

Titre **Guide de validation des méthodes d'analyses**

Objet Ce guide est destiné aux laboratoires de l'Anses dans le cadre des travaux sur les méthodes d'analyse quantitatives ou qualitatives qu'ils développent, adaptent ou optimisent dans l'exécution de leurs missions de laboratoires de référence nationaux ou de l'union européenne. Il est accompagné : 1) d'un glossaire des termes relatifs à la validation des méthodes ; 2) d'un modèle type pour la rédaction de méthodes d'analyses, accompagné d'un guide d'utilisation ; 3) d'un modèle de rapport de validation.

	<i>Nom</i>	<i>Fonction</i>	<i>Date</i>	<i>Signature</i>
<i>Rédaction</i>	Groupe de travail « GT Val 2 » coordonné par Michel Laurentie	Rédacteurs	15 octobre 2015	
<i>Vérification</i>	Pascale Parisot	Directrice générale adjointe en charge des laboratoires Pilote du processus PR3	16 octobre 2015	
	Olivier Pierson	Délégué à la qualité	27 octobre 2015	
<i>Approbation</i>	Marc Mortureux	Directeur général	28 octobre 2015	

Date d'application **28 octobre 2015**

Nature de la modification Document initial

Diffusion pour application Laboratoires

Sommaire

Table des figures	4
Table des tableaux	5
1. Avant-propos	6
2 Contexte et objectifs du guide	6
3. Domaine d'application	7
3.1 Cycle de vie d'une méthode.....	7
3.2 Périmètre du guide.....	8
4. Documents de référence	9
5. Glossaire	10
6. Démarche de validation	12
6.1 Expression du besoin (cahier des charges du client/prescripteur).....	12
6.2 Caractérisation.....	14
6.2.1 Caractérisation par approche modulaire	14
6.2.2 Approche globale versus approche par caractéristique par caractéristique.....	15
6.2.3 Caractérisation absolue versus caractérisation relative	16
6.2.4 Caractérisation intra-laboratoire versus interlaboratoires	16
6.3 Validation.....	17
7. Caractérisation des performances des méthodes	17
7.1 Prérequis statistiques.....	17
7.1.1 Données qualitatives	17
7.1.1.1 Type de données	17
7.1.1.2 Tableau de contingence.....	18
7.1.2 Méthodes quantitatives	18
7.2 Les supports de caractérisation	18
7.2.1 Types de supports de caractérisation.....	18
7.2.1.1 Qualité	18
7.2.1.2 Représentativité	19
7.2.2 Les supports de caractérisation « positifs »	20
7.2.3 Les supports de caractérisation « négatifs ».....	20
7.2.3.1 Les négatifs ou blancs	20
7.2.3.2 Les analytes non cibles.....	20
7.3 Caractéristiques de performance à déterminer selon le type de méthode.....	21
7.4 Modalités pratiques de caractérisation.....	22
7.4.1 Spécificité	22
7.4.2 Sensibilité	22
7.4.3 Fonction d'étalonnage/Efficacité (PCR)	23
7.4.4 Linéarité.....	23



7.4.5	Justesse	23
7.4.6	Fidélité	23
7.4.7	Répétabilité.....	24
7.4.8	Fidélité intermédiaire	24
7.4.9	Reproductibilité	25
7.4.10	Exactitude	25
7.4.11	Limites	25
7.4.11.1	Limite de détection	25
7.4.11.2	Limite de quantification	26
7.4.12	Autres types de limites.....	26
7.5	Outils de calculs pour la validation.....	26
8	Adéquation de la performance et validation	27
8.1	La démarche.....	27
8.2	Modèle de méthode d'analyse	27
8.3	Rapport de validation	27
9	Incertitude.....	28
10	Perspectives	31
11	Références bibliographiques	31
	Annexe « Membres du groupe GT Val 2 »	32
	Annexe « Bases statistiques pour les variables quantitatives».....	33
	Annexe « Spécificité ».....	36
	Annexe « Sensibilité ».....	38
	Annexe « Fonction d'étalonnage/Efficacité (PCR)»	39
	Annexe « Justesse »	41
	Annexe « Fidélité »	43
	Répétabilité.....	43
	Fidélité intermédiaire.....	44
	Reproductibilité	44
	Annexe « Exactitude »	46
	Exemple d'application : dosage de l'acrylamide dans le plasma de porc.....	48
	Annexe « Limites »	53
	Limite de détection.....	53
	Limite de quantification	56
	Annexe « Incertitude »	57
	Utilisation de l'incertitude élargie dans la prise de décision	57
	Calcul d'incertitude	58
	Modélisation de l'incertitude.....	59



Table des figures

Figure 1 - Cycle de vie d'une méthode d'analyse	7
Figure 2 - Illustration des notions de séries, niveaux et répétitions	12
Figure 3 - Schéma de l'expression du besoin et de ses conséquences entre le client / laboratoire	13
Figure 4 - Illustration de la méthodologie de caractérisation des caractéristiques de performance étudiées séparément	15
Figure 5 - Illustration de la méthodologie de caractérisation des caractéristiques de performance basée sur un critère unique	16
Figure 6 - Composantes croissantes de la variabilité : répétabilité, fidélité intermédiaire et reproductibilité	24
Figure 7 - Processus d'évaluation de l'incertitude	28
Figure 8 - Principales sources d'incertitudes et chapitres correspondants dans la norme NF EN ISO/CEI 17025	29
Figure 9 - Exemple d'informations fournies par une représentation graphique « box plot »	33
Figure 10 - Exemple de variances non homogènes	34
Figure 11 - Choix du facteur de pondération (w) en fonction de la relation entre la variance et le niveau de concentration	34
Figure 12 - Hypothèses statistiques requises pour les variables quantitatives	35
Figure 13 - Exemple de distributions de résidus (valeurs trouvées - valeurs attendues) obtenues pour différents modèles d'étalonnage	40
Figure 14 - Exemple de droite de justesse et son intervalle de confiance	42
Figure 15 – Illustration d'un profil d'exactitude	47
Figure 16 – Dosage d'acrylamide dans le plasma de porc : données brutes observées	49
Figure 17 - Profil d'exactitude réalisé avec les données brutes	50
Figure 18 - Droite de régression entre les concentrations introduites et les concentrations en retour	50
Figure 19 - Profil d'exactitude réalisé avec les données corrigées	51
Figure 20 - Importance de l'incertitude dans la prise de décision	57
Figure 21 - Modélisations par une fonction puissance du profil d'incertitude	60



Table des tableaux

Tableau 1 - Synthèse des résultats obtenus sous forme de tableau de contingence	18
Tableau 2 - Caractéristiques de performance à déterminer pour les méthodes qualitatives et les méthodes quantitatives	21
Tableau 3 - Synthèse tabulaire des caractéristiques de performances, de leurs valeurs prédéterminées, de leurs valeurs obtenues et de la décision sur leur validité	27
Tableau 4 - Synthèse des résultats obtenus lors de l'étude de caractérisation des performances	36
Tableau 5 - Intervalle de confiance à 95% d'un pourcentage de Sp	37
Tableau 6 Synthèse des résultats obtenus lors de l'étude de caractérisation des performances	38
Tableau 7 - Modèles usuels de fonctions d'étalonnages, calcul du modèle inverse, nombre minimal de niveau nécessaires, de répétitions par niveaux et nombre minimal de séries pour établir ce modèle	39
Tableau 8 - Différentes expressions de la justesse	41
Tableau 9 - Résultats des calculs du profil d'exactitude après correction des données brutes	52
Tableau 10 - Nombre de répétitions positives par niveau	54
Tableau 11 - Ensemble des performances obtenues pour déterminer les teneurs d'acrylamide dans le plasma de porc	59



1. Avant-propos

Le présent guide a été rédigé par le groupe de travail « GT Val 2 » constitué sur décision du directeur général n°2013/005. Les participants sont listés dans l'annexe « Membres du groupe GT Val 2 ». Dans le cadre de ce GT, une attention particulière a été donnée à la représentation de chacun des domaines d'activité de l'Agence.

Ce guide est destiné à l'usage des laboratoires de l'Anses dans le cadre des travaux de développement, d'adaptation ou d'optimisation de méthodes d'analyse résultant de leurs missions de laboratoires de référence nationaux ou de l'Union européenne.

Ce guide donne des lignes directrices générales sur le processus de validation d'une méthode, propose des modalités pratiques de caractérisation des performances, en présentant pour chaque caractéristique des exemples concrets. Ce guide ne se substitue pas aux référentiels réglementaires ou normatifs applicables aux différents domaines d'activité de l'Agence et utilisés dans les laboratoires.

Pour l'élaboration de ce document et conformément à son mandat, le groupe de travail s'est attaché à prendre en compte les différents types de méthodes (qualitative et quantitative) et à exploiter les documents déjà en vigueur au sein de l'Agence, sur la base d'un recensement effectué auprès des laboratoires. Il a pris en compte les plus récentes évolutions normatives sur ce sujet et a veillé à la compatibilité de ses recommandations avec les règles de l'accréditation proposées par le Comité français d'accréditation (Cofrac).

La parution de ce guide s'accompagne de la mise à disposition de documents associés qui ont été élaborés parallèlement par le groupe de travail, afin de faciliter la mise en application de certaines de ses recommandations sur une base harmonisée. Il s'agit :

- d'un glossaire des termes relatifs à la validation des méthodes – ANSES/PR3/7/01-01 ;
- d'un modèle-type pour la rédaction de méthodes d'analyses en vue de leur publication – ANSES/PR3/7/01-04, accompagné d'un guide d'utilisation – ANSES/PR3/7/01-05 ;
- d'un modèle de rapport de validation de méthode – ANSES/PR3/7/01-02.

Ce guide ainsi que ses documents associés sont destinés à être révisés régulièrement pour prendre en compte les évolutions normatives et les retours d'expérience des laboratoires de l'Anses.

2 Contexte et objectifs du guide

Dès sa création en 2011, le collège de la référence de l'Anses a identifié le thème de la validation des méthodes d'analyse parmi les sujets nécessitant un partage des pratiques et la définition d'orientations communes pour les laboratoires de l'Anses. Un premier groupe de travail a proposé un processus général de développement de méthode adapté au contexte de la référence, la définition d'une terminologie commune et la publication des méthodes destinées à être rendues officielles sur le site de l'Agence.

Ces travaux avaient également mis en évidence la nécessité d'approfondir plusieurs sujets, parmi lesquels figurent les aspects suivants :

- les modalités de caractérisation des performances d'une méthode d'analyse ;
- la constitution harmonisée des dossiers de validation ;
- la détermination d'une trame commune pour les méthodes d'analyses de l'Anses en vue de leur « publication » (au sens mise à disposition), après officialisation le cas échéant.



L'enquête de satisfaction menée en 2012 auprès des laboratoires agréés a confirmé le bien-fondé de ces axes complémentaires de travaux. En effet, ces laboratoires ont exprimé plusieurs attentes de la part des laboratoires nationaux de référence (LNR) concernant les modalités de communication et d'information des méthodes qui leur sont transférées : contenu des dossiers de validation, incluant leur adéquation avec les exigences du Cofrac ; modalités de transfert et d'appui technique ; modalités de diffusion des protocoles...

Le présent guide a pour objectif d'harmoniser au sein de l'Anses, dans le contexte des missions attachées aux mandats de référence, la démarche de mesure des performances (ou caractérisation) des méthodes analytiques, de validation de ces performances et de publication des méthodes ainsi validées.

Il ne présente que des recommandations dans le cadre de conditions idéales de caractérisation et qui peuvent ou doivent être adaptées en fonction des conditions réelles de terrain, de disponibilité de matériel, de couple hôte/matrice, le cas échéant de l'urgence de l'évaluation/validation, et des référentiels en vigueur dans les différents domaines d'activité...

3. Domaine d'application

3.1 Cycle de vie d'une méthode

La vie d'une méthode d'analyse est un processus évolutif qui suit différentes étapes [Hérou V. et al. (2013), Feinberg (2009)] pouvant être représentées par un cycle (figure 1) :

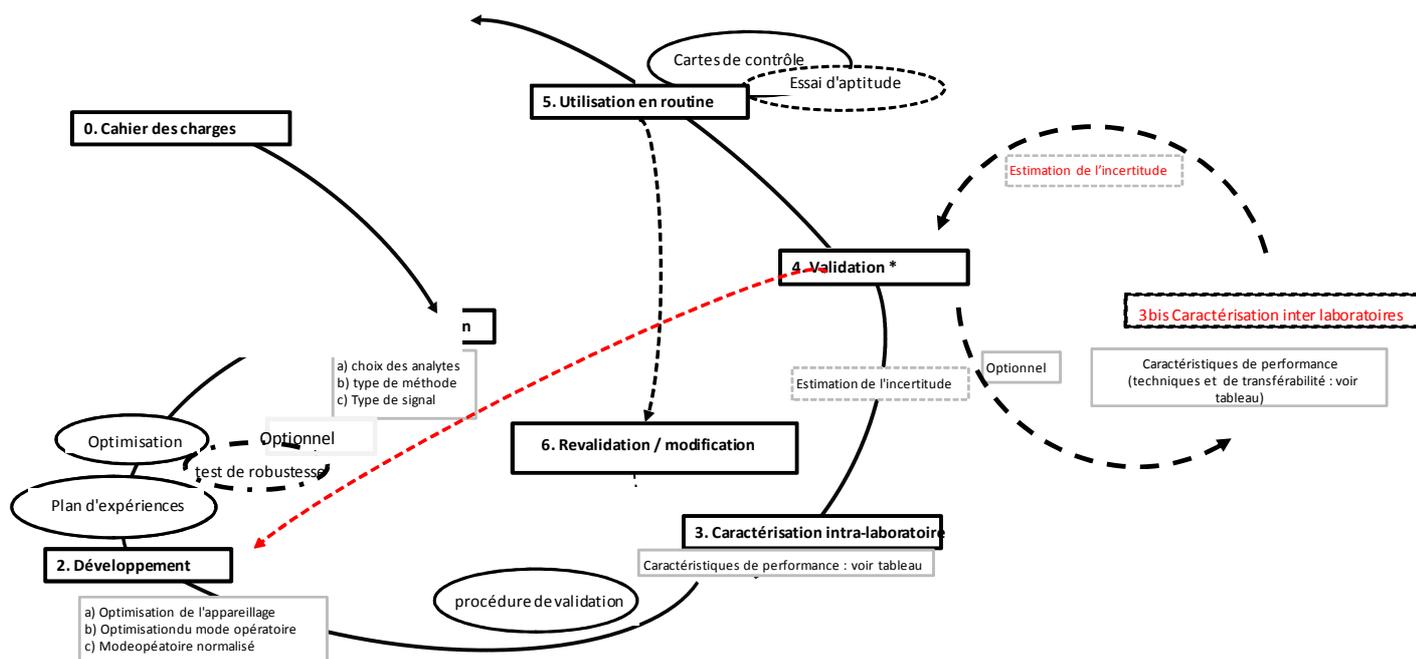


Figure 1 - Cycle de vie d'une méthode d'analyse.
(D'après Max Feinberg, 2009, modifié par le groupe GT Val 2)

Les principales étapes de ce cycle sont les suivantes :

- étape préliminaire (étape « 0 ») : expression du déploiement et de la mise en œuvre d'une méthode pour un client/prescripteur et/ou pour un usage spécifié sous la forme d'un cahier des charges ;
- étape 1 : phase de sélection des outils, des analytes, ... ;

- étape 2 : phase de développement / optimisation de la méthode (en l'absence de méthode préexistante pour le domaine d'application considéré) ;
- étape 3 : caractérisation intra-laboratoire et, au besoin, inter-laboratoires, de la méthode ;
- étape 4 : validation de la méthode développée au regard de l'usage attendu (voir étape 0)¹.

