les centiles 2,5% et 97,5% ont été calculés ;
- La valeur moyenne et l'écart-type ont été calculés à partir des sous-séries pour les paramètres (c'est-à-dire les QM et les centiles), ce qui a permis de calculer une incertitude étendue $2.u$ sur les QM et les centiles.

Dans cet exemple :

- Les équations (8) et (15) décrivent le comportement du phénomène. Nous disposons d'un modèle valide, comme demandé pour mettre en œuvre la méthode de Monte-Carlo ;
- L'équation (8) exige l'indépendance de u_i pour être valide. Nous pouvons alors valablement utiliser 5 nombres aléatoires pour chaque réalisation de Monte-Carlo ;
- La division des 200000 réalisations en 20 sous-groupes permet de calculer $2.u$ et de décider si le nombre de réalisations est suffisant ou non. Dans le cas présent, tous les u sont suffisamment faibles pour notre application, à l'exception des valeurs moyennes pour $n = 3$. Si ces valeurs avaient été nécessaires pour une utilisation quelconque, l'algorithme de Monte-Carlo devrait probablement être poursuivi pour obtenir des valeurs plus précises.

Les résultats sont présentés dans le Tableau 4 comme suit.

Tableau 4. Moyenne, limites inférieure et supérieure de l'intervalle IC95% des rapports σ/s en fonction de n (nombre de répétitions) et de N_t (nombre de termes dans l'équation de u_{total}).

N_t	n	3	4	5	6	8	10	13	16	20	25	32	40	50	
1	Moyenne	3,44	1,732	1,414	1,292	1,183	1,133	1,098	1,074	1,058	1,044	1,033	1,027	1,021	
	2u	0,26	0,011	0,005	0,002	0,001	0,001	0,001	0,001	0,001	0,000	0,000	0,000	0,000	
	IC95%-	0,520	0,567	0,598	0,626	0,663	0,688	0,718	0,738	0,759	0,781	0,802	0,819	0,836	
	2u	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,000	0,001	0,000	0,000
	IC95%+	6,272	3,761	2,877	2,455	2,038	1,819	1,658	1,544	1,468	1,389	1,328	1,284	1,246	
	2u	0,041	0,016	0,012	0,006	0,004	0,003	0,002	0,002	0,002	0,005	0,001	0,001	0,001	0,001
2	Moyenne	4,12	1,940	1,582	1,444	1,323	1,267	1,226	1,201	1,183	1,168	1,155	1,148	1,141	
	2u	0,26	0,011	0,005	0,002	0,001	0,001	0,001	0,000	0,001	0,000	0,000	0,000	0,000	
	IC95%-	0,714	0,748	0,773	0,792	0,823	0,845	0,870	0,887	0,904	0,923	0,941	0,956	0,970	
	2u	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,000	0,001	0,000	0,000	
	IC95%+	7,141	4,032	3,038	2,577	2,139	1,916	1,754	1,642	1,566	1,489	1,429	1,387	1,350	
	2u	0,053	0,017	0,011	0,007	0,004	0,003	0,002	0,002	0,002	0,005	0,001	0,001	0,001	0,001
3	Moyenne	4,19	1,970	1,608	1,467	1,344	1,287	1,246	1,220	1,202	1,186	1,174	1,166	1,160	
	2u	0,26	0,011	0,005	0,002	0,001	0,001	0,001	0,000	0,001	0,000	0,000	0,000	0,000	
	IC95%-	0,769	0,791	0,811	0,827	0,854	0,873	0,896	0,912	0,928	0,946	0,963	0,977	0,991	
	2u	0,002	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,000	0,000	
	IC95%+	7,287	4,056	3,056	2,591	2,152	1,929	1,768	1,656	1,580	1,504	1,444	1,402	1,366	
	2u	0,055	0,016	0,012	0,007	0,004	0,003	0,002	0,002	0,002	0,005	0,001	0,001	0,001	0,001
4	Moyenne	4,21	1,977	1,614	1,473	1,349	1,292	1,251	1,224	1,207	1,191	1,178	1,171	1,164	
	2u	0,26	0,011	0,005	0,002	0,001	0,001	0,001	0,000	0,001	0,000	0,000	0,000	0,000	
	IC95%-	0,789	0,804	0,822	0,836	0,862	0,880	0,903	0,918	0,934	0,951	0,969	0,983	0,996	
	2u	0,002	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,000	0,001	0,000	0,000	
	IC95%+	7,317	4,059	3,058	2,594	2,155	1,932	1,771	1,659	1,584	1,508	1,448	1,406	1,369	
	2u	0,057	0,016	0,012	0,007	0,004	0,003	0,002	0,002	0,002	0,005	0,001	0,001	0,001	0,001
5	Moyenne	4,21	1,979	1,615	1,474	1,350	1,293	1,252	1,226	1,208	1,192	1,179	1,172	1,165	
	2u	0,26	0,011	0,005	0,002	0,001	0,001	0,001	0,000	0,001	0,000	0,000	0,000	0,000	
	IC95%-	0,796	0,808	0,824	0,839	0,864	0,882	0,904	0,919	0,936	0,953	0,970	0,984	0,998	
	2u	0,002	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,000	0,000	0,000	0,000	
	IC95%+	7,322	4,060	3,059	2,595	2,156	1,933	1,772	1,660	1,585	1,509	1,449	1,407	1,370	
	2u	0,057	0,016	0,012	0,007	0,004	0,003	0,002	0,002	0,002	0,005	0,001	0,001	0,001	0,001

10.12 Analyse de la variance

L'analyse de la variance (ANOVA) est une technique statistique utilisée pour identifier et quantifier les effets aléatoires individuels d'une mesure afin qu'ils puissent être correctement pris en compte lors de l'évaluation de l'incertitude du résultat de la mesure.

Bien que les méthodes ANOVA soient applicables à un large éventail de mesures, elles sont axées sur les écarts-types et ne peuvent donc pas, à elles seules, identifier les effets systématiques susceptibles d'être présents.

L'annexe H du GUM [2] donne un aperçu de la manière dont les techniques d'ANOVA peuvent être utiles pour la détermination des incertitudes.

Comprendre les bases de l'ANOVA, en particulier le fonctionnement des degrés de liberté, aide à définir des plans d'expériences plus efficaces en centrant les tests là où ils sont nécessaires.

