

EXACTITUDE ET INTERVALLES STATISTIQUES EN VALIDATION DE METHODE

Catherine Yardin^a

Laboratoire national de métrologie et d'essais, Trappes, France

Summary. Accuracy is an important concept in method validation because it represents the global performance of the method. In general, accuracy is quantified by combining trueness and measurement precision. From the metrological point of view, it is not correct because accuracy is a qualitative concept. In method validation, practitioners use measurement error quantity instead of accuracy. They evaluate an interval $IM = X_{\text{average}} \pm k \cdot S_{FI}$ where X_{average} is the average of the measurement and S_{FI} the intermediate precision standard deviation. This interval is compared with an acceptance interval to validate the method. To choose the k value, two approaches are commonly used. The first approach takes $k = 2$. The second named "accuracy profile" evaluates a β -Expectation tolerance interval which contains in average a proportion β of measurement values. We compare these intervals to different statistical intervals (confidence, tolerance and prediction interval). Our goal is to give explanations on the interpretation of the β -Expectation tolerance interval in the context of validation and a more user-friendly expression.

1 Introduction

L'exactitude est une caractéristique importante en validation de méthode car elle permet d'évaluer de manière globale l'aptitude de la méthode. L'exactitude s'applique à un résultat individuel et elle comprend à la fois la justesse et la fidélité. Ces deux concepts sont bien connus : la justesse correspond à l'écart entre la moyenne et la valeur vraie alors que la fidélité correspond à la dispersion des résultats individuels autour de la moyenne. Néanmoins, l'évaluation de l'exactitude varie selon les référentiels et parfois, la différence est importante.

L'analyse du concept révèle des difficultés inhérentes à la définition énoncée dans le vocabulaire international de métrologie (VIM) [1]. L'exactitude y est définie comme une caractéristique qualitative qui ne peut pas être quantifiée. La justesse est elle aussi qualitative alors que la fidélité est quantitative. Cet article présente quelques-unes des ambiguïtés existantes (paragraphe 2) et justifie ainsi une interprétation cohérente avec la pratique du concept (paragraphe 3).

Le paragraphe 4 discute de l'évaluation de l'exactitude selon les normes NF T 90-210 [2] et NF V 03-110 [3]. La première valide la justesse avant de traiter l'exactitude. La deuxième évalue l'exactitude en tenant compte du nombre de mesures et du caractère estimé du biais. Le calcul est plus compliqué dans ce deuxième cas et parfois *opaque* pour le non initié. Cet article propose de le décomposer et des outils pour le simplifier.

2 l'exactitude, un concept guère applicable

2.1 Un concept difficile

Le VIM (§ 2.13) [1] définit l'exactitude comme l'« étroitesse de l'accord entre une valeur mesurée et une valeur vraie du mesurande » et précise dans la note 1 « l'exactitude de mesure n'est pas une grandeur et ne s'exprime pas numériquement ». L'exactitude est donc un concept qualitatif ; voire qualitatif ordonné : « un mesurage est quelquefois dit plus exact s'il fournit une plus petite erreur de mesure ».

Ce concept est difficile à appréhender pour les raisons suivantes :

- le type « qualitatif » n'est pas expliqué alors que le VIM traite essentiellement de concepts quantitatifs,
- il manque des exemples pratiques qui permettraient de « faire vivre le concept »,
- l'exactitude est associée à du quantifiable (l'erreur de mesure, la fidélité) et du non-quantifiable (la justesse)
- l'exactitude s'applique-t-elle à une méthode (mesurage), aux résultats produits par la méthode ou aux deux à la fois ?

2.2 Modifier la définition

^a Corresponding author : catherine.yardin@lne.fr

Une manière de s'approprier le concept est d'en modifier la définition. Par exemple, il peut être expliqué que l'exactitude n'est pas quantifiable car elle dépend de la valeur vraie du mesurande qui est inconnue (voir § 2.11 [1] sur la valeur vraie). En substituant une grandeur quantifiable à la valeur vraie, l'exactitude devient quantifiable.

Par exemple, la norme NF ISO 5725-1 [4] substitue « une valeur de référence acceptée » à la « valeur vraie » dans sa définition de l'exactitude. « le nombre infini de mesures » devient « un grand nombre de mesures » dans la définition de la justesse. Ainsi, la norme NF ISO 5725-1 utilise des concepts mesurables et quantifiables. Cette norme sert de référence (implicite) aux études de validation de méthode.