À l'issue de ces étapes, l'utilisation en routine de la méthode peut être envisagée (étape 5). La revue périodique de la méthode peut donner lieu à un besoin de revalidation ou d'un nouveau développement (étape 6).

3.2 Périmètre du guide

Le présent guide vise à présenter la démarche de validation, et en particulier :

- les modalités de définition des caractéristiques de performance ciblées, aussi appelées critères de performance : cahier des charges / revue de contrat avec le client / prescripteur (étape 0) ;
- les modalités selon lesquelles les méthodes peuvent être caractérisées en intra-laboratoire et / ou en inter-laboratoires en particulier (étape 3 et 3bis) :
 - la liste des caractéristiques de performance à estimer selon le type de méthode (qualitatif / quantitatif) ;
 - les modalités selon lesquelles ces caractéristiques peuvent être estimées ;
- la validation stricto sensu par comparaison des valeurs des caractéristiques de performance aux critères cibles prédéterminés (ou valeurs d'acceptabilité) (étape 4).

Il s'applique aux laboratoires de l'Anses qui ont à réaliser des études de validation pour les méthodes :

- qu'ils conçoivent et développent ;
- qu'ils adaptent à partir d'une méthode développée en externe si cette méthode est reconnue ;
- qu'ils mettent en œuvre sur la base d'une méthode externe, si cette méthode n'est pas reconnue (au sens du document Cofrac LAB REF 08, § 2.1).

Il s'applique aux méthodes destinées à être utilisées en interne ou à être transférées à d'autres laboratoires.

Il s'applique à l'ensemble des domaines couverts par les laboratoires de l'Anses, tout en respectant les exigences particulières de ces domaines.

Le présent guide vise à satisfaire aux exigences du paragraphe 5.4.5 de la norme NF EN ISO / CEI 17025 « Validation des méthodes » et au paragraphe 4.3.6 de la norme NF EN ISO / CEI 17043. La mise en application des autres exigences de la norme NF EN ISO / CEI 17025, en particulier la qualification du matériel et du personnel (chapitres 5.2 et 5.5) sont un prérequis à l'évaluation de la performance de la méthode analytique et à leur validation. Ils font partie du processus qui contribue à la maîtrise de la performance d'une méthode analytique.

Ce guide ne traite ni du développement, ni de l'optimisation des méthodes d'analyse (étape 2).

De même, ce guide ne traitera pas de la réalisation et de l'analyse statistique des Essais Inter-Laboratoires d'Aptitude (EILA) qui sont spécifiquement décrites dans le guide Anses dédié aux statistiques pour les EILA (ANSES/PR3/9/01).

¹ L'astérisque en étape 4 matérialise le fait qu'une première validation doit être entreprise à l'issue de la caractérisation intra-laboratoire avant d'engager la phase de caractérisation inter laboratoires par un Essai Inter-Laboratoire de Validation (EILV).



4. Documents de référence

- NF EN ISO / CEI 17025 : 2005 Exigences générales concernant la compétence des laboratoires d'étalonnages et d'essais (Indice de classement : X 50-061)
- NF EN ISO / CEI 17043 : 2010 Évaluation de la conformité. Exigences générales concernant les essais d'aptitude (Indice de classement : X 50-096)
- NF EN FDIS 16140-2 Microbiologie de la chaîne alimentaire — Validation des méthodes — Partie 2: Protocole pour la validation de méthodes alternatives (commerciales) par rapport à une méthode de référence.
- NF T 90-210 : 2009 Qualité de l'eau – Protocole d'évaluation initiale des performances d'une méthode dans un laboratoire
- NF V 03-110 : 2010 Analyse des produits agricoles et alimentaires – Protocole de caractérisation en vue de la validation d'une méthode d'analyse quantitative par construction du profil d'exactitude
- NF ISO 3534-1 : 2007 Statistiques. Vocabulaire et symboles. Partie 1 : Termes statistiques généraux et termes utilisés en calcul de probabilité. (Indice de classement : X 06-002-1)
- NF ISO 3534-2 : 2005 Statistiques. Vocabulaire et symboles. Partie 2 : statistiques appliquées (Indice de classement : X 06-002-6)
- NF ISO 5725-1 : 1994 Application de la statistique – Exactitude (justesse et fidélité) des résultats et méthodes de mesure – Partie 1 : Principes généraux et définitions (NF X 06-041-1).
- NF ISO 5725-2 : 1994 Application de la statistique – 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 (NF X 06-041-2).
- NF ISO 5725-3 : 1994 Application de la statistique – Exactitude (justesse et fidélité) des résultats et méthodes de mesure – Partie 3 : Mesures intermédiaires de la fidélité d'une méthode de mesure normalisée (NF X 06-041-3).
- NF ISO 5725-4 : 1994 Application de la statistique – Exactitude (justesse et fidélité) des résultats et méthodes de mesure – Partie 4 : Méthodes de base pour la détermination de la justesse d'une méthode de mesure normalisée (NF X 06-041-4).
- NF U 47-600-1 : Méthodes d'analyse en santé animale PCR : Exigences et recommandations pour la mise en œuvre de la PCR en santé animale (Février 2015).
- NF U 47-600-2 : Méthodes d'analyse en santé animale PCR : Exigences et recommandations pour le développement et la validation de la PCR en santé animale (Février 2015).
- ISO 78-2 :1999 Chimie- plans de normes – Partie 2 : Méthodes d'analyse chimique
- VIM ou Guide ISO / CEI 99 : 2012 (JCGM 200) Vocabulaire international de métrologie – Concepts fondamentaux et généraux et termes associés.
- Guide GUM (JCGM 100) Evaluation of measurement data — Guide to the expression of uncertainty in measurement, 2008.
- Manuel des tests de diagnostic et des vaccins pour les animaux terrestres « Section 3-6 Lignes directrices de validation » 2014, OIE
- EPPO PM7/98. Specific requirements for laboratories preparing accreditation for a plant pest diagnostic activity Bulletin de l'OEPP, Volume 40, Issue 1, pages 5–22, Avril 2010
- EPPO PM7/76. Use of EPPO diagnostic protocols. Bulletin de l'OEPP. Volume 40, Issue 3, <<http://onlinelibrary.wiley.com/doi/10.1111/epp.2010.40.issue-3/issuetoc>> pages 350–352, Décembre 2010



AFNOR. GA V03-042 Guide d'application aux normes NF EN ISO 24276, NF EN ISO 21569, NF EN ISO 21570, NF EN ISO 21571 et leurs amendements 1. Méthodes d'analyses pour la détection des organismes génétiquement modifiés et produits dérivés. Décembre 2013.

ENGL / EURL for GM food and feed. Definition of minimum performance requirements for analytical methods of GMO testing. Avril 2015.

http://gmo-crl.jrc.ec.europa.eu/doc/MPR%20Report%20Application%2020_10_2015.pdf

XP CEN ISO/TS 19036 : Août 2006 « Microbiologie des aliments - Lignes directrices pour l'estimation de l'incertitude de mesure pour les déterminations quantitatives »

Amendé par : XP CEN ISO / TS 19036 / A1 : Février 2009 « Microbiologie des aliments - Lignes directrices pour l'estimation de l'incertitude de mesure pour les déterminations quantitatives - AMENDEMENT 1 : incertitude de mesure sur les faibles taux »

NF ISO 21748 : Décembre 2010 « 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 »

5. Glossaire

Les définitions des termes introduits dans le présent guide sont données dans le glossaire ANSES/PR3/7/01-01. Cependant, pour un confort de lecture de ce guide, les principales notions sont reprises ci-après :

Caractérisation : La caractérisation est « l'estimation de la valeur, du nombre, de l'importance ou de la grandeur des choses ». La démarche de caractérisation se conçoit essentiellement dans un but de décision, mais la décision ne fait pas partie de la caractérisation.

Ainsi, la distinction entre la caractérisation d'une méthode et sa validation procède de la même logique que celle de la séparation classiquement retenue entre évaluation du risque et gestion du risque (article 3 du règlement (CE) n°178 / 2002).

La caractérisation d'une méthode est la détermination ou l'estimation statistique des caractéristiques de performance de cette méthode.

La performance des méthodes peut être caractérisée par :

- des caractéristiques techniques : elles caractérisent la méthode au regard des résultats donnés par la méthode par comparaison avec le statut réel (ou supposé) des échantillons. Exemples de caractéristiques de performance techniques des méthodes : la sensibilité, la spécificité, l'exactitude, la répétabilité et la reproductibilité... (voir tableau 2) ;
- des caractéristiques de mise en œuvre (coût, simplicité, rapidité ou délai...) : elles visent à définir si la méthode proposée peut être facilement utilisée en routine et / ou transférée à d'autres laboratoires.

Validation : Selon la norme NF EN ISO / CEI 17025, il s'agit de la « confirmation par examen et l'apport de preuves objectives du fait que les exigences particulières en vue d'une utilisation prévue déterminée sont remplies ». En fait, selon la norme U47-600-1, cette confirmation « consiste à comparer les valeurs des critères de performance déterminées au cours de l'étude de caractérisation de la méthode à celles attendues ou assignées au préalable (limites d'acceptabilité, objectifs à atteindre), puis à déclarer la méthode d'analyse valide ou non valide ».

Caractéristiques : Qualité qui peut être attribuée à une méthode d'analyse et, éventuellement, quantifiée ou qualifiée par un ou plusieurs paramètres de validation.

Note : il est dit qu'une méthode est spécifique, fidèle, répétable, robuste mais aussi facile d'application, écoresponsable, rapide etc.



Paramètres : Grandeur à estimer à partir d'une série de mesures.

Note 1 : Le biais absolu, l'écart-type de répétabilité, les coefficients d'une courbe d'étalonnage, la moyenne sont des paramètres, alors qu'une valeur de référence est une donnée.

Note 2 : En statistique, un paramètre désigne une quantité choisie pour expliciter les caractéristiques d'une population (ex : moyenne, écart-type, médiane, proportion... en tant que quantité réelle théorique, c'est-à-dire portant sur une population). Dans un modèle mathématique, on parlera plutôt de coefficient du modèle (voir la définition NF ISO 3534-1, 2.9).

Note 3 : À partir d'un échantillon de valeurs numériques, on peut estimer les coefficients d'un modèle soit de façon ponctuelle, soit par intervalle de confiance.

Note 4 : Un ou plusieurs paramètres de validation peuvent définir une caractéristique de performance unique (comme pour établir un profil d'exactitude). La plupart sont des moyennes et des écarts-types qui servent notamment au calcul d'intervalles de dispersion ou d'incertitude.

Note 5 : Un paramètre de validation peut aussi servir à effectuer un test de comparaison ou un test d'équivalence.

Critère (de validation) : valeur limite spécifiée pour une caractéristique de performance (NF ISO 3534-2 paragraphe 3.1.3).

Note 1 : Un critère peut être quantitatif ou qualitatif. Par exemple, pour la linéarité ou la spécificité l'objectif est du type conforme / non conforme, alors que pour un coefficient de variation de la répétabilité, il est fixé comme un pourcentage.

Note 2 : Plusieurs documents officiels définissent des critères en utilisant des termes spécifiques.

- a) Au niveau européen, la Décision Européenne 657 / 2002 introduit les Limites de performance minimum requises (LPMR) ;
- b) Au niveau international, la Commission CCMAS du *Codex Alimentarius* définit l'approche par critères ;
- c) Au niveau français, la norme NF V 03-110 parle de limite d'acceptabilité et la norme NF T 90-210 d'écart maximum acceptable ;
- d) dans le domaine des OGM, le réseau ENGL a établi des lignes directrices concernant les critères minimums de performance des méthodes analytiques avec des valeurs cibles acceptables pour chacune des caractéristiques évaluées (Hougs, 2011).

Exemples : La fidélité d'une méthode de quantification des résidus de médicaments doit avoir, selon la réglementation européenne, un coefficient de variation de la répétabilité $\leq 20\%$ et une justesse comprise entre 70% et 120%.

Méthode qualitative : méthode qui, à partir d'une quantité déterminée d'échantillon, permet la mise en évidence de la présence d'un analyte et fournit une réponse en termes de présence/absence.

Méthode quantitative : méthode d'analyse qui mesure la quantité ou la fraction pondérale d'un analyte de manière à pouvoir l'exprimer sous forme de valeur numérique dans les unités appropriées.