11 Évaluation de la qualité des incertitudes déterminées par le laboratoire

11.1 Introduction

Deux méthodes différentes peuvent être utilisées pour vérifier la qualité de la détermination des incertitudes :

- ✚ La première est basée sur l'un des principes de la norme ISO 21748 [11], qui stipule que s_R (reproductibilité de la méthode d'essai) est une bonne base pour l'estimation d'une incertitude ;
- ✚ La seconde est basée sur les recommandations de la norme ISO 13528 [3], qui propose un score ζ pour évaluer les incertitudes lors des EA, qui peut être adapté pour une utilisation au sein d'un seul laboratoire.

Les deux options peuvent être mises en œuvre par un seul laboratoire ou par un organisateur d'EA pour les participants à ses programmes.

11.2 Utilisation de l'écart-type de reproductibilité

Lorsque l'écart-type de reproductibilité s_R peut être connu à partir de sources externes (généralement d'une CIL ou de la documentation scientifique ou de la normalisation), il peut servir de base à l'incertitude.

Dans la plupart des cas, le laboratoire est censé se considérer comme un laboratoire "moyen" et, par conséquent, son incertitude n'est pas significativement différente de s_R .

Dans certains cas, le laboratoire peut avoir de bonnes raisons de se considérer comme un laboratoire qui s'écarte significativement de la moyenne. Typiquement, il peut s'agir :

- ✚ D'un laboratoire de référence, qui met en œuvre des ressources meilleures que ce qui est demandé dans le document de référence (typiquement, un équipement d'essai plus précis que ce qui est demandé, des conditions environnementales plus serrées que ce qui est demandé (par exemple, des intervalles de température de $\pm 1^\circ\text{C}$ par rapport aux $\pm 5^\circ\text{C}$ spécifiés), etc. ...) ;

Par exemple, il peut s'agir d'un laboratoire national destiné à produire des MRC.

- ✚ D'un laboratoire de routine, qui met en œuvre des moyens qui ne répondent pas complètement aux exigences du document de référence mais dont la pratique est suffisamment bonne pour l'utilisation prévue des résultats d'essai.

Par exemple, un laboratoire d'usine destiné à vérifier la conformité des produits, lesquels, d'après la connaissance du processus de fabrication, ont très peu de risques de ne pas répondre aux exigences spécifiées.

Dans tous les cas, le rapport u_i/s_R devrait correspondre à ce qui est attendu du laboratoire :

- ✚ $u_i/s_R < 1$ pour les laboratoires de référence ;
- ✚ $u_i/s_R \cong 1$ pour les laboratoires moyens ;
- ✚ $u_i/s_R > 1$ pour les laboratoires qui ne répondent pas complètement aux exigences.

11.3 Utilisation d'un score ζ adapté

La méthode utilisant les ζ -scores décrite dans la norme ISO 13528 [3] peut être étendue pour évaluer la qualité des incertitudes déterminées par le laboratoire. Pour ce faire, l'équation (1) peut être reformulée en équation (16), comme suit :

$$\zeta = \frac{x_i - X_{RM}}{\sqrt{u_{RM}^2 + u_i^2}} \quad (16)$$

où x_i est un résultat quelconque obtenu par le laboratoire sur un MR,

X_{RM} est la valeur attribuée au MR,

u_{RM} est l'incertitude sur X_{RM} ,

et u_i est l'incertitude déclarée par le laboratoire concernant ce résultat d'essai.

Comme indiqué au § 7.2, il n'est pas nécessaire que le MR soit un MRC, mais il doit s'agir d'un matériau dont le X_{RM} est connu par une autre source que les essais réalisés au sein du laboratoire (c'est-à-dire au moyen d'une CIL, d'une autre source de connaissances, d'une collaboration entre plusieurs laboratoires, etc. ...).

Bien entendu, les scores ζ ne doivent pas être calculés à partir d'ensembles de résultats d'essais qui ont été utilisés pour calculer u_i . Un tel calcul constituerait une sorte d'auto-confirmation circulaire.

En raison de leur définition, on peut considérer que les scores ζ suivent une distribution gaussienne dont la valeur moyenne est de 0 et l'écart type de 1. Les limites habituelles (c'est-à-dire 2 et 3) peuvent alors être utilisées pour évaluer les valeurs u_i . De plus, lorsqu'une quantité suffisante de scores ζ a été calculée à partir d'une série de valeurs x_i , l'écart-type s_ζ des valeurs ζ peut être calculé, et ce s_ζ ne devrait pas s'écarter significativement de 1. Les recommandations du § 10.7 peuvent être utilisées pour vérifier cela.

Lorsque U intègre un terme de biais (sous la forme $U = B + k \cdot u$), l'équation (16) doit être un peu adaptée. La façon la plus simple de le faire est d'utiliser les valeurs u_i calculées avec l'équation $u_i = B/k + u$, de sorte que B et u représentent correctement leurs sources d'incertitude correspondantes. Dans ce cas, la seule limite qui ait un sens pour ζ est k , c'est-à-dire 2 dans la plupart des cas.

L'utilisation de l'équation (16) sur une seule valeur u_i permet uniquement de vérifier si elle est sous-estimée. Lorsque la valeur u_i est surestimée, le score ζ correspondant devient très faible, mais aucun test statistique ne peut le mettre en évidence.

Au contraire, le calcul des valeurs s_ζ permet de vérifier à la fois la sous-estimation et la surestimation des valeurs u_i , mais seulement à un niveau qui implique plusieurs déterminations d'incertitude :

- ✚ La surestimation de u_i conduit à des valeurs s_ζ significativement inférieures à 1 et la sous-estimation de u_i conduit à des valeurs s_ζ significativement supérieures à 1 ;
- ✚ Les valeurs s_ζ n'ont de sens que pour un ensemble de plusieurs déterminations d'incertitude.

L'évaluation des valeurs de s_ζ permet donc d'évaluer si un problème général d'évaluation des incertitudes se pose, par exemple pour EA une méthode particulière de détermination de ces incertitudes.

11.4 Résultats de l'évaluation des incertitudes

Les Figure 7.a à i fournissent des représentations graphiques des évaluations de u_i effectuées conformément aux propositions du présent § 11, à partir d'une sélection de programmes d'EA effectués par CompaLab en 2023. En plus de ces représentations graphiques, le Tableau 5 fournit les écarts types s_ζ des scores ζ pour cette sélection.