2.3 L'exactitude dans les documents de référence

Une autre manière de procéder est de s'interroger sur l'utilisation du concept en validation de méthode. Or, le concept n'est pas toujours présent dans les référentiels de portée générale.

Par exemple, la norme NF EN ISO/CEI 17025 [5] cite l'exactitude comme caractéristique d'une méthode ; par contre, elle ne cite pas la justesse. Le guide Eurachem [6] discute de la confusion générée par les concepts (exactitude, justesse) et plutôt que de modifier les définitions de [1], il propose d'utiliser une évaluation pratique. Pour la notion de justesse, cette évaluation pratique est le biais. En revanche, ce guide ne propose pas une analyse globale avec une étude de l'exactitude (la fidélité est évaluée séparément). Plusieurs auteurs ont attiré l'attention sur ces ambiguïtés des définitions du VIM [1]. Le lecteur intéressé peut consulter les références [7], [8].

2.4 Quelle exactitude utilisée pour valider une méthode ?

Comme évoqué en introduction, l'utilisation de l'exactitude répond au besoin de valider la méthode avec un critère global et quantifiable. On pourrait reformuler le concept comme la norme NF ISO 5725-1, mais il n'est pas indispensable d'utiliser l'exactitude. Aussi, cet article propose-t-il de lui substituer le concept d'erreur de mesure.

Le VIM ([1] § 2.16) donne la définition suivante de l'erreur de mesure : « différence entre la valeur mesurée d'une grandeur et une valeur de référence ». Il est précisé dans la note 1 (alinéa a) que dans le cas où la valeur de référence est un étalon ou une valeur conventionnelle, l'erreur de mesure est connue. Par ailleurs, le VIM cite comme synonyme le terme "erreur" qui est très proche du terme "écart" utilisé en validation de méthode.

Généralement, la performance exigée pour la méthode est formulée comme un écart acceptable autour de la valeur de référence [3]. L'exigence d'un niveau d'exactitude peut être une interprétation inscrite dans des documents normatifs ou des guides. Aussi, traduit-on correctement

l'exigence en se référant à l'erreur plutôt qu'à l'exactitude de la méthode.

Pour conclure ce paragraphe, nous rappellerons que compte tenu des ambiguïtés indiquées précédemment, il est nécessaire de compléter voire de corriger la définition de l'exactitude [1]. Dans le cas où le concept est qualitatif (comme indiqué dans la définition), il serait profitable d'en indiquer la raison et d'introduire des exemples illustratifs. Il peut être envisagé d'ajouter une note prohibant l'utilisation « quantitative » du concept. Dans le cas où le concept est quantitatif (comme validé par l'usage), il est souhaitable d'utiliser des définitions cohérentes comme par exemple, celles de la norme NF ISO 5725-1.

3 Valider une méthode avec l'erreur de mesure

3.1 De la définition à l'expression

Le VIM [1] ne propose pas d'expression mathématique permettant de quantifier l'erreur de mesure, mais on peut aisément déduire de la définition la formule suivante :

$$E = X - X_{\text{réf}} \quad (1)$$

où E, X, $X_{\text{réf}}$ sont respectivement l'erreur de mesure, le résultat de mesure et une (la) valeur de référence.

L'erreur de mesure comprend une composante systématique ([1] § 2.17) et une composante aléatoire ([1] § 2.19). L'erreur systématique est estimée par le biais de mesure. L'erreur aléatoire est supposée nulle en moyenne avec une dispersion évaluée par la fidélité. Aussi, l'équation (1) peut être reformulée en distinguant ces deux composantes :

$$E = (X - X_{\text{moyen}}) + (X_{\text{moyen}} - X_{\text{réf}}) \quad (2)$$

où le premier terme représente l'écart à la moyenne et le deuxième terme, le biais.

L'équation (2) justifie l'utilisation de l'intervalle (3) pour estimer l'erreur de mesure de la méthode.

3.2 Estimer et anticiper

L'équation (2) est valable lorsque l'on dispose d'une valeur de référence du même niveau que la valeur mesurée. Lors de l'étude de validation, des mesures sont réalisées ponctuellement dans le temps (moment de la validation) et l'espace (nombre limité de points du domaine d'application de la méthode). Pour ces points, l'erreur de mesure, le biais et la fidélité sont déterminés. Mais l'objectif est de valider la méthode pour l'ensemble des mesures réalisées et à réaliser. Sous cet angle, les résultats ponctuels obtenus lors de la validation ne sont qu'un échantillon au temps t de cet ensemble de résultats. Aussi, la validation est-elle menée en utilisant des outils (statistiques) d'estimation et de prévision.