Note : Le terme méthode semi-quantitative est parfois employé. Une méthode semi-quantitative est une méthode qualitative utilisant une variable qualitative ordonnée croissante ou décroissante (exemple : dilution décroissante $\frac{1}{2}$; $\frac{1}{4}$; $\frac{1}{8}$; $\frac{1}{16}$; $\frac{1}{32}$...). L'utilisation de ce terme conduit à des confusions et il convient de ne pas l'utiliser au profit du terme « méthode qualitative ordonnée ».

Séries, niveaux et répétitions : les plans d'expériences réalisés impliquent l'utilisation de termes tels que séries, niveaux et répétitions. Ces termes sont définis dans le glossaire et représentés graphiquement sur la figure 2.



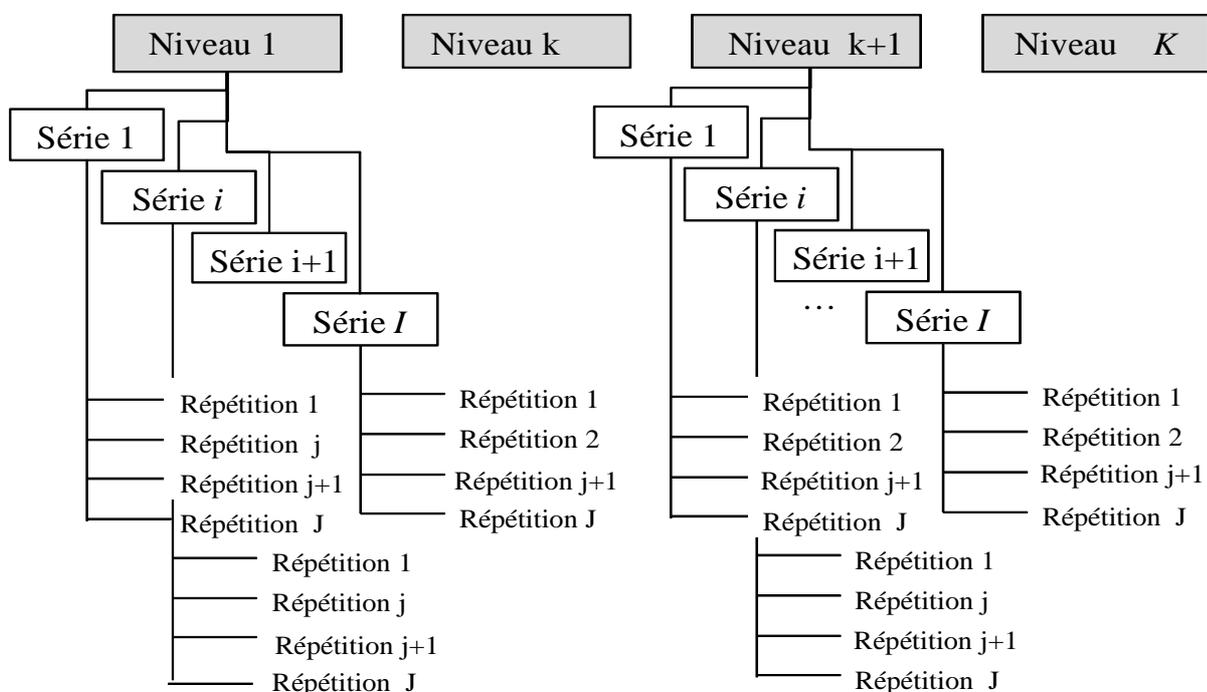


Figure 2 - Illustration des notions de séries, niveaux et répétitions.

Note : Autant que possible, les échantillons utilisés pour la caractérisation des performances des méthodes doivent être préparés de façon indépendante, de façon à limiter la propagation d'une erreur systématique et donc de minimiser l'incertitude.

6. Démarche de validation

La validation d'une méthode d'analyse correspond à une reconnaissance de son aptitude à satisfaire à l'usage attendu en routine. Elle s'effectue en confrontant des valeurs observées de ses caractéristiques de performance avec les valeurs des critères de validation pour les conditions dans lesquelles la méthode est utilisée. Le cas échéant, les caractéristiques de mise en œuvre pourront également être prises en compte pour la validation d'une méthode.

La démarche de validation repose donc sur 3 grandes étapes :

- la définition des critères de validation, ou caractéristiques de performance ciblées, en tant que traduction de l'expression des besoins du client/prescripteur ;
- la caractérisation de la méthode ;
- la validation *stricto sensu*.

6.1 Expression du besoin (cahier des charges du client/prescripteur)

Parmi les missions des laboratoires nationaux de référence telles que listées dans le Code rural et de la pêche maritime (CRPM, article R202-5), figure notamment le développement ou l'optimisation de méthodes d'analyse et l'appui scientifique et technique à la tutelle. Dès lors, les besoins méthodologiques à des fins analytiques officielles de l'État sont adressés aux LNR. Ces deux missions relèvent également des missions dévolues aux laboratoires de référence de l'Union européenne (LRUE), selon le règlement 882/2004, Article 32, le prescripteur des LRUE étant la DG SANTÉ.

Partant du principe que les méthodes (officielles) doivent être aptes à l'usage attendu en routine pour pouvoir être validées (voir ci-dessous), la phase de concertation préalable entre le LNR et le prescripteur est une phase indispensable à la réussite de tout projet analytique. Il convient, lors de



cette première étape, de définir précisément les attentes explicites et implicites du client, qui peuvent souvent se déduire de l'utilisation prévue pour la méthode. À titre d'exemples, on peut citer les méthodes suivantes :

- ✓ utilisées pour des analyses en gestion de foyers de maladies animales ou végétales, ou pour l'investigation de toxi-infections alimentaires, par des laboratoires de 1^{ère} intention, et qui pourront être adaptées à ces usages si elles permettent d'obtenir des résultats rapides, de façon peu onéreuse, et si l'objectif est de pouvoir obtenir des résultats sur des grands nombres d'échantillons et rapidement afin de définir le périmètre d'infection ;
- ✓ utilisées à l'import, pour la détection d'un organisme nuisible de quarantaine non présent sur le territoire ; elles devront être les plus sensibles possible pour éviter l'introduction d'un organisme de quarantaine et donner des résultats relativement rapides pour permettre une libération des lots de marchandises consignées ;
- ✓ utilisées exclusivement pour des analyses de confirmation ou de caractérisation plus poussée des agents pathogènes, elles devront alors présenter les meilleures caractéristiques de performance possibles (telles que limite de détection ou quantification, sensibilité / spécificité), même si le coût analytique est élevé ou si la mise en œuvre requiert un haut niveau de technicité dans les laboratoires de référence.

Ce travail de concertation entre le client/prescripteur et le prestataire vise donc à traduire les besoins du client / prescripteur en objectifs précis de performance de la méthode, notamment en ce qui concerne les taux de faux négatifs (sensibilité) et faux positifs (spécificité), limites de détection / quantification, exactitude acceptable *a priori*, et d'aboutir ainsi à la formalisation d'un « cahier des charges de la méthode ».

La figure 3 illustre cette formalisation du cahier des charges :

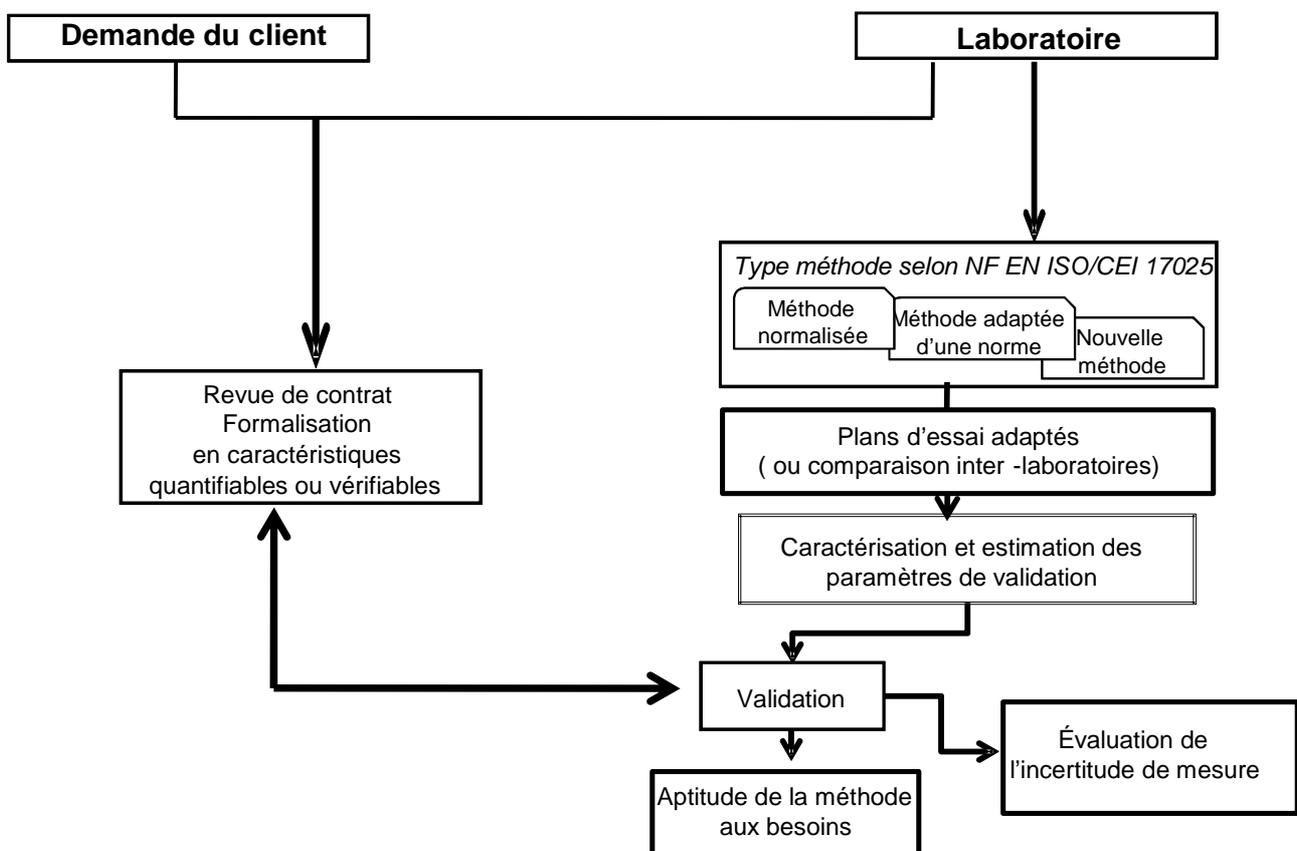


Figure 3 - Schéma de l'expression du besoin et de ses conséquences entre le client / laboratoire.
(Communication personnelle de Max Feinberg, modifié par le GT Val 2)

Toutefois, comme les exemples précédents ont pu le montrer, au-delà des caractéristiques de performance techniques, d'autres caractéristiques (rapidité, coûts, délais, simplicité...) entrent en ligne de compte dans la définition de l'aptitude d'une méthode à son usage. Cet usage peut d'ailleurs orienter le LNR vers une méthode plus qu'une autre, dans un contexte où il est souvent amené à faire un choix et un compromis entre plusieurs caractéristiques (ex : sensibilité vs. spécificité).

Souvent, en se fondant sur sa connaissance du pathogène, de l'épidémiologie, du réseau des laboratoires concernés, le LNR est en mesure de traduire les besoins de son commanditaire sans qu'une discussion formelle soit nécessaire. Dans ce cas, il est néanmoins recommandé que le LNR obtienne un accord du commanditaire sur le « cahier des charges ». L'aptitude des méthodes à atteindre les limites et critères définis dans la réglementation constitue nécessairement un besoin implicite du prescripteur que constituent les autorités compétentes.

Note : Exemples de formalisation de besoins, implicites ou explicites :

- Exigences réglementaires ou normatives : Règlement CE 2002/657, Règlement CE 2073/2005, NF U 47-600, Norme NF EN ISO 16140-partie 2, guide GA V03-042 (décembre 2014), lignes directrices 3.6 OIE sur la validation (2014).
- Santé des végétaux : les besoins généraux du client (DGAI) sont exprimés à travers un cahier des charges pour la validation des méthodes d'analyse officielles signé conjointement par le client (DGAI / SDQPV) et le prestataire (le LNR). Si le contenu reste adaptable à chaque cas particulier en fonction de l'organisme nuisible concerné, sa situation épidémiologique, les demandes urgentes, etc., il fixe un cadre général précisant les caractéristiques de choix des méthodes à sélectionner selon leur usage.
- OGM : lignes directrices de l'ENGL fixant les critères minimum de performance des méthodes analytiques (ENGL / EURL for GM food and feed, avril 2015).

D'autres aspects peuvent être envisagés dans la revue de contrat comme les contraintes de délais, les ressources nécessaires et disponibles. Cependant, ils ne sont pas considérés dans ce guide.

6.2 Caractérisation

Plusieurs approches sont possibles mais l'évaluation des caractéristiques de performance est généralement réalisée « critère par critère » ou de façon « globale ». Il existe également la possibilité de faire une approche dite « modulaire », c'est-à-dire des validations par étapes du processus analytique, comme d'abord l'étape d'extraction puis l'étape de détection.

6.2.1 *Caractérisation par approche modulaire*

En règle générale, pour satisfaire aux besoins de son client, le laboratoire est amené à devoir caractériser un processus analytique dans son intégralité, i.e. depuis la réception de l'échantillon jusqu'au résultat final. Dans ce cas, la méthode couvre l'intégralité des étapes du processus analytique et la caractérisation portera sur ce périmètre (taux de faux positifs/négatifs, teneur en analyte,...).

Dans certains domaines et certains cas, il peut être intéressant de procéder à une caractérisation des étapes du processus de manière séparée. Cette approche sera particulièrement utile par exemple lorsque plusieurs analyses font appel à une même technique sur un même échantillon à la suite d'une étape de préparation commune. C'est l'approche dite « modulaire », au sens « par module » (Holst-Jensen and Berdal, 2004).