Il convient de noter que :

- ✚ Dans cet exercice, les s_ζ représentent les écarts types des scores ζ des participants et non d'une série de scores ζ provenant d'évaluations effectuées dans un même laboratoire. Par conséquent, un écart significatif de s_ζ à 1 concerne la globalité des participants plutôt que chacun d'entre eux ;
- ✚ En raison de cette situation, certains scores ζ sont apparus comme des valeurs aberrantes qui ont eu un impact important sur le calcul de s_ζ . Ces valeurs aberrantes ont donc été supprimées des données avant les calculs. Leur nombre pour chaque programme est mentionné dans le tableau 5 ;
- ✚ Afin de faciliter la lecture, les Figure 7.a à i et le Tableau 5 ont été classés par valeurs croissantes de s_ζ .

Tableau 5. Écarts types s_ζ des scores ζ pour une sélection de programmes d'EA de 2023

Programme d'EA	Analyse chimique - Acier faiblement allié - Teneur en carbone %	Essai de traction - Résistance à la traction - Rm	Essai de traction - Limite d'élasticité Rp0,2	Essai de dureté Vickers - HV10	Essai de choc Charpy - Énergie – KV2	Essai de choc Charpy - Expansion latérale	Force de cisaillement des treillis soudés	Essai de traction - Allongement à la rupture A5d	Essai de corrosion Huey - Aciers inoxydables - Taux de corrosion (48h)
Nombre de participants	118	26	24	81	25	19	30	23	25
Nombre de u_i déclarées	78	15	12	51	16	10	16	13	10
Percentage	66%	58%	50%	63%	64%	53%	53%	57%	40%
Nb d'aberrants pour les scores ζ	3	1	2	2	0	0	0	1	1
s_ζ	0,90	1,07	1,35	1,56	2,14	2,41	2,75	3,02	3,42

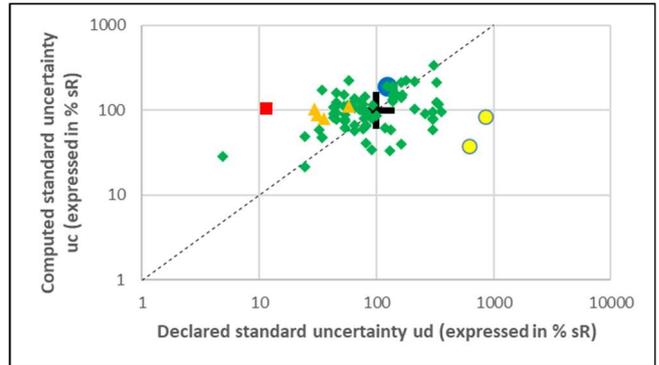
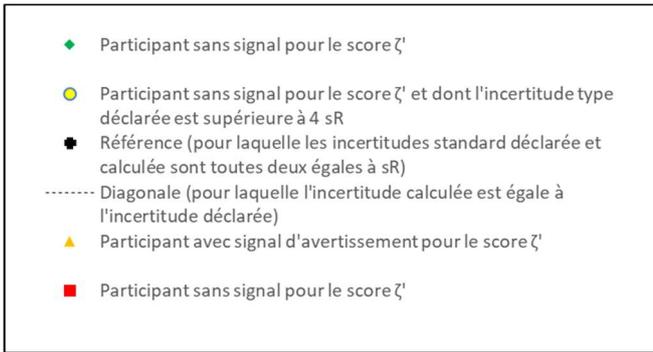


Figure 7.a : Résultats 2023 d'évaluation des incertitudes pour la teneur en carbone d'un acier faiblement allié.

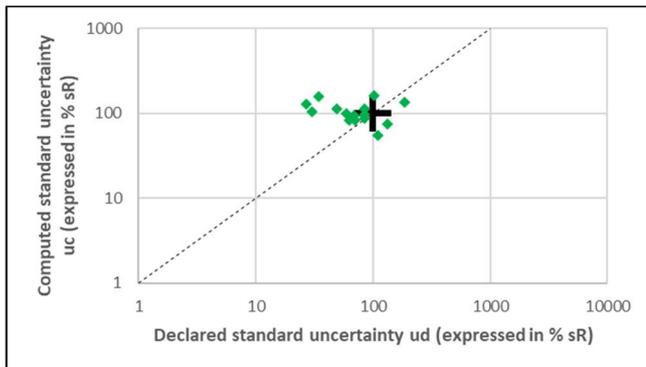


Figure 7.b : Résultats 2023 d'évaluation des incertitudes pour la résistance à la traction R_m .

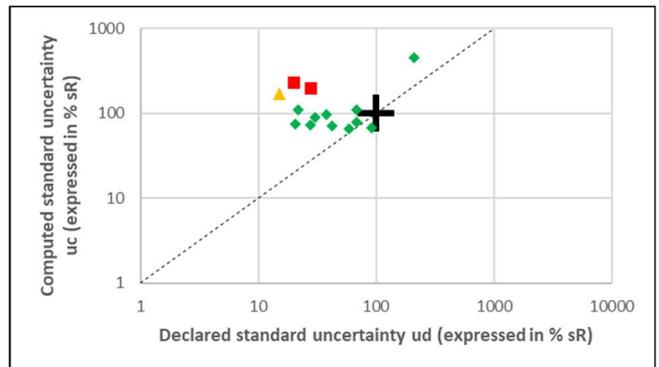


Figure 7.c : Résultats 2023 d'évaluation des incertitudes pour la limite d'élasticité $R_{p0.2}$.

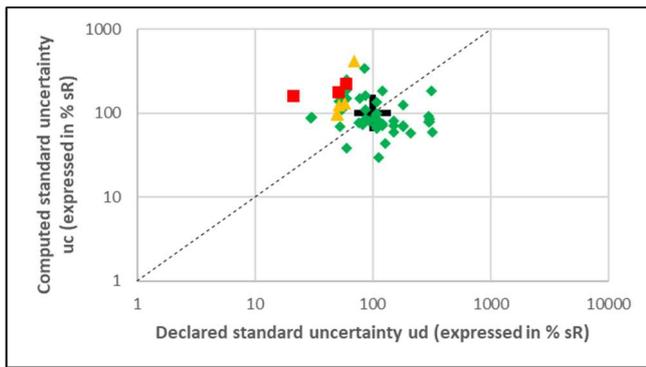


Figure 7.d : Résultats 2023 d'évaluation des incertitudes pour la dureté Vickers HV_{10} .