3.3 Intervalles statistiques

Il est habituel en statistique de représenter l'étendue des valeurs par un intervalle du type :

$$IM = X_{\text{moyen}} \pm k.S \quad (3)$$

où S est un écart-type et k un coefficient qui dépend de facteurs comme le nombre de mesures, l'écart-type de la moyenne, la loi de distribution des mesures... Trois types d'intervalle statistique sont utilisés pour caractériser une population : l'intervalle de confiance qui porte sur un paramètre (ex : moyenne), l'intervalle de dispersion qui porte sur une proportion donnée de la population et l'intervalle de prévision qui contient un nouvel échantillon de la population [10].

En métrologie, le résultat est exprimé sous la forme (3) avec $k = 2$ et S correspondant à l'incertitude-type. Cet intervalle est proche, d'après l'objectif visé, d'un intervalle de dispersion.

En validation de méthode, l'aspect prévision est privilégié. S'étant assuré que les résultats testés sont valides, on souhaite déterminer un intervalle dans lequel se situeront nos futurs résultats. L'intervalle de prévision répond à cet objectif.

4 Evaluation de l'erreur de mesure (exactitude) dans deux normes de validation de méthode

Ce paragraphe discute de l'évaluation de l'erreur de mesure selon deux normes bien connues des laboratoires. La première norme NF T 90-210 [2] est dédiée aux analyses de l'eau et elle propose une approche critère par critère de la validation. La deuxième norme NF V 03-110 [3] est dédiée au domaine de l'agro-alimentaire et elle propose une approche globale de la validation.

L'erreur de mesure est évaluée et vérifiée selon le même principe. A l'étape 1, un intervalle de valeurs, symétrique autour de la valeur moyenne, est construit :

$$IM = X_{moyen} \pm k \times S_{FI} \quad (4)$$

S_{FI} est l'écart-type de fidélité intermédiaire calculé selon la formule :

$$S_{FI} = \sqrt{S_r^2 + S_B^2} \quad (5)$$

Il combine l'écart-type de répétabilité S_r et l'écart-type S_B représentant l'effet des facteurs comme le temps, l'appareil, l'opérateur, ...

A l'étape 2, l'intervalle IM est comparé à l'intervalle des valeurs acceptables noté IA. Si IM est inclus dans IA, l'erreur de mesure n'est pas trop élevée et la méthode est validée.

4.1 Construction de l'intervalle IM

La méthode de construction de l'intervalle IM diffère dans les deux normes et conduit à des valeurs distinctes du coefficient k utilisé dans (4).

. IM selon NF T 90-210 : le biais de (mesure de) la méthode est d'abord vérifié. Dans le cas où il est acceptable, l'intervalle IM est construit avec $k = 2$. Le choix de cette valeur n'est pas expliqué dans la norme.

$$IM = X_{moyen} \pm 2 \times S_{FI} \quad (6)$$

. IM selon NF V 03-110 : IM est un intervalle statistique nommé intervalle de tolérance d'espérance β . Par exemple, β fixé à 80% signifie que les intervalles ainsi construits contiennent en moyenne 80% des valeurs. L'intervalle réellement construit lors de l'étude de validation contiendra une proportion plus faible ou plus élevée que 80%.

Dans cette norme, la valeur de k est le produit de deux termes :

- le coefficient t de la loi de Student qui dépend du nombre de degrés de liberté ν
- un facteur correctif appliqué à l'écart-type S_{FI}

$$k = t_\nu \times \sqrt{1 + \frac{1}{I \times J \times B^2}} \quad (7)$$

où :

$R = S_r^2 / S_B^2$ est le rapport des variances
 B^2 est un coefficient donné par :

$$B^2 = \frac{R + 1}{J \times R + 1} \quad (8)$$

I est le nombre de séries (modalités du facteur étudié)

J est le nombre de répétitions

ν est le nombre de degrés de liberté donné par (9) :

$$\nu = \frac{(R + 1)^2}{\frac{(R + \frac{1}{J})^2}{I - 1} + \frac{1 - \frac{1}{J}}{IJ}}$$

Cette expression de k résulte de deux approximations : l'utilisation de la loi de Student et le calcul des degrés de liberté ν (9) basé sur la méthode de Welch-Satterthwaite [9]. ν (9) n'est généralement pas un nombre entier et le quantile t de Student est obtenu en tronquant ν à l'entier le plus proche ou par interpolation linéaire.