Par exemple, pour un échantillon, réaliser une extraction d'ADN de p prises d'essai, puis n PCR sur chaque extrait d'ADN : dans ce cas, le module « extraction » pourra être évalué / caractérisé spécifiquement avant évaluation / caractérisation de chaque système PCR pour ce type d'extrait.

À signaler que cette approche modulaire, pour tentante qu'elle soit, peut présenter des risques ; les étapes pouvant interagir, au sens des interactions rencontrées dans les plans d'expérience. Donc, dans l'hypothèse où un laboratoire applique ce type d'approche, il convient de procéder si possible à une vérifier la compatibilité des modules entre eux et ainsi de valider globalement le processus analytique retenu.

6.2.2 Approche globale versus approche par caractéristique par caractéristique

Ce document présente les modalités de détermination des caractéristiques de performance d'une méthode. Un plan d'expérience unique au laboratoire peut permettre la détermination simultanée de plusieurs caractéristiques de performance. La possibilité de les traiter séparément existe également.

Plusieurs textes réglementaires ou normatifs définissent une approche de caractérisation fondée sur des caractéristiques de performance et leurs valeurs limites, sans toutefois préciser les moyens pour obtenir ces caractéristiques. Cette approche consiste à traiter séparément chaque caractéristique et elle peut être qualifiée de méthodologie « caractéristique par caractéristique ». Des textes normatifs (NF ISO 5725) décrivent les approches statistiques utilisées pour estimer ces caractéristiques de performances.

La figure 4 illustre une telle démarche :

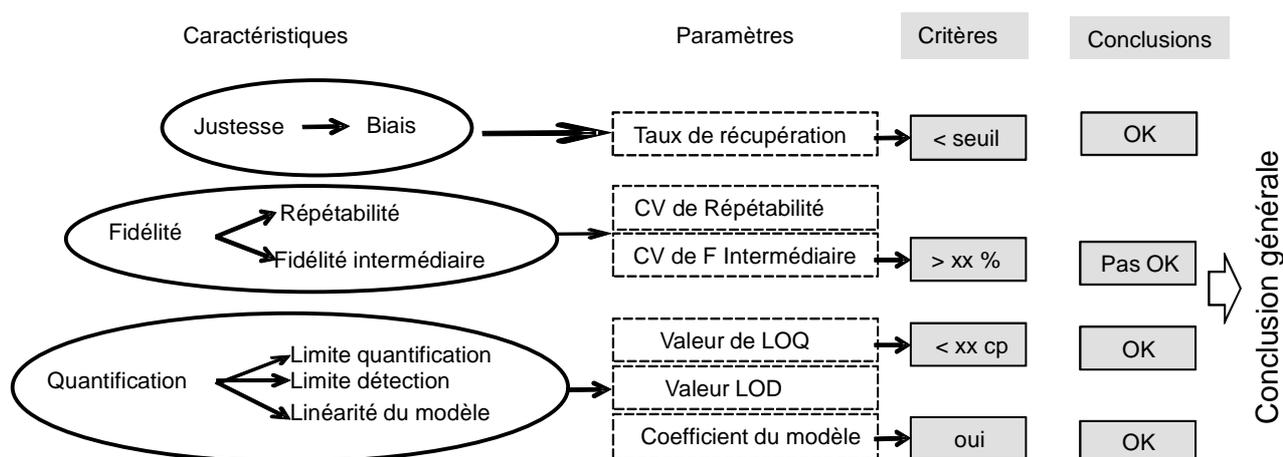


Figure 4 - Illustration de la méthodologie de caractérisation des caractéristiques de performance étudiées séparément.

(Communication personnelle de Max Feinberg, modifiée par le GT Val 2)

Cette approche fournit des informations intéressantes mais elle présente parfois comme inconvénient majeur de déboucher sur des conclusions contradictoires. Par exemple, certaines caractéristiques sont valides alors que d'autres ne le sont pas et il devient difficile de statuer sur la validation globale de la méthode analytique.

Par opposition, une caractérisation globale peut être réalisée sur une seule caractéristique de performance et la validité de la méthode sera statuée par rapport à un intervalle d'acceptabilité (figure 5). La caractéristique de performance majeure pour cette approche est l'exactitude. En effet, cette caractéristique est une combinaison de la justesse (erreur systématique) et de la fidélité (erreurs aléatoires). De plus, elle satisfait généralement à la demande du client qui veut un résultat

le plus exact possible. D'un simple point de vue métrologique, on peut signaler la concomitance entre le concept d'exactitude et le paramètre bien connu d'incertitude.

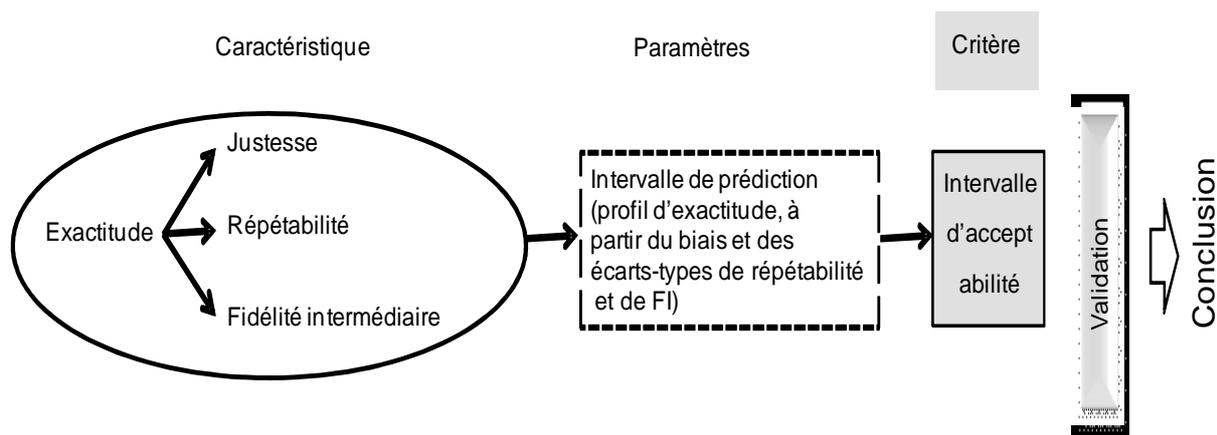


Figure 5 - Illustration de la méthodologie de caractérisation des caractéristiques de performance basée sur un critère unique .

(Approche globale, communication personnelle de Max Feinberg modifiée par le GT Val 2)

Pour les méthodes quantitatives, il existe un outil qui permet graphiquement de vérifier l'adéquation de l'exactitude des niveaux étudiés et l'intervalle d'acceptabilité : le profil d'exactitude. Sa construction repose sur la détermination classique des paramètres de performance mais qui sont combinés pour faciliter la prise de décision sur la validité de la méthode. Le profil d'exactitude est un outil de prédiction pour la réalisation des mesures en routine, mais également un outil de diagnostic des sources d'erreurs à contrôler en routine.

Il est important de souligner que l'approche globale permet de prendre en compte la totalité des processus de la méthode, comme cela se passera en routine.

En parallèle, une approche globale peut être appliquée pour les méthodes qualitatives. Elle est fondée sur la limite ou la probabilité de détection (LOD / POD).

6.2.3 Caractérisation absolue versus caractérisation relative

S'agissant plus spécifiquement du terme « relatif », cet adjectif est utilisé pour qualifier une caractéristique de performance d'une méthode dite alternative calculée par rapport à celle fournie par une méthode de référence et non par rapport à un statut connu des échantillons (approche déclinée pour les caractéristiques de performance concernés).

Note 1 : Lorsqu'une caractéristique de performance est calculée par rapport à celle d'une méthode de référence sur des échantillons de statuts incertains, les désaccords éventuellement observés entre les deux méthodes ne sont pas nécessairement liés à des inexactitudes de la méthode alternative ; le calcul de caractéristiques de performance relatives peut pénaliser des méthodes plus performantes que la méthode de référence en vigueur.

Note 2 : Les calculs de sensibilité / spécificité sont fondés sur des pourcentages de concordance entre les résultats attendus avec la méthode de référence et les résultats obtenus. Il est possible de vérifier que la distribution des valeurs observées n'est pas purement aléatoire à l'aide de tests statistiques.

6.2.4 Caractérisation intra-laboratoire versus interlaboratoires

La méthodologie de la caractérisation des critères de performance d'une méthode comporte deux étapes qui peuvent être effectuées l'une à la suite de l'autre, ou une seule en fonction des objectifs et, le cas échéant, du degré d'urgence de la validation.

La première étape est normalement de caractériser au niveau intra-laboratoire les paramètres de performance de la méthode à évaluer : la sensibilité (relative), la spécificité (relative), l'exactitude (relative), la limite de détection (relative), la fidélité (répétabilité, reproductibilité intra-laboratoire = fidélité intermédiaire).

La seconde étape repose sur l'organisation d'un essai inter-laboratoires pour vérifier les paramètres de performance obtenus en intra-laboratoire et tester également la reproductibilité, et selon les besoins d'autres caractéristiques telles que la robustesse de la méthode, certaines composantes de sa « déléguabilité » ou « transférabilité » (ex : facilité d'utilisation)...

D'autres paramètres et critères de performance (coût, durée, justesse...) peuvent également être calculés en fonction des besoins du client.

6.3 Validation

L'étape de validation consiste à confronter les valeurs des paramètres de performances obtenues pour les caractéristiques étudiées (cf. § 6.2), soit par critère, soit de manière globale aux critères de validation prédéterminés tels que définis lors de la revue de contrat (cf. § 6.1).

7. Caractérisation des performances des méthodes

Selon les méthodes analytiques étudiées, quantitatives ou qualitatives, il sera associé, à chaque caractéristique de performance, un ou des paramètres qui seront l'expression de la grandeur à estimer à partir de la série de mesures. Par exemple, la justesse aura pour paramètres le biais ou le taux de recouvrement ; la fidélité sera exprimée par un écart-type ou un coefficient de variation.

Note : Les valeurs des caractéristiques de performance peuvent varier selon le plan d'expérience appliqué et la méthode utilisée pour la calculer. Il convient donc, dans tout rapport de caractérisation (ou dans les publications rédigées), de bien préciser le plan d'expérience et la méthodologie de calcul.

7.1 Prérequis statistiques

Les analyses statistiques et en particulier l'utilisation de tests statistiques requièrent des hypothèses fondamentales sur les variables observées. Ce chapitre décrit les hypothèses de bases pour les variables utilisées dans les méthodes quantitatives et qualitatives. Il couvre également les notions nécessaires à la réalisation des plans d'expériences.

Note : les échantillons utilisés pour la caractérisation des performances des méthodes doivent être préparés de façon indépendante (voir figure 2).

7.1.1 Données qualitatives

7.1.1.1 Type de données

Dans le domaine des méthodes d'analyses qualitatives, deux types de variables sont rencontrés : les variables qualitatives nominales et les variables qualitatives ordinales.

Les variables qualitatives nominales sont des variables qualitatives dont la modalité ne permet pas de connaître la modalité suivante. Par exemple, les variables présence / absence, ou les variables +/-, sont des variables qualitatives nominales.

Les variables qualitatives ordinales ou les variables « semi-quantitatives », sont des variables qualitatives ordonnées. La modalité permet de connaître la modalité suivante. Par exemple, une classification 0, 1, 2, 3 ou bien une variable issue de dilution 1/256, 1/128, 1/64, 1/32, 1/16 sont des variables qualitatives ordonnées.



Ces variables n'ont pas de loi de distribution, et de façon classique, il est recommandé, quand cela est possible, de les transformer en variable quantitative discrète ou continue, par exemple en nombre de présents ou absents, en pourcentages etc. Après transformation, les hypothèses de normalité, indépendance et homogénéité des variances sont applicables (voir données quantitatives).

7.1.1.2 Tableau de contingence

Le calcul de certaines caractéristiques de performance des méthodes qualitatives fait appel aux notions d'accords / désaccords positifs et négatifs. Ces notions sont matérialisées par un tableau de contingence (tableau 1).

Tableau 1 - Synthèse des résultats obtenus sous forme de tableau de contingence.

		Statut de l'échantillon	
		Positif	Négatif
Résultat donné par la méthode à caractériser	Positif	PA = accords positifs = vrais positifs	PD = déviations positives = faux positifs
	Négatif	ND = déviations négatives = faux négatifs	NA = accords négatifs = vrais négatifs
Sommes		N+	N-

7.1.2 Méthodes quantitatives

Trois hypothèses doivent être respectées quand les variables sont quantitatives continues (concentration, quantité, valeur de Ct, logarithme du nombre d'unité formant colonie, réponse instrumentale...) : normalité, indépendance, homogénéité des variances.

Chacune des 3 hypothèses est présentée dans l'annexe « Bases statistiques pour les variables quantitatives ».

7.2 Les supports de caractérisation

7.2.1 Types de supports de caractérisation

Les supports de caractérisation comprennent les témoins positifs, les témoins négatifs, les matériaux de référence ou tout autre échantillon utilisé pour caractériser la méthode.

Les valeurs calculées pour les différentes caractéristiques d'une méthode et donc les conclusions de la validation sont directement liées au choix des supports utilisés pour réaliser l'étude de caractérisation. C'est pourquoi le choix de ces supports doit être judicieusement déterminé selon leur qualité et représentativité.

7.2.1.1 Qualité

Les matériaux de référence utilisés pour la caractérisation des méthodes doivent être de qualité adaptée, et leurs qualités intrinsèques connues. Ainsi, de manière générale, dans la mesure où ils existent et sont accessibles, le laboratoire utilisera par ordre de préférence des matériaux de référence certifiés externes, des matériaux de référence externes (par exemple issus d'études interlaboratoires) ou internes ou des échantillons de statut connu.