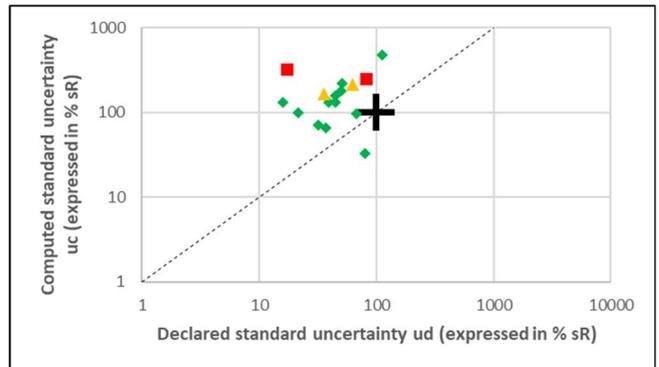


Figure 7.e : Résultats 2023 d'évaluation des incertitudes pour l'essai de flexion par choc Charpy, Energie KV_2 .

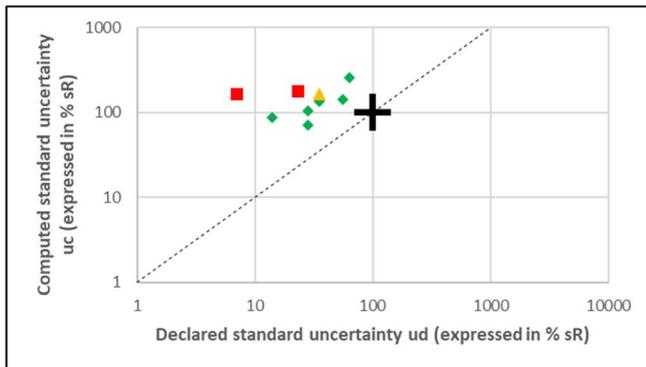


Figure 7.f : Résultats 2023 d'évaluation des incertitudes pour l'essai de flexion par choc Charpy, Expansion latérale LE .

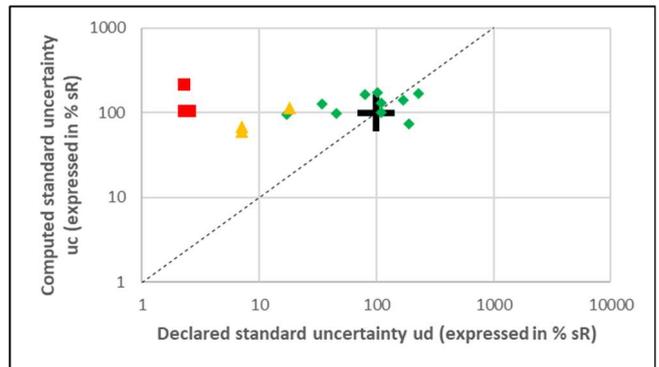


Figure 7.g : Résultats 2023 d'évaluation des incertitudes pour la force de cisaillement d'un joint soudé.

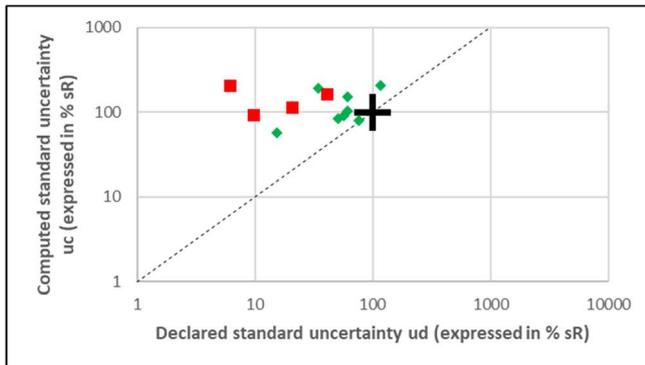


Figure 7.h : Résultats 2023 d'évaluation des incertitudes pour l'essai de traction, Allongement à rupture A_{5d} .

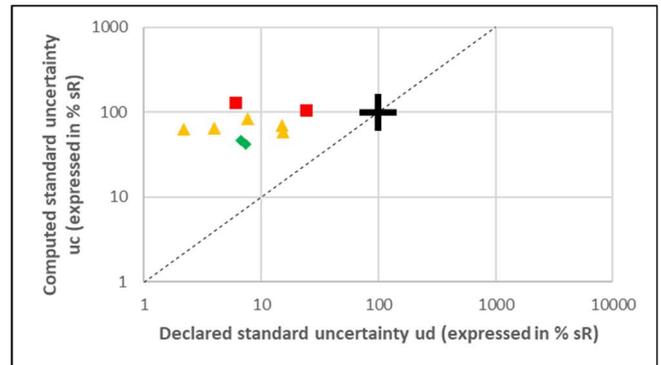


Figure 7.i : Résultats 2023 d'évaluation des incertitudes pour la vitesse de corrosion intergranulaire lors d'un essai Huey sur acier inoxydable.

Ces résultats permettent d'observer que :

- ✚ Les conclusions du § 3.2 sont confirmées par cette série de résultats supplémentaires. Les méthodes d'essai pour lesquelles des MRC existent ou qui sont principalement métrologiques sont celles pour lesquelles les incertitudes u_i sont les mieux déterminées. Parmi celles pour lesquelles des MRC existent (c'est-à-dire la chimie, la dureté, les essais de traction et l'essai Charpy), les méthodes non destructives sont celles pour lesquelles les valeurs de u_i sont les mieux déterminées ;
- ✚ Les écarts-types s_z sont un bon paramètre pour déterminer dans quelle mesure la méthode est difficile pour la détermination des incertitudes associées : plus s_z est grand, plus la moyenne u_i/s_R est faible, plus la proportion d'alertes est importante, c'est-à-dire plus les incertitudes u_i des participants sont sous-estimées ;
- ✚ Plus la détermination des incertitudes u_i est difficile, moins la proportion de laboratoires qui les déterminent est importante ;
- ✚ Une remarque particulière doit être faite pour l'essai de cisaillement sur treillis soudés, qui est évidemment une méthode totalement "technologique", mais pour laquelle une grande partie des laboratoires ont déterminé leurs incertitudes u_i à partir des résultats des CIL auxquelles ils ont participé antérieurement. On constate alors que, dans ce cas, l'utilisation des résultats de CIL est le seul moyen d'obtenir des déterminations correctes de ces incertitudes.