Ainsi, IM est donné par :

$$IM = X_{moyen} \pm t_\nu \times \sqrt{1 + \frac{1}{I \times J \times B^2}} \times S_{FI} \quad (10)$$

Dans sa présentation, la norme utilise le concept d'écart-type de l'intervalle de tolérance pour désigner d'une manière globale le terme :

$$S_{IT} = \sqrt{1 + \frac{1}{I \times J \times B^2}} \times S_{FI} \quad (11)$$

Mais de l'avis de l'auteur, cette présentation qui simplifie les calculs, pour les programmer dans un tableur, ne facilite pas la compréhension de l'intervalle.

4.2 Simplifier le calcul de IM selon NF V 03-110

. Cas où la fidélité intermédiaire correspond à la répétabilité ($S_{FI} = S_r$) : il n'y a pas d'influence des facteurs considérés. Alors $R = 0$, $B^2 = 1$ et k devient :

$$k = t_v \times \sqrt{I + \frac{I}{I \times J}} \quad (12)$$

D'où l'intervalle IM :

$$IM = X_{moyen} \pm t_v \times \sqrt{I + \frac{I}{I \times J}} \times S_r \quad (13)$$

Il est important de noter que dans ce cas, l'écart-type de répétabilité S_r a été estimé avec l'ensemble des mesures $I \times J$. Le nombre de degrés de liberté :

$$v = I \times J - 1 \quad (14)$$

atteint ici sa valeur maximale.

L'expression (13) peut être reformulée en distinguant l'écart-type de la moyenne et l'écart-type d'une valeur individuelle :

$$IM = X_{moyen} \pm t_v \times \sqrt{S_r^2 + \frac{S_r^2}{I \times J}} \quad (15)$$

Le terme S_r^2 est la variance d'une valeur individuelle et le terme $S_r^2/I \times J$ est la variance de la moyenne. Comme la moyenne a été estimée, il est correct de considérer son incertitude lors de la construction de l'intervalle.

. Expliciter les composantes de l'écart-type : en substituant les deux variances S_r^2 et S_B^2 au rapport R et au coefficient B^2 , l'équation (10) se réécrit :

$$IM = X_{moyen} \pm t_v \times \sqrt{S_{FI}^2 + \frac{(S_B^2 + \frac{S_r^2}{J})}{I}} \quad (16)$$

Elle comprend la variance d'une valeur individuelle S_{FI}^2 et la variance de la moyenne donnée par :

$$\frac{(S_B^2 + \frac{S_r^2}{J})}{I} \quad (17)$$

4.3 Utiliser les caractéristiques d'un intervalle statistique

Généralement, un intervalle statistique dépend de nombreux facteurs. On peut citer : le nombre de mesures, le nombre de degrés de liberté, le caractère estimé de certaines composantes, ... Cette caractéristique est analysée dans le cas de l'intervalle IM (10) de la norme NF V 03-110 [3].

Dans le cas simple où la fidélité correspond à la répétabilité ($S_{FI} = S_r$), l'intervalle est évalué sans recourir à des approximations (calcul de k et du nombre de degrés de liberté). L'intervalle IM correspond alors à un intervalle de prévision (ou de prédiction) tel qu'il est défini dans la norme NF ISO 16269-8 [10]. Il est possible

d'utiliser les abaques de cette norme pour déterminer la valeur de k (12).

Autrement, la largeur de l'intervalle IM augmente avec l'effet du facteur représenté par S_B^2 . Plus S_B^2 est importante et plus le nombre de degrés de liberté diminue. Il tend vers une limite basse de valeur ($I - 1$).

Comme la difficulté de cette méthode réside principalement dans le calcul du nombre de degrés de liberté, on peut utiliser les deux valeurs extrêmes de $v - (I \times J - 1)$ et $(I - 1)$ - pour évaluer approximativement la largeur de l'intervalle.

Pour diminuer l'intervalle IM, le nombre et la répartition des mesures doivent être choisis en fonction de l'effet du facteur. Quand le facteur a peu d'effet, il faut augmenter le nombre de répétitions J et dans le cas contraire, il faut augmenter le nombre de séries I .

Ces actions n'influencent pas l'intervalle IM (6) de la norme NF T 90-210 car celui-ci ne dépend pas du nombre de mesures. Sous cet angle, on peut affirmer que (6) n'est pas un intervalle statistique.