Rappel de définitions (issues du glossaire) :

- ✓ Matériau de référence (MR) : Matériau, suffisamment homogène et stable quant à une ou plusieurs propriétés spécifiées, qui a été préparé pour être adapté à son utilisation prévue dans un processus de mesurage.
- ✓ Matériau de référence certifié (MRC) : matériau de référence caractérisé par une procédure métrologiquement valide applicable à une ou plusieurs propriétés spécifiées et accompagné d'un certificat qui donne la valeur de la propriété spécifiée, son incertitude associée, et une expression de la déclaration de traçabilité métrologique.
Note : Les matériaux biologiques peuvent être issus des collections internationales ou nationales.
- ✓ Matériau de référence externe (MRE) : matériau de référence dont la (les) valeur(s) de consensus a (ont) été déterminée(s) à la suite d'études interlaboratoires.
- ✓ Matériau de référence interne (MRI) : matériau de référence dont la valeur de référence est attribuée par l'utilisateur : par comparaison aux valeurs certifiées ou aux valeurs de consensus des MRC ou MRE; par ajout d'une quantité connue de l'analyte à la matrice exempte de cet analyte.

Note : Par extension, sont inclus dans cette catégorie les échantillons de routine dont le statut a été précédemment déterminé par une ou plusieurs méthodes.

Remarques :

- (1) Des supports de statut inconnu peuvent également dans certains cas être utilisés lors d'études de caractérisation mais sont théoriquement réservés aux études réalisées par comparaison (ex : méthode alternative à une méthode de référence). Voir paragraphe 6.2.3. sur la comparaison relative.
- (2) Le laboratoire doit intégrer la portée et les limites des certificats des matériaux utilisés. À titre d'exemple, dans le domaine des OGM, les témoins positifs sont certifiés pour une cible et un certain pourcentage mais non certifiés pour l'absence d'autres cibles, ce qui peut entraîner des problèmes de spécificité lors des tests de caractérisation.
- (3) Il existe aussi des contrôles qualité internes constitués de la prise d'essai à analyser préparée spécifiquement et uniquement pour une série analytique. Ce type d'échantillon ne nécessite pas obligatoirement d'étude de stabilité et d'homogénéité. La valeur cible peut être la valeur nominale de la supplémentation.

7.2.1.2 Représentativité

Comme toute étude statistique, une étude de caractérisation ne peut pas nécessairement porter sur tous les objets / individus... du domaine d'application souhaité et étudié. Les supports de comparaison, en tant qu'échantillons au sens statistique du terme, devront donc être choisis (nature, caractéristique, concentration) afin d'être les plus représentatifs possibles des matrices à analyser de sorte que les conclusions dressées à l'issue de la caractérisation soient valides pour l'ensemble du domaine de validation (voir ci-dessous supports de comparaison « positifs » et « négatifs »).

Dans le cas des études de caractérisation, en revanche, la représentativité des échantillons n'est souvent pas basée sur un tirage aléatoire au sein de la population d'étude mais selon des critères de choix (stratification).

ex :

- prendre des eaux plates et gazeuses pour étudier les eaux embouteillées ;
- prendre des espèces représentatives au sein d'un genre (bactérien, animal, végétal,...)
- ...



7.2.2 Les supports de caractérisation « positifs »

Les supports de comparaison « positifs » sont ceux choisis pour leur propriété connue (ou supposée) de contenir l'analyte cible. Deux cas sont à considérer :

- o les matrices naturellement contaminées ;
- o les matrices (blanches ou non : voir ci-dessous) supplémentées d'une quantité connue (si possible) de l'analyte.

Lors d'études de caractérisation, les matrices naturellement contaminées seront généralement privilégiées mais les secondes seront envisagées dès lors que les premières ne sont pas disponibles.

Par ailleurs, plus les sources de variabilité (analyte, matrice...) sont importantes plus il faut augmenter le nombre d'échantillons. Pour une variabilité liée à l'analyte, 10 échantillons positifs au minimum sont en général requis. Toutefois dans le cas d'analytes dont les échantillons sont rares, ce nombre pourra être réduit sans toutefois pouvoir être nul. Pour une variabilité liée aux matrices, il est recommandé de tester des matrices d'origines biologiques différentes.

Lorsque plusieurs échantillons de référence sont disponibles, le laboratoire les choisira de manière à avoir la plus grande variabilité intra spécifique possible : origines géographiques variées, plantes hôtes différentes, dates de prélèvements éloignées etc.

7.2.3 Les supports de caractérisation « négatifs »

7.2.3.1 Les négatifs ou blancs

La notion de blanc, d'échantillon témoin et de témoins négatifs est définie dans de nombreuses lignes directrices, de référentiels européens ou nationaux et dans des réglementations. Dans la norme NF ISO T90-210 (2009), la définition du blanc est la suivante : « essai réalisé en l'absence de matrice (blanc réactif) ou sur une matrice qui ne contient pas l'analyte (blanc matrice). Il est indispensable que le laboratoire précise de quel blanc il s'agit ». Cette définition précise l'existence de plusieurs types de blanc : le blanc réactif et le blanc matrice.

À ces deux types de blanc, il convient d'en ajouter un troisième : le « blanc instrumental » résultant du bruit de fond de l'appareillage utilisé.

L'estimation du blanc est primordiale car il est souvent utilisé pour déterminer la limite de détection selon le domaine analytique. Dans ce cas, le blanc utilisé est le blanc matrice et à défaut le blanc réactif.

Note 1 : Dans certains cas, il n'est pas possible d'avoir de « vrai » blanc, c'est-à-dire un blanc ayant un bruit de fond négligeable lié à la présence du contaminant étudié (chimique, microbiologique, etc.).

Note 2 : Dans certains types de méthodes, l'appareil utilisé définit un bruit de fond (exemple : PCR temps réel et baseline, absorbance pour le « témoin tampon »...) ; il appartient à l'utilisateur de préciser si la valeur de ce dernier doit être retirée ou non des mesures pour les échantillons.

7.2.3.2 Les analytes non cibles

Les analytes non cibles sont également des supports de caractérisation « négatifs » au sens où ils ne contiennent pas l'analyte cible et donc que la réponse attendue avec la méthode caractérisée est négative. Ils se distinguent des « blancs » précédemment définis par le fait qu'ils contiennent un ou plusieurs analytes distincts de celui ou ceux étudiés mais dont l'interférence éventuelle avec le test doit être évaluée (cf. spécificité).



Critères de choix des analytes non cibles :

Les échantillons seront choisis de façon à ce que les analytes soient le plus proche possible de l'analyte cible (même sous-genre, genre ou famille). Le plus grand nombre possible d'analytes proches sera testé. Les analytes les plus éloignés mais susceptibles d'être rencontrés fréquemment dans les conditions de prélèvement (par exemple mêmes espèces hôtes, même marchandise, même région géographique...) seront également sélectionnés.

Pour chaque analyte proche, plusieurs échantillons seront testés si possible, en retenant les mêmes critères de choix que pour les échantillons positifs.

7.3 Caractéristiques de performance à déterminer selon le type de méthode

Les caractéristiques de performance des méthodes qualitatives et quantitatives sont résumées dans le tableau 2. Ces caractéristiques ne doivent pas être systématiquement étudiées et documentées mais l'absence d'une ou plusieurs caractéristiques doit être justifiée dans le rapport de caractérisation et de validation de méthode d'analyse.

Tableau 2 - Caractéristiques de performance à déterminer pour les méthodes qualitatives et les méthodes quantitatives.

Étape du processus de validation	Caractéristique	Caractéristique de performance à déterminer en fonction du type de méthode		
		Qualitative	Quantitative	
Caractérisation Intra labo	Spécificité	X	X	
	Sensibilité	X	(X)	
	Fonction d'étalonnage/efficacité (PCR)		X	
	Fidélité	Répétabilité	(X)	X
		Fidélité intermédiaire	X	X
	Justesse	sans objet	X	
	Exactitude (justesse + fidélité) ^a	sans objet	X	
	Linéarité	sans objet	X	
	Limite de	quantification	sans objet	X
détection		X	(X)	
Domaine de validité	X	X		
Caractérisation interlaboratoires	Reproductibilité	X	X	
	Répétabilité	(X)	X	
	LOD	(X)	sans objet	
	LOQ	sans objet	(X)	
	Spécificité ^b	X	sans objet	
	Sensibilité ^c	X	sans objet	
	Autres caractéristiques non techniques à définir selon les points critiques et le cahier des charges (délai, rapidité, efficacité, coût)	X	X	

sans objet : caractéristique non pertinente (x) : les caractéristiques inscrites entre parenthèses sont recommandées.

a : pour les méthodes analytiques quantitatives, l'exactitude représente toujours la combinaison de la justesse et de la fidélité.

b : ou taux de faux positifs dans certains référentiels

c : ou taux de faux négatifs dans certains référentiels



Note : les caractéristiques listées dans le tableau ci-dessus sont au minimum requises pour une caractérisation complète de la méthode. Toutefois, selon les référentiels applicables au domaine concerné, d'autres caractéristiques pourront être exigées (ex : robustesse, exactitude relative,...).

Note : si la méthode doit être appliquée à un ensemble de matrices et que les critères de validation n'ont pas été atteints pour la totalité des matrices alors le laboratoire peut conclure en proposant une aptitude de la méthode dans un domaine de validité plus restreint que celui initialement envisagé (domaine de validation).

Le domaine d'application « possible » est défini à partir du domaine de validité et par la justification de la représentativité des supports de caractérisation par rapport à l'ensemble des matrices sur laquelle la méthode sera appliquée. Ce domaine d'application possible issu de l'étude de caractérisation / validation est ensuite utilisable comme étant celui de la méthode rédigée et proposée (voir modèle de méthode – ANSES/PR3/7/01-04). Les laboratoires utilisateurs de cette méthode devront la mettre en œuvre dans le domaine d'application établi.

Les caractéristiques de performance non techniques font également partie de la décision finale pour la validation de la méthode. En effet la facilité de la mise en œuvre, la faisabilité et l'efficacité sont des caractéristiques qui sont à prendre en compte. Cependant, elles ne feront pas l'objet d'un chapitre particulier dans ce guide.

7.4 Modalités pratiques de caractérisation

7.4.1 Spécificité

La spécificité d'une procédure analytique quantitative est la capacité à établir de manière univoque l'existence de la substance à analyser en présence d'autres composants potentiellement présents. Une procédure d'analyse est dite « spécifique » lorsqu'elle permet de garantir que le signal mesuré provient seulement de la substance à analyser ou qu'elle permet de mesurer quantitativement un paramètre physico-chimique ou un groupement fonctionnel d'une ou de plusieurs substance(s) dans l'échantillon.

Selon cette définition, il s'agit bien de démontrer que la ou les substance(s) quantifiée(s) au sein de la matrice sont bien le ou les analyte(s) recherché(s). Très souvent, la spécificité se fonde sur une absence d'interférences. Elle peut être démontrée par différents moyens. C'est ainsi qu'elle peut être inhérente à la technique (par exemple : identification par spectrométrie infrarouge ou spectrométrie de masse), elle peut être obtenue par séparation préalable (par exemple : chromatographie), de façon mathématique (par exemple : résolution d'équations simultanées) ou de façon biochimique (réactions à l'aide d'enzymes).

Dans le domaine des méthodes qualitatives, deux types de spécificité sont classiquement définis : la spécificité analytique et la spécificité diagnostique. La différence entre ces deux spécificités repose sur l'origine du support utilisé. Pour la spécificité analytique, les supports sont des supports de référence et pour la spécificité diagnostique, les supports sont issus du terrain. Dans le domaine de la microbiologie des aliments, le terme « exclusivité » est également utilisé quand le calcul est réalisé sur des souches pures, et le terme « spécificité » est utilisé lorsqu'il est réalisé sur des supports de type aliments.

La spécificité correspond au pourcentage de résultats négatifs trouvés parmi les négatifs attendus. Les calculs sont fournis dans l'annexe « Spécificité ».

Dans tous les cas, la spécificité vise à évaluer la capacité de la méthode soumise à validation à ne pas donner de réponse quand la cible n'est pas présente.

7.4.2 Sensibilité

Comme pour la spécificité, deux types de sensibilité sont généralement définis : la sensibilité analytique et la sensibilité diagnostique. La différence entre ces deux sensibilités repose là-encore



sur l'origine du support utilisé. Pour la sensibilité analytique, les supports sont des supports de référence et pour la sensibilité diagnostique, les supports sont issus du terrain. Dans le domaine de la microbiologie des aliments, le terme « inclusivité » est utilisé quand le calcul est réalisé sur des souches pures, et le terme « sensibilité » est utilisé lorsqu'il est réalisé sur des supports de type aliments.

La sensibilité correspond au pourcentage de résultats positifs trouvés parmi les résultats positifs attendus. Les calculs sont fournis dans l'annexe « Sensibilité ».

Dans tous les cas, la sensibilité vise à évaluer la capacité de la méthode soumise à validation à donner une réponse positive quand la cible est présente.

7.4.3 Fonction d'étalonnage/Efficacité (PCR)

Pour un grand nombre de méthodes quantitatives, la quantification se fait à l'aide d'une fonction étalonnage. C'est un élément important de la mesure car la fonction d'étalonnage doit être établie de façon exacte et doit permettre de sélectionner un modèle d'étalonnage qui sera utilisé en routine. Le modèle doit décrire au mieux la relation entre la variable dépendante Y, généralement une réponse instrumentale, et la variable X qui représente la quantité d'analyte et qui est une variable indépendante génératrice de la réponse Y. Des exemples de fonctions d'étalonnage usuelles sont présentés dans l'annexe « Fonction d'étalonnage, efficacité (PCR) ». Il est également précisé dans cette annexe le nombre de niveaux et de répétitions nécessaires pour obtenir une bonne estimation des paramètres du modèle.