12 Conclusions

Une étude des résultats d'évaluation des incertitudes lors des CIL organisées par CompaLab a montré que la plupart des incertitudes déterminées par les participants sont largement sous-estimées. À cet égard, on peut distinguer trois catégories de méthodes d'essai :

- ✚ Les méthodes d'essai qui sont principalement métrologiques ou pour lesquelles des MRC sont disponibles (typiquement les essais chimiques) ;
- ✚ Les méthodes d'essai qui comprennent une part importante de sources qualitatives d'incertitudes (typiquement les essais de traction) ;
- ✚ Les méthodes d'essai dont les incertitudes sont principalement régies par des sources qualitatives (typiquement les essais de corrosion).

Sans surprise, les incertitudes sont globalement bien déterminées pour certaines des méthodes d'essai de la première catégorie, alors qu'elles sont globalement sous-estimées d'un facteur 1 pour 10 ou plus pour certaines des méthodes d'essai de la troisième catégorie.

Cette situation semble provenir d'un choix massif des laboratoires d'utiliser la méthode B du GUM pour déterminer leurs incertitudes, quelle que soit la méthode d'essai. Plusieurs raisons peuvent expliquer cette situation :

- ✚ Il y a encore une vingtaine d'années, la plupart des laboratoires n'étaient pas familiarisés avec les techniques de détermination des incertitudes. Ils étaient alors enclins à sous-traiter ce travail à des consultants, qui venaient principalement de la métrologie où la méthode GUM B est généralement efficace ;
- ✚ La méthode B du GUM peut être mise en œuvre dans un seul laboratoire, tandis que les autres méthodes nécessitent une collaboration directe ou indirecte (c.-à-d. en utilisant un MR) entre plusieurs laboratoires.

Pendant, les méthodes GUM, en particulier la méthode B, sont efficaces dans le domaine de la métrologie (où des étalons sont toujours disponibles), mais ne le sont pas en présence de sources qualitatives significatives d'incertitude, ce qui est assez courant dans les essais en laboratoire. C'est pour cette raison que la question du biais est plus importante :

- ✚ La question du biais est plus importante dans les essais en laboratoire que dans l'étalonnage ;
- ✚ Les laboratoires sont enclins à considérer comme négligeables certaines sources qualitatives d'incertitude simplement parce que leur contribution est difficile à déterminer par cette méthode.

De plus, certaines questions plus ou moins spécifiques aux essais ne sont pas abordées dans le GUM (biais de la méthode d'essai, effet du matériau, homogénéité interne des éprouvettes, effet de la gamme des mesurandes, préparation des éprouvettes, ...) et le GUM ne fournit pas suffisamment de recommandations sur ces questions.

D'autre part, les informations provenant d'autres sources, généralement les résultats provenant de CIL et les résultats des programmes de surveillance de la qualité mis en œuvre par le laboratoire peuvent être réutilisés directement ou comme matière première pour les expériences selon la méthode A du GUM afin de déterminer les incertitudes d'une manière qui prend mieux en compte les questions énumérées ci-dessus et, par conséquent, de fournir de bien meilleures estimations des incertitudes. En outre, la réutilisation de ces données nécessite beaucoup moins de ressources internes (temps et argent) que la mise en œuvre de la méthode B du GUM.

Lorsqu'il est important que les incertitudes soient bien déterminées, une grande expérience collaborative selon la méthode B du GUM (c'est-à-dire une CIL dédiée) devrait être organisée, dont les résultats peuvent également être ensuite utilisés pour mettre en œuvre des programmes très efficaces de surveillance interne de la qualité des essais, comme cela est demandé pour l'accréditation des laboratoires.

La détermination des incertitudes devrait toujours commencer par :

- ✚ Une clarification de l'utilisation prévue des incertitudes à déterminer ;
- ✚ Une collecte d'informations publiques ou internes concernant la fidélité des essais (en particulier à partir des anciens résultats de CIL) ;
- ✚ Une estimation du type de méthode d'essai (fortement qualitative ou pas)

La méthode la plus appropriée pour déterminer les incertitudes dépend fortement des réponses à ces questions. Et dans la plupart des cas, la réponse n'est pas la "méthode B du GUM" !

13 Références

- [1] ISO/CEI 17025:2017 : Exigences générales concernant la compétence des laboratoires d'étalonnages et d'essais
- [2] BIPM, JCGM 100 : Évaluation des données de mesure — Guide pour l'expression de l'incertitude de mesure, 2008, DOI : https://www.bipm.org/documents/20126/2071204/JCGM_100_2008_F.pdf/53384399-1e6d-b598-dd17-

[2831a6b5b812?version=1.14&t=1696944988349&download=true](https://www.iso.org/standard/72411.html)

également publié par l'ISO sous la référence ISO/Guide 98.

- [3] ISO 13528:2022 : Méthodes statistiques utilisées dans les essais d'aptitude par comparaison interlaboratoires
- [4] BIPM, JCGM 200 : Vocabulaire international de métrologie – Concepts fondamentaux et généraux et termes associés, 2008,
DOI : https://www.bipm.org/documents/20126/2071204/JCGM_200_2012.pdf/f0e1ad45-d337-bbeb-53a6-15fe649d0ff1?version=1.16&t=1659082802818&download=true
également publié par l'ISO sous la référence ISO/Guide 99.
- [5] ISO 3534-2:2006 : Statistique — Vocabulaire et symboles – Partie 2 : Statistique appliquée
- [6] ISO 5725-1:2023 : Exactitude (justesse et fidélité) des résultats et méthodes de mesure – Partie 1: Principes généraux et définitions
- [7] ISO 6507-1:2023 : Matériaux métalliques — Essai de dureté Vickers – Partie 1 : Méthode d'essai
- [8] ISO 5725-2 : Exactitude (justesse et fidélité) des résultats et méthodes de mesure – Partie 2 : Méthode de base pour la détermination de la répétabilité et de la reproductibilité d'une méthode de mesure normalisée
- [9] ASTM E691-22 : Standard Practice for Conducting an Interlaboratory Study to Determine the Precision of a Test Method
- [10] ISO 33405:2024 : Matériaux de référence — Approches pour la caractérisation et l'évaluation de l'homogénéité et la stabilité
- [11] ISO 21748:2017 : Lignes directrices relatives à l'utilisation d'estimations de la répétabilité, de la reproductibilité et de la justesse dans l'évaluation de l'incertitude de mesure
- [12] ISO 5725-3:2023 : Exactitude (justesse et fidélité) des résultats et méthodes de mesure – Partie 3 : Fidélité intermédiaire et plans alternatifs pour les études collaboratives
- [13] ISO 3:1973 : Nombres normaux — Séries de nombres normaux
- [14] ISO 33406:2024 : Approches pour la production de matériaux de référence avec des propriétés qualitatives
- [15] ISO 16269-6:2014 : Interprétation statistique des données – Partie 6 : Détermination des intervalles statistiques de dispersion
- [16] ISO 2854:1976 : Interprétation statistique des données — Techniques d'estimation et tests portant sur des moyennes et des variances
- [17] David Luengo, Luca Martino, Mónica Bugallo, Víctor Elvira and Simo Särkkä, "A survey of Monte Carlo methods for parameter estimation" EURASIP Journal on Advances in Signal Processing, Article 25, May 2020
DOI : <https://doi.org/10.1186/s13634-020-00675-6>
- [18] BIPM, JCGM 101, Évaluation des données de mesure — Supplément 1 du "Guide pour l'expression de l'incertitude de mesure" — Propagation de distributions par une méthode de Monte Carlo, 2008 (en anglais)
DOI : https://www.bipm.org/documents/20126/2071204/JCGM_101_2008_E.pdf/325dcaad-c15a-407c-1105-8b7f322d651c?version=1.15&t=1696950361579&download=true