4.4 Du besoin de justifier la construction de l'intervalle IM

L'intervalle IM (10) de la norme NF V 03-110 [3] est un intervalle de confiance pour la valeur d'une future mesure. Il représente le cas où le futur échantillon comporte une seule valeur dans la norme [10]. Si cet intervalle de confiance est inclus dans l'intervalle d'acceptabilité IA, alors la méthode produit des résultats valides. La valeur de $\beta = 80\%$ représente le niveau de confiance associé à cet intervalle.

L'intervalle IM (6) de la norme NF T 90-210 [2] n'est pas un intervalle statistique. Il ne correspond ni à l'intervalle de prévision ni à l'intervalle de dispersion ni à l'intervalle de confiance. Il est peut-être un intervalle d'incertitude ; mais il n'est pas possible d'analyser cette éventualité sans éléments d'information supplémentaires (par exemple, définir le mesurande considéré).

Quelle signification attribuée à cet intervalle au regard de la méthode à valider, des résultats produits lors de la validation et des futurs résultats ? En l'absence d'explication, il est légitime de s'interroger sur la fiabilité du processus de validation réalisé selon la norme NF T 90-210.

5 Conclusions et perspectives

Le concept d'« exactitude » est trop ambigu pour convenir à l'exercice pratique que représente la validation d'une méthode. Comme les analystes et les responsables techniques estiment généralement l'erreur de mesure à la place de l'exactitude, il est aisé d'adopter ce concept dans les référentiels dédiés à la validation. La révision des termes « justesse » et « exactitude » définis dans le VIM faciliterait cette tâche. A cet effet, l'auteur propose de

compléter leur définition avec des exemples illustratifs et une note prohibant l'utilisation quantitative de ces concepts.

Pour valider une méthode avec le critère de l'erreur de mesure, l'approche « profil d'exactitude » est complète et correcte : application de la théorie statistique, prise en compte de l'incertitude de la moyenne et du nombre de mesures à travers le nombre de degrés de liberté. L'intervalle IM (10) associé est expliqué et justifié.

Ce travail devrait être réalisé pour l'intervalle IM (6) construit selon la norme NF T 90-210 [2]. A défaut, il est permis de douter de la fiabilité de la validation réalisée avec cette méthode.

Enfin, l'erreur de mesure n'est pas le seul critère global utilisé pour valider une méthode. De nombreux laboratoires d'étalonnage estiment l'incertitude d'étalonnage sur la gamme de mesure considérée. Cette incertitude ne doit pas excéder une valeur spécifiée pour que la méthode soit validée.

Il serait intéressant de comparer ces deux critères de validation : erreur de mesure et incertitude. Comme il est nécessaire d'estimer l'incertitude (des résultats issus) de la méthode, on peut soulever la question de la coexistence de deux critères globaux pour valider une méthode.

Références

1. Vocabulaire International de métrologie – Concepts fondamentaux et généraux et termes associés (VIM), Août (2011)
2. Norme NF T 90-210 – Protocole d'évaluation initiale des performances d'une méthode dans un laboratoire (2009)
3. Norme NF V 03-110 – Analyse des produits agricoles et alimentaires – Protocole de caractérisation en vue de la validation d'une méthode d'analyse (2010)
4. Norme NF ISO 5725-1 – Exactitude (justesse et fidélité) des résultats et méthodes de mesure – Partie 1 : Principes généraux et définitions – Décembre (1994)
5. Norme NF EN ISO/CEI 17025 – Exigences générales concernant la compétence des laboratoires d'étalonnage et d'essais – Septembre (2005)
6. Eurachem – The Fitness for purpose of analytical methods – A laboratory guide to method validation and related topics – Second edition (2014)
7. X. Fuentes-Arderiu - “An inconsistency between definitions of “measurement accuracy” and “measurement error” in the VIM - *Accred Qual Assur*, **19**:241 (2014)
8. E. Prenesti, F. Gosmaro - “Trueness, precision, and accuracy: a critical overview of the concepts as well as proposals for revision” – *Accred Qual Assur*, **20**:33-40 (2015)
9. Robert W. Mee – β -Expectation and β -Content Tolerance Limits for Balanced One-Way ANOVA

Random Model – *Technometrics*, **VOL. 26**, N° 3, August (1984)

10. Norme NF ISO 16269 – 8 - Interprétation statistique des données – partie 8 : Détermination des intervalles de prédiction – Décembre (2004)