7.4.4 Linéarité

Le terme « linéarité » concerne la relation entre les concentrations calculées à l'aide de la fonction d'étalonnage et les concentrations théoriques. Il peut exister une confusion car la linéarité peut aussi être évaluée pour la fonction d'étalonnage lorsqu'il s'agit d'un modèle linéaire. Comme cela est décrit dans l'annexe « Fonction d'étalonnage, efficacité (PCR) », la fonction d'étalonnage n'est pas forcément linéaire, ce qui n'empêche pas d'avoir une bonne quantification. Dans le contexte de la validation, l'étude de la linéarité porte sur la proportionnalité entre les concentrations calculées par la fonction d'étalonnage et les concentrations théoriques. L'étude de la linéarité est en fait une évaluation du biais de justesse et sera décrite dans l'annexe concernant la justesse.

7.4.5 Justesse

La justesse exprime la différence entre un ensemble de valeurs mesurées à l'infini et une valeur de référence. La vraie valeur de la justesse est déterminée par rapport à un matériau de référence certifié (MRC). Dans le cas d'absence de MRC, une valeur qui est acceptée comme une valeur conventionnellement vraie peut être utilisée.

La justesse est étudiée pour chaque niveau de concentration avec ou sans la matrice. Ces niveaux sont communément appelés « standards de validation ». Il existe plusieurs modes d'expression de la justesse car le paramètre de validation qui la caractérise est le biais, qui peut être exprimé de façon absolue ou relative.

L'annexe « Justesse » décrit ces différents modes d'expression et présente un exemple de correction dans le cas de la présence d'un biais de justesse (effet matrice par exemple).

7.4.6 Fidélité

La fidélité exprime l'étroitesse de l'accord entre une série de mesures provenant de multiples prises d'essai d'un même échantillon homogène dans des conditions prescrites. La fidélité peut être évaluée à trois niveaux (figure 6) : la répétabilité, la fidélité intermédiaire (intra-laboratoire) et la reproductibilité (interlaboratoires).



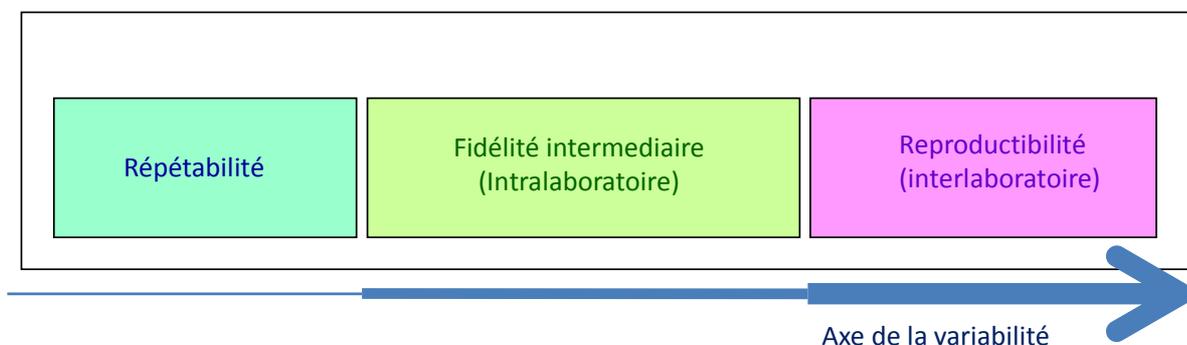


Figure 6 - Composantes croissantes de la variabilité : répétabilité, fidélité intermédiaire et reproductibilité.

La fidélité traduit uniquement la distribution des erreurs aléatoires et n'a aucune relation avec la valeur vraie ou spécifiée. La mesure de fidélité est calculée à partir de l'écart-type des résultats d'essais et peut être exprimée sous forme de coefficient de variation (CV).

Il convient si possible d'étudier simultanément la répétabilité et la fidélité intermédiaire, même si ces caractéristiques sont décrites séparément dans les normes, réglementations ou référentiels (voir annexe « Fidélité »).

7.4.7 Répétabilité

La répétabilité est définie par les conditions de répétabilité qui regroupent les facteurs intra-séries (mêmes conditions opératoires, dont même technicien, et court intervalle de temps entre les répétitions). Il est nécessaire d'avoir des répétitions par niveau pour identifier et quantifier les sources d'erreur. Le nombre de répétitions minimal par niveau est de 2, mais il peut être porté à 3 si le nombre de niveaux étudiés est faible ($n \leq 3$).

Le nombre de séries et le nombre de répétitions par niveau sont liés et dépendent de la variabilité de la méthode. De plus, si la variabilité issue de la répétabilité est faible par rapport à celle générée par la fidélité intermédiaire, le plan d'expérience devra, par exemple, inclure plus de séries et simplement un nombre minimal de 2 répétitions par niveau. Selon les domaines, le nombre de répétitions et le nombre de séries peuvent être définis dans des lignes directrices, normes ou réglementations.

Une étude pilote ou de pré-validation est conseillée pour pouvoir optimiser le plan expérimental en fonction des sources de variations.

7.4.8 Fidélité intermédiaire

La fidélité intermédiaire (appelée reproductibilité intra-laboratoire en microbiologie des aliments, cf. ISO/TS 19035 sur l'incertitude de mesure) est estimée par la répétition des séries, afin de pouvoir quantifier la contribution de cette source d'erreur sur la variabilité totale (maximiser les sources de variations dans un même laboratoire, dont différents techniciens, pour effectuer les répétitions). Le nombre minimal de séries est de 3, quel que soit le nombre de niveaux étudiés.

Le nombre de séries et le nombre de répétitions par niveau sont liés et dépendent de la variabilité de la méthode. Selon les domaines, le nombre de répétitions et le nombre de séries peuvent être définis dans des lignes directrices, normes ou réglementations.

Une étude pilote ou de pré-validation est conseillée pour pouvoir optimiser le plan expérimental en fonction des sources de variations.

7.4.9 Reproductibilité

L'étude de la reproductibilité permet de caractériser la variabilité totale incluant la variabilité liée aux laboratoires utilisant la méthode. Cette caractérisation se fait au moyen d'un essai inter-laboratoires de validation. Il ne faut pas confondre ce type d'essai inter-laboratoires avec un essai inter-laboratoires d'aptitude qui permet d'évaluer l'aptitude des laboratoires. Les six fascicules de la norme NF ISO 5725 décrivent les processus et les calculs statistiques à mener. Ces calculs sont également repris dans le chapitre 6 de « Labo-Stat : Guide de validation des méthodes d'analyse » (M. Feinberg, 2009, ed. Lavoisier). Un résumé est fourni dans l'annexe « Fidélité ».

7.4.10 Exactitude

L'exactitude exprime l'étroitesse de l'accord entre le résultat et la valeur de référence acceptée, aussi appelée « valeur vraie ». L'étroitesse de l'accord ainsi observée est la résultante de la somme des erreurs systématiques et aléatoires, en d'autres termes l'erreur totale liée au résultat. Par conséquent, l'exactitude est l'expression de la combinaison de la justesse et de la fidélité. De ce fait, c'est la caractéristique de performance majeure qui est étudiée dans une approche globale.

Cette étude peut se faire par l'utilisation d'un outil graphique qui repose sur une approche statistique fondée sur le calcul d'un intervalle de dispersion. Cet outil est le profil d'exactitude (cf. paragraphe 6.2.2) et il permet de statuer sur la validation de la performance de la méthode. Il a été proposé en 2003 par une commission de la SFSTP et il est utilisé dans de nombreux domaines tels que la physico-chimie, mais également le dénombrement bactérien par méthode classique, l'ELISA et la PCR quantitative. Il est reconnu aux niveaux des instances de normalisation ISO et AFNOR et a été introduit dans des normes AFNOR telles que les normes NF V03-110, NF U 47-600-2, ainsi que dans la nouvelle version de la NF EN FDIS 16140-2.

Cette approche est décrite dans l'annexe « Exactitude ». Il y est également présenté un exemple d'application.

7.4.11 Limites

Les limites de détection (LOD) et/ou de quantification (LOQ) sont des valeurs systématiquement rapportées dans les dossiers de validation. Les définitions sont nombreuses et il en a été recensé plus d'une cinquantaine (Currie, 2004).

7.4.11.1 Limite de détection

De nombreuses réglementations, lignes directrices ou normes proposent des méthodes d'estimation de la LOD. Toutefois, toutes visent à déterminer la limite en-dessous de laquelle l'analyte est considéré comme « non détecté ». Pour ce faire, toutes se fondent sur l'analyse d'échantillons contenant l'analyte à différents niveaux de concentration, mais des différences portent sur :

- le nombre de niveaux (1, 3, 5...) et de répétitions par niveaux (3, 6, 10, 60...);
- le critère de décision, notamment la probabilité de détection associée ou son intervalle de confiance : $LOD_{50\%}$, $LOD_{95\%}$,
- les modalités d'exploitation des résultats issus de ces plans d'expérience avec des interprétations :
 - o fondées sur le ratio entre le nombre de répétitions donnant une réponse positive et le nombre total de répétitions, par niveau : ex : 10+/20 ;
 - o graphiques fondées sur une modélisation de la probabilité de détection en fonction de la concentration de l'analyte (ex : POD).



Dans tous les cas, il convient qu'un laboratoire amené à déterminer une LOD pour une méthode, précise les modalités selon lesquelles la valeur a été obtenue et la probabilité de détection associée.

Note 1 : Pour les méthodes qualitatives, en l'absence de réglementation, lignes directrices ou de norme applicable au domaine analytique considéré, le GT propose de calculer la LOD selon la norme NF EN ISO16140-2, utilisant l'approche fondée sur la probabilité de détection (POD, modèle linéaire généralisé).

Note 2 : Pour les méthodes quantitatives, en l'absence de réglementation, lignes directrices ou de norme applicable au domaine analytique considéré, le GT propose d'utiliser préférentiellement soit une méthode de calcul de la LOD fondée sur le rapport signal sur bruit ou bien une estimation à partir de la fonction d'étalonnage.

7.4.11.2 Limite de quantification

Il n'existe pas de définition précise de la limite de quantification (LOQ). Généralement, la LOQ représente la plus faible concentration dans un échantillon qui puisse être quantifiée avec une fidélité et une justesse acceptables dans des conditions expérimentales indiquées.

Comme pour la LOD, les critères de justesse et de fidélité retenus pour fixer cette valeur de LOQ doivent être clairement définis (ex : CV de répétabilité <25% et biais / erreur de justesse <30%...). La valeur peut être déduite à partir de l'approche globale fondée sur le profil d'exactitude, ou par l'estimation sur une plage d'étude de ces deux caractéristiques séparément.

Remarque : dans certains domaines, la LOQ est calculée à partir du rapport signal sur bruit et exprimée comme un multiple de l'écart-type signal/bruit.

En l'absence de réglementation, de lignes directrices ou de norme dans le domaine analytique considéré, le GT propose de calculer la LOQ préférentiellement par une approche globale (profil d'exactitude).

Les calculs concernant la LOQ sont présentés dans l'annexe «Limites».

7.4.12 Autres types de limites

En fonction des référentiels et des réglementations d'autres limites peuvent être exigées. Par exemple la décision CE/2002/657 introduit deux limites à documenter dans le cadre des résidus de médicaments vétérinaires : la limite de décision $CC\alpha$ et la capacité de détection $CC\beta$:

- 1) La limite de décision ($CC\alpha$) est la limite à partir de laquelle il peut être décidé si un échantillon est réellement non conforme avec un risque d'erreur α .
- 2) La capacité de détection ($CC\beta$) est la plus petite concentration qui puisse être détectée, identifiée et/ou quantifiée avec un risque d'erreur β .

Les calculs de ces limites sont décrits dans la décision CE/2002/657 mais sont initialement issus de la norme NF EN ISO 11843.

7.5 Outils de calculs pour la validation

De nombreux outils sont disponibles pour effectuer les calculs statistiques. L'utilisation de logiciels statistiques commerciaux (SAS, Statgraphics, Systat...) ou en accès libre (ex : R) est possible pour faire les analyses et les graphiques basiques (Boxplot, comparaison de moyenne (test de Student, Anova...), comparaison de variances (test de Fisher, Bartlett, Levène, Cochran, Grubbs, Dixon ...), les calculs d'intervalle de confiance... Il existe également des logiciels intégrés dédiés à la validation.



Il est également possible d'utiliser des feuilles de calcul du logiciel Microsoft Excel. Des feuilles ont été développées pour le calcul du profil d'exactitude (feuille associée à la norme NF V 03-110) ou pour le calcul de LOD pour les méthodes qualitatives

(<http://www.wiwiss.fu-berlin.de/fachbereich/vwl/iso/ehemalige/wilrich/index.html>).

Elles sont disponibles sur le site intranet de la plateforme Anses/PAS.

Si des feuilles de calcul sont développées en interne, le GT rappelle que ces feuilles doivent être validées sur le plan statistique (calculs) mais également sur le plan informatique (environnement...) et sur le plan qualité (traçabilité).

8 Adéquation de la performance et validation

8.1 La démarche

L'étape finale est la comparaison des valeurs de performance obtenues par la méthode aux critères de performance prédéterminés. Pour conclure sur les différentes caractéristiques de performance, il est proposé de faire une synthèse tabulaire (tableau 3) et de fournir une conclusion sur la validité du critère.

Tableau 3 - Synthèse tabulaire des caractéristiques de performances, de leurs valeurs prédéterminées, de leurs valeurs obtenues et de la décision sur leur validité.

Caractéristique de performance (Paramètre)	Valeur cible prédéterminée (cahier des charges)	Valeur obtenue à l'issue de la caractérisation	Conclusion (OK, domaine d'application à revoir...)

Afin de globaliser la démarche, un modèle commun de présentation de méthode et un rapport de caractérisation et de validation / vérification sont proposés.

8.2 Modèle de méthode d'analyse

Un modèle de méthode d'analyse a été développé simultanément à la rédaction du présent guide. Ce modèle est accompagné d'un guide qui fournit des indications utiles pour son utilisation. Fondé sur les normes en la matière (norme ISO 78-2 notamment), ce modèle permet d'identifier la méthode, d'en fournir un descriptif (domaine d'application, réactifs, appareillage, type d'échantillons etc.), de détailler son mode opératoire, de préciser les résultats sur ses performances ainsi que des préconisations pour sa mise en œuvre.