Annexe :

Exemples of implementing different methods for the determination of uncertainties

Exemples de résultats de CIL disponibles publiquement :

Chimie des métaux : ISO/TR 9769:2018 - Aciers et fontes — Vue d'ensemble des méthodes d'analyse disponibles

Beaucoup de normes ASTM concernant des méthodes d'essais fournissent des informations relatives à la répétabilité et la reproductibilité déterminées par des CIL, et notamment les ASTM A90, ASTM B487, ASTM B499, ASTM C736, ASTM C792, ASTM C1135, ASTM D522, ASTM D523, ASTM D638, ASTM D696, ASTM D785, ASTM D790, ASTM D792, ASTM D882, ASTM D1005, ASTM D1505, ASTM D2202, ASTM D2240, ASTM D2457, ASTM D2584, ASTM D2794, ASTM D3359, ASTM D3418, ASTM D3895, ASTM D4366, ASTM D5630, ASTM D6980, ASTM E8, ASTM E10, ASTM E18, ASTM E21, ASTM E23, ASTM E45, ASTM E92, ASTM E96, ASTM E112, ASTM E228, ASTM E308, ASTM E384, ASTM E399, ASTM E606, ASTM E647, ASTM E698, ASTM E793, ASTM E831, ASTM E930, ASTM E1131, ASTM E1356, ASTM E1641, ASTM E2041, ASTM E2550, ASTM G1, ASTM G28, ASTM G48.

Il convient de prêter attention au fait que l'ASTM E177 (§ 10) définit r (répétabilité) et R (reproductibilité) comme des valeurs qu'environ 95 % des paires de résultats d'essais obtenus respectivement dans des conditions de répétabilité ou de reproductibilité ne dépassent pas. Par conséquent, $r = 2,8.s_r$ et $R = 2,8.s_R$. Les raisons en sont fournies dans la même référence.

Exemple de plan d'expériences aléatoires pour une expérience exhaustive selon la méthode A du GUM

Exemple d'un laboratoire préparant lui-même les éprouvettes, utilisant 2 méthodes d'essai, 8 machines d'essai et employant 5 opérateurs. Il a été décidé que :

- 2 types de matériaux (moyen et difficile) ;
- 5 niveaux de mesurandes ;

devraient être inclus dans l'expérience. Le plan d'expériences devrait donc inclure :

- 2 méthodes d'essai ;
- 2 types de matériaux (moyens et difficiles) ;
- 5 niveaux de mesurandes ;
- 3 sources de préparation d'éprouvettes (ses propres installations et, dans la mesure du possible, 2 sources externes) ;
- 8 machines d'essai ;
- 5 opérateurs ;
- 2 conditions environnementales (c'est-à-dire, réalisation des essais à 2 périodes différentes de l'année) ;
- 2 répétitions, si l'ET de répétabilité est nécessaire.

Soit un nombre total de $2 \times 2 \times 5 \times 3 \times 8 \times 5 \times 2 = 9600$ tests. Ce nombre d'essais est techniquement et économiquement impossible à mettre en œuvre. De plus, les MR ne sont pas disponibles pour toutes les combinaisons de types de matériaux et de mesurandes. Il est donc décidé d'utiliser un plan d'expérience aléatoire qui tient compte des facteurs suivants :

- Les proportions réelles des combinaisons de méthodes d'essai - machines d'essai - niveaux de mesurandes - opérateurs que le laboratoire utilise pour produire les résultats d'essai ;
- Utiliser des coefficients de pondération pour rendre l'expérience plus représentative des opérations réelles du laboratoire. Pour ce faire, des coefficients de 0,3 à 3 ont été attribués à chaque combinaison, avec 1 pour les combinaisons moyennes ;

- Niveaux des mesurandes 1 - 3 - 10 - 30 - 100. Toutes les machines d'essai ne sont pas en mesure de traiter tous les niveaux de mesurandes ;
- Nécessité d'utiliser les MR, mais sur seulement 20 % des résultats d'essai (ce qui est suffisant pour obtenir une précision acceptable dans la détermination du biais). En outre, les MR ne sont disponibles que pour le "matériau moyen" et les niveaux de mesurande 3 et 30 ;
- Nécessité de déterminer la répétabilité, c'est-à-dire le nombre de répétitions, mais sur seulement 20% des combinaisons (ce qui est suffisant pour obtenir une précision acceptable dans la détermination du biais).

La matrice des combinaisons et la pondération ont été définies comme suit :

Source d'incertitude	Occurrence	Pondération
Type de matériau	Moyen	0,9
	Difficile	0,1
Préparation	Interne	0,333
	Externe 1	0,333
	Externe 2	0,333
Conditions environnementales	Hiver	0,5
	Eté	0,5
Répétitions	Avec	0,2
	Sans	0,8

Méthode d'essai	Machine d'essai	Opérateur	Mesurande	Pondération
A	1	Albert	1 to 30	0,5
A	1	Benoît	1 to 30	2
A	1	Camille	1 to 30	0,5
A	2	Albert	3 to 100	2
A	2	Benoît	3 to 100	0,5
A	2	Daniel	3 to 100	0,5
A	2	Etienne	3 to 100	0,5
A	3	Albert	1 to 1000	0,3
A	3	Benoît	1 to 1000	0,3
A	3	Camille	1 to 1000	0,3
A	3	Daniel	1 to 1000	0,3

Méthode d'essai	Machine d'essai	Opérateur	Mesurande	Pondération
A	3	Etienne	1 to 1000	0,3
A	4	Daniel	10 to 100	2
A	4	Etienne	10 to 100	0,5
B	4	Daniel	10 to 100	0,5
B	4	Etienne	10 to 100	2
A	5	Albert	3 to 30	0,3
A	5	Camille	3 to 30	0,3
B	5	Albert	3 to 30	1
B	5	Benoît	3 to 30	0,5
B	5	Camille	3 to 30	2
B	5	Daniel	3 to 30	1
B	5	Etienne	3 to 30	0,5
A	6	Alfred	3 to 100	0,5
A	6	Benoît	3 to 100	0,5
A	6	Camille	3 to 100	0,5
B	6	Alfred	3 to 100	0,5
B	6	Camille	3 to 100	3
B	7	Daniel	3 to 30	2
B	7	Etienne	3 to 30	1
B	8	Daniel	10 to 100	1
B	8	Etienne	10 to 100	3

Cela donne $118 \times 2 \times 3 \times 2 = 1416$ possibilités parmi lesquelles 100 seront choisies au hasard, dont 25 seront choisies au hasard pour être répétées pour calculer s_r . Les essais sur les mesurandes 3 et 30 des "matériaux moyens" seront effectués sur MR afin de déterminer le biais.

La sélection aléatoire pondérée peut être réalisée en suivant les étapes suivantes :

- ✚ Construire la matrice de toutes les combinaisons possibles, y compris le poids global pour chacune d'entre elles (ce poids est obtenu par la multiplication de tous les poids individuels, exemple : la combinaison "moyenne - interne - hiver - A - 1 - Albert - 1" a un poids global de $w = 0,9 \times 0,333 \times 0,5 \times 0,5 = 0,075$);
- ✚ Construire la matrice pondérée, qui répète chaque combinaison dans un nombre aléatoire pondéré de fois. Un facteur de multiplication m (égal pour toutes les combinaisons) peut être nécessaire pour obtenir un nombre entier approprié pour chaque combinaison (exemple avec un nombre aléatoire 0,44, $w = 0,075$ et $m = 100$, $0,44 \times 0,075 \times 100 = 3,3$ à arrondir à 3);
- ✚ Sélectionner au hasard le nombre souhaité de combinaisons à tester dans cette dernière matrice.

Un exemple de plan d'expérience résultant de 100 combinaisons à tester, parmi lesquelles les 25 premières doivent être répétées, est fourni ci-après.

Méthode d'essai	Machine d'essai	Opérateur	Mesurande	Type de matériau	Préparation	Conditions environnementale
A	1	Albert	10	Difficile	Externe 2	Été
A	2	Albert	10	Moyen	Externe 1	Hiver
A	1	Camille	10	Moyen	Externe 1	Hiver
A	3	Benoît	10	Moyen	Externe 2	Été
B	6	Camille	10	Moyen	Externe 1	Hiver
B	8	Etienne	30	Moyen	Externe 1	Été
A	3	Etienne	100	Moyen	Externe 1	Été
B	8	Etienne	10	Moyen	Interne	Été
A	3	Camille	10	Moyen	Externe 2	Été
B	8	Daniel	100	Difficile	Externe 2	Hiver
B	4	Daniel	100	Difficile	Interne	Été
A	5	Albert	10	Difficile	Externe 2	Hiver
B	8	Daniel	10	Moyen	Externe 1	Hiver
B	8	Etienne	10	Moyen	Interne	Été
A	1	Albert	30	Moyen	Interne	Été
A	4	Etienne	100	Moyen	Externe 1	Hiver
A	1	Albert	10	Difficile	Externe 2	Hiver
B	5	Camille	3	Moyen	Externe 1	Hiver
B	6	Camille	100	Moyen	Externe 2	Hiver
B	7	Etienne	10	Moyen	Externe 2	Été
B	7	Daniel	10	Moyen	Externe 2	Hiver
B	6	Camille	3	Difficile	Externe 2	Été
B	5	Camille	10	Moyen	Externe 1	Été
B	5	Camille	3	Moyen	Interne	Été
B	8	Daniel	30	Moyen	Externe 1	Été
A	2	Benoît	100	Moyen	Externe 2	Été

Méthode d'essai	Machine d'essai	Opérateur	Mesurande	Type de matériau	Préparation	Conditions environnementale
B	6	Alfred	3	Moyen	Externe 2	Hiver
A	1	Benoît	1	Moyen	Externe 2	Hiver
B	7	Etienne	3	Moyen	Externe 1	Été
B	8	Daniel	10	Moyen	Externe 2	Été
B	8	Daniel	100	Moyen	Interne	Été
B	8	Etienne	30	Moyen	Externe 1	Été
A	2	Albert	30	Difficile	Interne	Hiver
B	6	Camille	30	Moyen	Externe 1	Été
B	8	Etienne	10	Moyen	Externe 2	Été
A	3	Albert	10	Moyen	Externe 1	Été
B	6	Camille	3	Difficile	Externe 2	Été
B	5	Camille	10	Moyen	Interne	Hiver
A	2	Daniel	30	Moyen	Externe 1	Été
B	7	Etienne	30	Moyen	Externe 1	Été
B	6	Camille	100	Difficile	Externe 2	Été
B	5	Benoît	30	Moyen	Externe 2	Hiver
B	7	Etienne	30	Moyen	Interne	Hiver
A	4	Etienne	30	Difficile	Externe 1	Été
B	5	Camille	10	Moyen	Externe 1	Été
A	3	Etienne	100	Moyen	Externe 2	Hiver
B	6	Camille	10	Moyen	Interne	Été
A	5	Camille	30	Moyen	Externe 1	Été
A	3	Etienne	100	Moyen	Externe 2	Hiver
A	5	Camille	30	Moyen	Externe 2	Hiver
A	2	Albert	3	Moyen	Externe 1	Hiver
A	1	Benoît	3	Moyen	Externe 2	Hiver