De plus, ce modèle prévoit d'assurer la traçabilité de l'historique de la méthode, en répertoriant les versions successives et en distinguant le niveau des modifications effectuées sur les différentes versions, en mineures ou majeures.

8.3 Rapport de validation

Une trame de rapport de validation de méthode a également été élaborée en lien avec le présent guide. Cette trame comporte sept paragraphes et la possibilité d'intégrer des annexes. Les paragraphes permettent de :

- décrire la méthode, le domaine d'application et les critères de performance cibles ;



- décrire les résultats pour les différentes caractéristiques étudiées, pour des caractérisations intra ou inter-laboratoires ;
- fournir une incertitude de mesure, une conclusion sur la validation de la méthode et une bibliographie associée.

Les annexes permettent de fournir les résultats détaillés des caractéristiques et des paramètres associés qui ont été étudiés.

9 Incertitude

La norme NF EN ISO/CEI 17025 prévoit que les laboratoires d'étalonnage ou d'essais doivent disposer de procédures permettant d'estimer l'incertitude. Si le guide EURACHEM propose une procédure de détermination de l'incertitude, elle reste assez complexe et il convient de l'adapter à chaque domaine de la chimie et, en particulier, en utilisant les données issues de la mesure de la performance des méthodes. L'estimation de l'incertitude va au-delà du calcul d'un intervalle (de confiance ou de répartition) car toutes les sources d'erreurs doivent être prises en compte. La figure 7 résume l'ensemble des étapes généralement proposées pour estimer l'incertitude.

Dans le domaine de la microbiologie des aliments, il existe une approche normative (XP CEN/ISO TS 19036) pour calculer l'incertitude de mesure pour les analyses quantitatives.

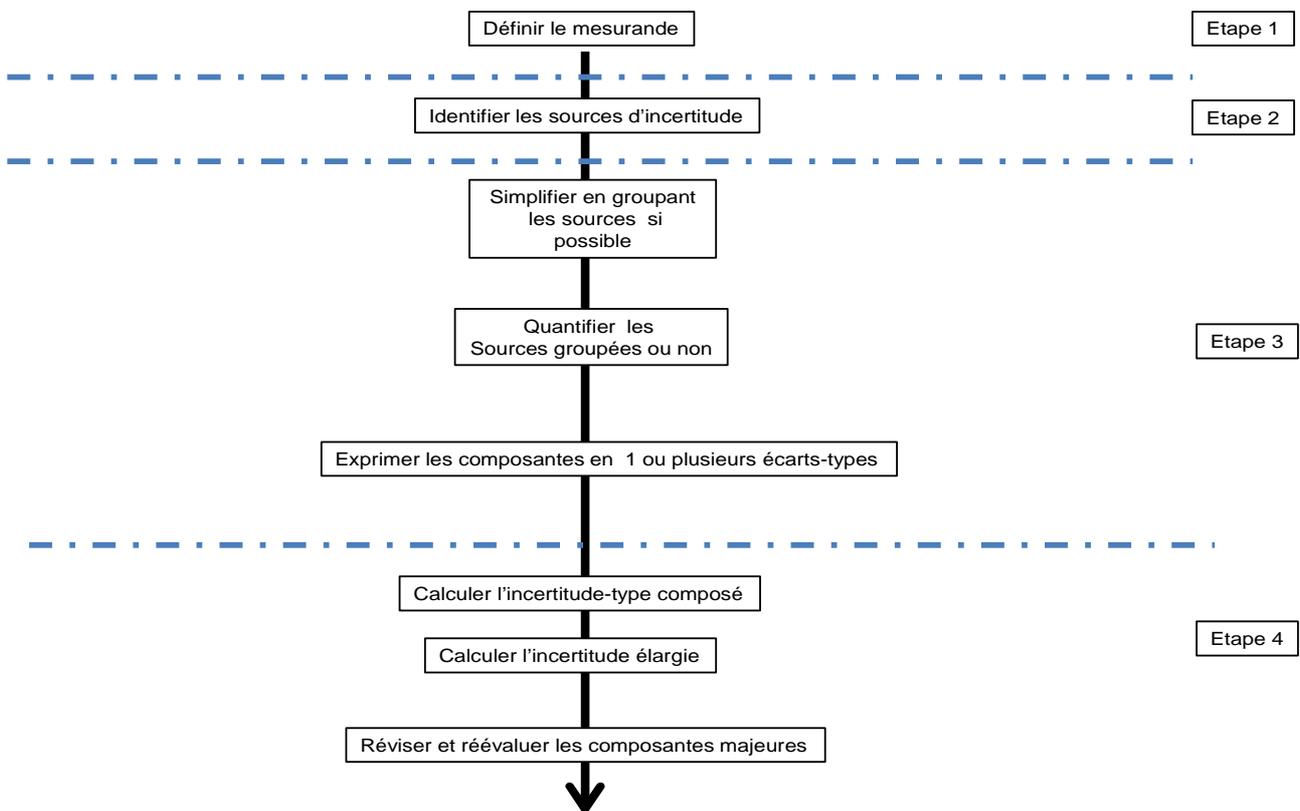


Figure 7 - Processus d'évaluation de l'incertitude.

(D'après Max Feinberg, 2009, et modifié par le groupe GT Val 2)

La première étape est de définir le mesurande, en portant une attention particulière sur le processus de calcul qui va permettre de passer à la deuxième étape qui est l'identification des sources d'incertitudes.

En utilisant le principe du diagramme général des causes et effets de type Ishikawa, les principales sources peuvent être regroupées en 5 modalités ou règle des 5 M : Main d'œuvre, Moyen, Milieu, Matière, Méthode. Le diagramme est présenté sur la figure 8.



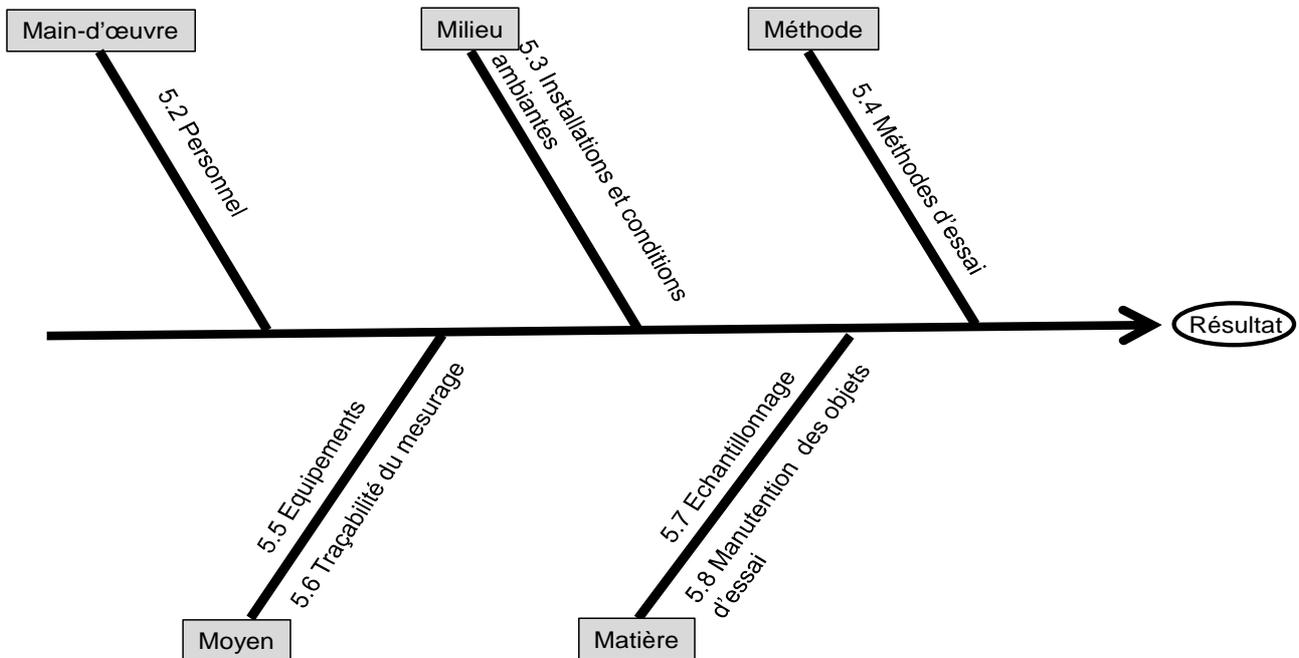


Figure 8 - Principales sources d'incertitudes et chapitres correspondants dans la norme NF EN ISO/CEI 17025.

(D'après Max Feinberg, 2009, modifié par le groupe GT Val 2)

En ce qui concerne la main d'œuvre, il faut considérer tout ce qui est lié aux opérateurs, ou tout personnel intervenant dans ou sur le processus de réalisation de la méthode d'analyse. La source d'erreur concernant les moyens intègre les équipements de mesure, la pureté des réactifs, etc. Pour ce qui concerne le milieu, il faut prendre en compte les installations mais également une partie aléatoire. Les principales sources d'erreur dans la matière sont à la fois dues à des erreurs dans l'échantillonnage, mais également dans les conditions de conservation ou toute autre source liée à l'échantillon (variation biologique, etc.). Enfin, pour ce qui concerne la méthode, les sources d'erreur doivent inclure l'ensemble des facteurs inhérents à la méthode : préparation de l'échantillon, matériel de laboratoire (ex : verrerie), stœchiométrie des réactions, processus de calcul, correction du blanc, etc.

L'étape 3 consiste à quantifier l'incertitude à partir des sources identifiées. Le principe de base est de quantifier l'incertitude par un ou des écarts-types qui sont appelés incertitudes-types et notés u . Les sources d'erreur étant généralement multiples, l'incertitude globale appelée incertitude type composée correspond à la combinaison de l'ensemble des sources d'erreur. Par convention, elle est notée u_C .

Deux modes d'estimation de l'incertitude type sont décrits dans le guide pour l'expression de l'incertitude de mesure (GUM) :

- l'approche dite de type A, approche statistique basée sur une série d'observations,
- l'approche de type B, basée sur des approches probabilistes.

Approche de type A :

Les données qui peuvent être utilisées ont des origines diverses : elles peuvent être issues d'essais inter-laboratoires, d'études de robustesse, d'études antérieures, d'essais d'aptitude ou bien d'études de performance de méthode. Cette dernière origine est conseillée car elle permet de documenter les sources d'incertitudes dans des conditions telles que les conditions de répétabilité et / ou de fidélité intermédiaire. L'utilisation de l'approche globale et en particulier du profil

d'exactitude permet d'estimer l'incertitude type et d'estimer une fonction d'incertitude (Gassner, 2014). Le principe et les calculs sont détaillés dans l'annexe « Incertitude ».

Approche de type B :

Il faut disposer, pour cette approche, d'une relation mathématique entre toutes les grandeurs qui interviennent dans le mesurage (VIM). Par exemple, la relation établie entre la concentration et une réponse instrumentale. Cependant, cette approche est peu utilisée dans le domaine des méthodes quantitatives car cette relation n'explique qu'une partie de l'incertitude et il est préféré une approche de type A pour estimer l'incertitude type composée.

Dans le cas d'une méthode qualitative, l'incertitude ne peut pas être chiffrée mais elle doit être documentée et les sources d'erreur doivent être identifiées.

Il est également possible de combiner, pour calculer l'incertitude type, l'approche de type A et l'approche de type B.

Le calcul de l'incertitude type composée et de l'incertitude élargie constituent la quatrième étape. Il existe globalement deux méthodes : une méthode générale et une méthode par discrétisation.

La méthode générale repose sur l'équation de propagation des variances. Cela traduit la combinaison des différentes incertitudes-types estimées. Les cas les plus simples, c'est-à-dire les cas additifs et les cas multiplicatifs sont décrits dans l'annexe « Incertitude ».

Note : La méthode par discrétisation (Kragten, 1994) n'est pas développée dans ce guide.

Le groupe GT Val 2 préconise l'approche globale pour calculer l'incertitude type composée.

Dans le domaine de la microbiologie des aliments, pour les méthodes quantitatives, l'approche normative 'XP CEN/ISO TS 19036 est semi-globale : elle combine une approche de type A (écart-type de reproductibilité du résultat final de l'analyse) et une approche de type B (composante liée à la distribution des bactéries en suspension, selon la loi de Poisson, devenant prépondérante aux bas niveaux de contamination). Trois options sont décrites pour l'estimation de l'écart-type de reproductibilité s_R , classées par ordre décroissant de priorité :

- s_R intra-laboratoire, avec plan d'expérience permettant d'inclure la composante de l'incertitude liée à la prise d'essai dans l'échantillon pour essai, importante en cas de contamination hétérogène de l'échantillon (matrices solides) ;
- s_R inter-laboratoires dérivée d'un EILV, sous certaines conditions décrites dans la norme ;
- s_R inter-laboratoires dérivée d'un EILA, sous certaines conditions décrites dans la norme.

Il est classique de fournir, associé aux résultats rendus, l'incertitude élargie. Cette incertitude élargie, notée U, est dérivée de l'incertitude type composée par l'expression :

$$U = k \times u_c()$$

Où $u_c()$ est l'incertitude type composée et k le facteur d'élargissement.

Par convention le facteur k est fixé à 2, indiquant que 95 % des valeurs possibles sont dans U, et il est fixé à 3 s'il est souhaité que 99 % des valeurs possibles soient incluses dans U.

La connaissance de l'incertitude élargie est recommandée pour pouvoir interpréter un résultat et prendre une décision en connaissance des causes d'incertitude. Un exemple est présenté dans l'annexe « Incertitude ».

Une autre approche de l'estimation de l'incertitude de mesure est l'utilisation de données issues d'un EILV comme décrit dans la norme NF ISO 21748.

L'incertitude élargie peut être fixée dans des normes ou référentiels réglementaires mais elle doit être vérifiée.



10 Perspectives

Ce guide a pour vocation d'aider et d'harmoniser les approches nécessaires à la validation des méthodes quantitatives et qualitatives développées dans les laboratoires de l'Anses.