Méthode d'essai	Machine d'essai	Opérateur	Mesurande	Type de matériau	Préparation	Conditions environnementale
A	4	Etienne	100	Moyen	Externe 1	Hiver
A	6	Camille	30	Moyen	Externe 2	Hiver
B	8	Daniel	30	Moyen	Externe 1	Eté
A	1	Benoît	1	Moyen	Externe 2	Hiver
A	5	Camille	30	Moyen	Interne	Hiver
B	7	Etienne	3	Moyen	Interne	Eté
B	4	Daniel	10	Moyen	Interne	Hiver
A	3	Camille	30	Difficile	Interne	Hiver
B	6	Camille	3	Moyen	Externe 1	Hiver
B	7	Daniel	3	Moyen	Interne	Hiver
B	6	Alfred	100	Moyen	Externe 2	Eté
A	5	Albert	10	Difficile	Externe 1	Hiver
B	8	Daniel	10	Moyen	Externe 1	Eté
B	6	Camille	100	Moyen	Externe 1	Eté
B	7	Etienne	10	Moyen	Interne	Eté
A	2	Daniel	30	Moyen	Externe 2	Eté
B	5	Albert	30	Moyen	Interne	Hiver
B	5	Benoît	30	Moyen	Interne	Eté
A	1	Albert	30	Moyen	Interne	Eté
B	8	Etienne	30	Moyen	Externe 1	Eté
B	4	Etienne	30	Moyen	Externe 2	Hiver
A	3	Etienne	100	Moyen	Externe 1	Eté
B	4	Etienne	10	Moyen	Externe 2	Hiver
A	6	Benoît	30	Difficile	Externe 2	Hiver

Méthode d'essai	Machine d'essai	Opérateur	Mesurande	Type de matériau	Préparation	Conditions environnementale
B	4	Daniel	10	Moyen	Externe 1	Hiver
B	8	Etienne	30	Moyen	Interne	Eté
A	1	Albert	30	Difficile	Interne	Hiver
A	2	Albert	30	Moyen	Externe 1	Hiver
A	6	Alfred	100	Moyen	Externe 2	Eté
B	4	Etienne	10	Moyen	Externe 2	Eté
A	2	Albert	30	Moyen	Interne	Eté
B	8	Daniel	100	Difficile	Externe 1	Hiver
B	8	Daniel	100	Moyen	Externe 2	Hiver
B	5	Camille	10	Moyen	Interne	Hiver
A	1	Benoît	1	Moyen	Externe 1	Hiver
A	6	Alfred	100	Difficile	Externe 2	Hiver
A	5	Albert	3	Moyen	Externe 2	Hiver
A	6	Alfred	3	Moyen	Interne	Eté
A	3	Benoît	3	Moyen	Externe 1	Eté
B	7	Daniel	10	Moyen	Externe 2	Hiver
A	3	Benoît	100	Moyen	Externe 1	Eté
A	2	Albert	30	Moyen	Externe 2	Hiver
B	5	Daniel	3	Moyen	Externe 2	Hiver
A	2	Benoît	10	Moyen	Interne	Hiver
B	5	Etienne	3	Moyen	Interne	Hiver
A	1	Camille	3	Moyen	Externe 1	Hiver
B	6	Alfred	100	Moyen	Interne	Eté
A	1	Albert	3	Moyen	Externe 2	Eté

Le nombre total de combinaisons par source d'incertitudes est le suivant :

Source d'incertitude	Occurrence	Nombre total	Nombre total de combinaisons répétées
Type de matériau	Moyen	84	19
	Difficile	16	6
Préparation	Interne	27	5
	Externe 1	35	10
	Externe 2	38	10
Conditions environnementales	Hiver	50	11
	Eté	50	14

<i>Source d'incertitude</i>	<i>Occurrence</i>	<i>Nombre total</i>	<i>Nombre total de combinaisons répétées</i>
Méthode	A	46	10
	B	54	15
Machine	1	12	4
	2	10	1
	3	10	3
	4	9	2
	5	17	4
	6	17	3
	7	9	2
	8	16	6
Opérateur	Albert	17	5
	Benoît	12	1
	Camille	23	8
	Daniel	18	5
	Etienne	24	6
Mesurande	1	3	0
	3	18	3
	10	30	14
	30	29	3
	100	20	5

Exemple d'étude en 2 étapes

Une autre façon de procéder pour l'exemple supérieur consisterait à mettre en œuvre deux étapes, comme suit :

1. Effectuer 8 à 10 tests pour chaque source d'incertitude et calculer le biais et l'erreur aléatoire correspondants ;
2. Sélectionner une, deux ou trois sources principales d'incertitude et effectuer une série de tests sur des combinaisons aléatoires limitées aux sources d'incertitude qui contribuent principalement à l'incertitude globale.

Le plan d'expérience pourrait typiquement être le suivant :

1. 8 essais sur le matériau habituel et 2 sur le matériau difficile ;
2. 3 essais sur chacune des sources de préparation des éprouvettes ;
3. 5 essais pour deux moments d'essai ;
4. 5 essais pour chaque méthode d'essai ;
5. 1 essai pour chaque machine d'essai ;
6. 2 essais pour chaque opérateur ;
7. 2 essais pour le mesurande "3", 3 essais pour le mesurande "10", 3 essais pour le mesurande "30", 2 essais pour le mesurande "3", sur RM pour les mesurandes 3 et 30.

Pour chacun d'entre eux, toutes les autres sources d'incertitude doivent être identiques, dans la mesure du possible (pour une méthode d'essai, il se peut qu'aucune des machines et aucun des opérateurs ne soit capable de produire des résultats d'essai pour les deux méthodes).

Exemples de mise en œuvre de la méthode B du GUM

Les documents suivants contiennent des exemples de déterminations d'incertitudes utilisant la méthode B du GUM :

- 🚧 GUM [2] Annexe H : plusieurs exemples dont un relatif à une méthode d'essai (dureté Rockwell C) ;*
- 🚧 ISO 148-1, concernant l'essai d'impact Charpy ;*
- 🚧 ISO 4545, concernant l'essai de dureté Knoop ;*
- 🚧 ISO 6506-1, concernant l'essai de dureté Brinell ;*
- 🚧 ISO 6507-1, concernant l'essai de dureté Vickers ;*
- 🚧 ISO 6508-1, concernant l'essai de dureté Rockwell ;*
- 🚧 ISO 6892-1, concernant l'essai de traction des métaux ;*
- 🚧 ISO/TR 15263, concernant l'essai de traction des métaux.*