Cependant certains points n'ont pas été abordés comme :

- le transfert des méthodes analytiques
- la validation de méthodes d'analyses multi-analytes
- l'utilisation de carte de contrôle en routine (contrôle qualité)
- de nouvelles approches développées pour estimer l'incertitude.

Ces points pourront faire l'objet de compléments ultérieurs.

Ce guide sera révisé périodiquement pour intégrer les évolutions techniques des domaines couverts, les évolutions des normes et pour intégrer les nouvelles approches développées en statistiques.

11 Références bibliographiques

Code rural et de la pêche maritime, article R202-5, (Juillet 2012)

Currie L.A. ., 2004, Detection and quantification limits: basic concepts, international harmonization and outstanding ("low-level") issues, 2004, Applied Radiation and Isotopes, 61, 145-149

Eurachem/Citac Guide, 2000, Quantifying Uncertainty in Analytical Measurement, 2000, Second edition, final draft. Traduit par le LNET et accessible sur le site www.lne.fr

Feinberg M., 2009, Labo-Stat : Guide de validation des méthodes d'analyse, Lavoisier, Tec&Doc

Gassner A.L., Schappler J., Feinberg M., Rudaz S., 2014, Derivation of uncertainty functions from validation studies in biological fluids: application to the analysis of caffeine and its major metabolites in human plasma samples, 2014, J. Chrom A, 1353, 121-130

Héreau, V. Anthoine G. Processus français d'officialisation des méthodes d'analyses dans le domaine de la santé des végétaux. Euroréférence n°9 Spécial « végétal ». Printemps 2013. 16-19

Holst-Jensen A and Berdal K.G., 2004, The modular analytical procedure and validation approach and the units of measurement for genetically modified materials in foods and feeds, 2004, Journal of AOAC International, 87, 4, 927-936

Hougs L, Zel J, Burns M, Ciabatti I, Mankertz J, Narendja F, Sandberg M, Schulze M, Scholtens I, Charels D, Charles-Delobel C, Luque-Perez E, Mazzara M, Savini C, Weber T, Plan D et Van den Eede G, (2011)., Guidance document from the European Network of GMO laboratories (ENGL): Verification of analytical methods for GMO testing when implementing interlaboratory validated methods. Luxembourg: Publications Office of the European Union.

Kragten J., 1994, Calculating standard deviation and confidence intervals with the universally applicable spreadsheet technique, 1994, The Analyst, 119, 2161-2166.

Schwartz, 1993, Méthodes statistiques à l'usage des médecins et des biologistes, 4ème Ed. Paris, Flammarion, Médecine-Sciences, 1993.

Thompson M., 2000, Recent trends in inter-laboratory precision at ppb and sub-ppb concentrations in relation to fitness for purpose criteria in proficiency testing. 2000, The Analyst; 125, 385-386.

Wilrich, C., Wilrich, P.T., 2009, Estimation of the POD function and the LOD of a qualitative microbiological measurement method, 2009, Journal of AOAC International, 92, 1763-1772



Annexe « Membres du groupe GT Val 2 »

Animateur :

Michel LAURENTIE, Laboratoire de Fougères, responsable de la plate-forme PAS

Membres experts :

Martine CHERBONNEL, Laboratoire de Ploufragan-Plouzané

Guillaume DUFLOS, Laboratoire de sécurité des aliments, site de Boulogne-sur-Mer

Valérie GAUDIN, Laboratoire de Fougères

Vincent HÉRAU, Laboratoire de santé végétale, site d'Angers

Frédéric HOMMET, Laboratoire de sécurité des aliments, site de Maisons-Alfort

Bertrand LOMBARD, Laboratoire de sécurité des aliments, site de Maisons-Alfort

Marina NICOLAS, Laboratoire de sécurité des aliments, site de Maisons-Alfort

Dominique PESSEL, Laboratoire de Fougères

Julie PETTON, Laboratoire de Ploufragan-Plouzané

Christophe ROSIN, Laboratoire d'hydrologie de Nancy

Audrey SCHMITZ, Laboratoire de Ploufragan-Plouzané

Éric VERDON, Laboratoire de Fougères

Représentante de la délégation à la qualité :

Catherine de MÈREDIEU

Représentantes de la direction des laboratoires :

Véronique DA-RIZ

Barbara GOUGET



Annexe « Bases statistiques pour les variables quantitatives »

Les ouvrages statistiques classiques considèrent que les variables quantitatives continues doivent respecter trois hypothèses statistiques pour pouvoir être analysées de façon correcte. Ces trois hypothèses sont : la normalité, l'indépendance des mesures et l'homogénéité des variances par niveau.

- Normalité des mesures: la distribution des observations doit suivre, par niveau, une loi normale.

Sur le plan statistique, il est difficile de vérifier correctement la normalité des données car il faut un nombre important de données. Cette hypothèse sera donc considérée comme acquise, exceptée dans quelques cas comme les dénombrements, les données issues de PCR quantitative ou les celles issues de dosage RIA qui sont transformées, généralement logarithmique, avant d'être exploitables. La non-normalité des données peut aussi être liée à la présence de valeurs atypiques. Il est possible de vérifier par un graphique adapté la présence de telles valeurs. L'utilisation de graphique de type box plot (boîte à moustache) est très commode. Un exemple est fourni sur la figure 9.

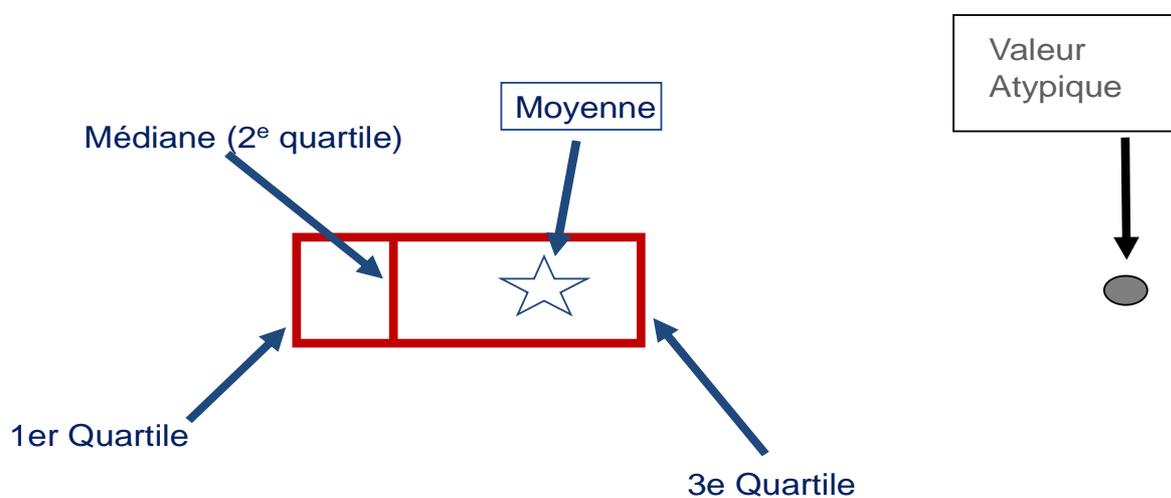


Figure 9 - Exemple d'informations fournies par une représentation graphique « box plot ».

Des tests statistiques, comme ceux de Grubbs ou de Dixon, peuvent être utilisés. L'utilisation d'un graphique ou d'un test statistique de présence de valeurs atypiques ne justifie pas l'exclusion de ces valeurs. Ces valeurs ne pourront être exclues que si des raisons (analytique, technique...) autres que des raisons statistiques existent.

- Indépendance des mesures : elle doit être assurée en inter-niveau et en intra-niveau.

L'indépendance des données est obtenue en veillant à ce que chaque échantillon, chaque répétition et chaque injection soit indépendante. Par exemple, chaque prise d'essai doit être réalisée de façon indépendante. De même, il faut faire attention à faire de vraies répétitions et non simplement répéter deux injections avec un passeur d'échantillons. Il faut également veiller aux dilutions en cascades, largement utilisées pour préparer des solutions-filles à partir de solution-mère ou pour diluer en série un échantillon. La propagation de l'erreur n'est pas forcément maîtrisée, rendant les résultats inexploitables.

- Homogénéité de la variance par niveau : l'hypothèse d'égalité des dispersions par niveau n'est vraiment importante que pour l'étalonnage et lorsqu'on applique la méthode des moindres carrés.

L'utilisation de différents niveaux de concentration d'étalonnage implique que la variance de la réponse instrumentale soit du même ordre de grandeur. Bien souvent, la variance n'est pas

constante en fonction de la concentration. Pour s'en assurer, une représentation graphique de la variance par niveau en fonction des niveaux peut être réalisée. La figure 10 montre un exemple d'inégalité des variances (non homogène) sur un domaine de concentration assez large.

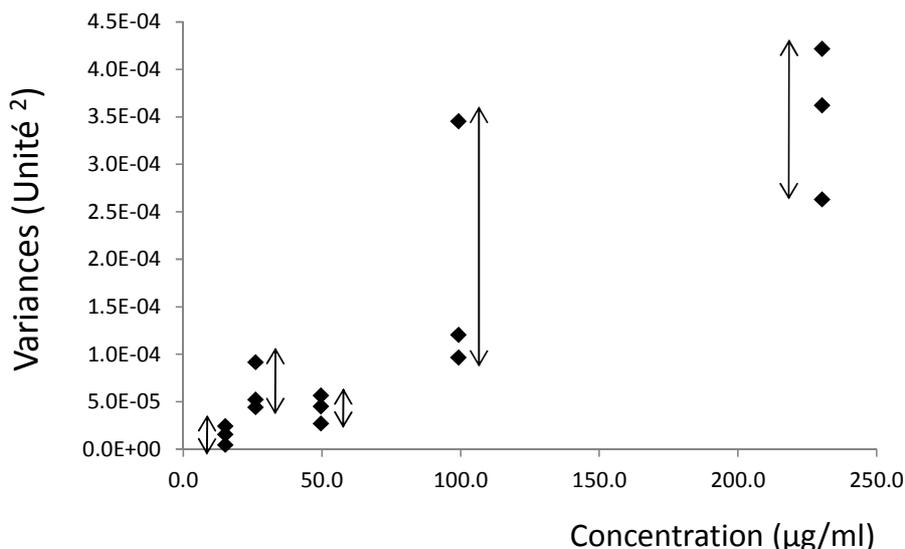


Figure 10 - Exemple de variances non homogènes.

La mise en évidence de variances non homogènes peut également se faire à l'aide de test statistique comme celui de Cochran ou de Lévène.

Si les variances ne sont pas homogènes, il faut avoir recours, soit à une transformation des données, soit à une pondération, afin de pouvoir stabiliser les variances et appliquer la méthode des moindres carrés sans problème.

La transformation de données la plus classique est la transformation logarithmique décimale. D'autres modèles de transformation existent, mais le GT Val 2 conseille d'utiliser la transformation logarithmique.

Une deuxième possibilité est l'utilisation de la pondération. Le principe est de donner du poids à une des deux variables (X ou Y) pour compenser l'effet de l'hétérogénéité. Il existe plusieurs types de pondération. La figure 11 illustre les différents types de pondération.

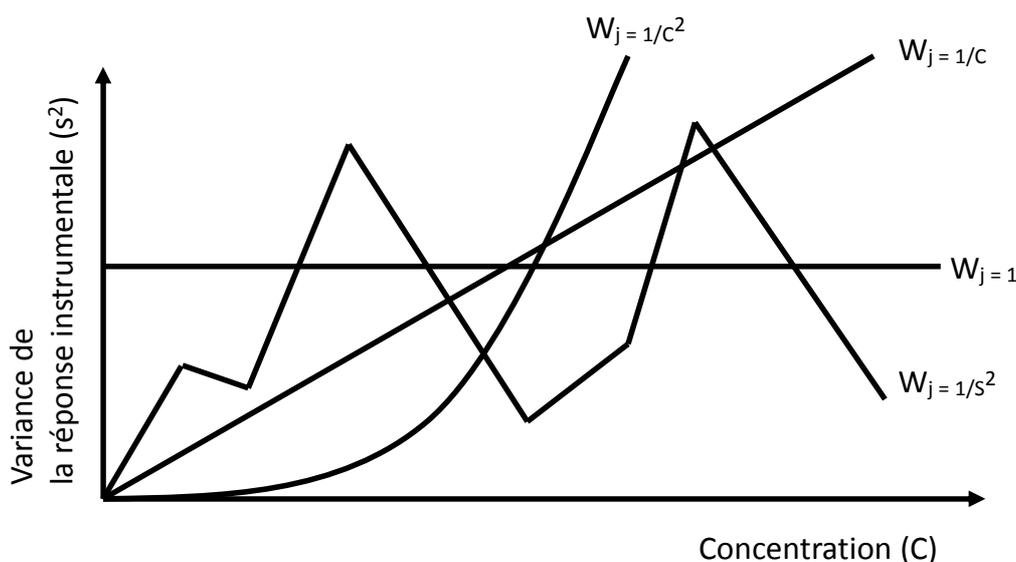


Figure 11 - Choix du facteur de pondération (w) en fonction de la relation entre la variance et le niveau de concentration.

Il existe quatre types de pondération (notées w_i) :

Les variances sont constantes, il n'y a pas besoin de pondération ($w_i = 1$).

Les variances (s_i^2) sont proportionnelles à la concentration (C_i), en conséquence la pondération sera fixée à $w_i = 1/C_i$.

Les sont proportionnelles au carré de la concentration, en conséquence la pondération sera fixée à $w_i = 1/C_i^2$

Les variances ne sont pas dans un rapport de proportionnalité avec la concentration, en conséquence la pondération sera fixée à $w_i = 1/s_i^2$

Les logiciels associés aux appareils de mesure proposent parfois ces pondérations et même l'imposent. Attention car imposer une pondération peut avoir l'effet contraire de l'effet recherché. Il faut également faire attention car l'association d'une transformation et d'une pondération ne conduit pas forcément à stabiliser la variance.

Sur la figure 12, l'ensemble de ces hypothèses de base est résumé.

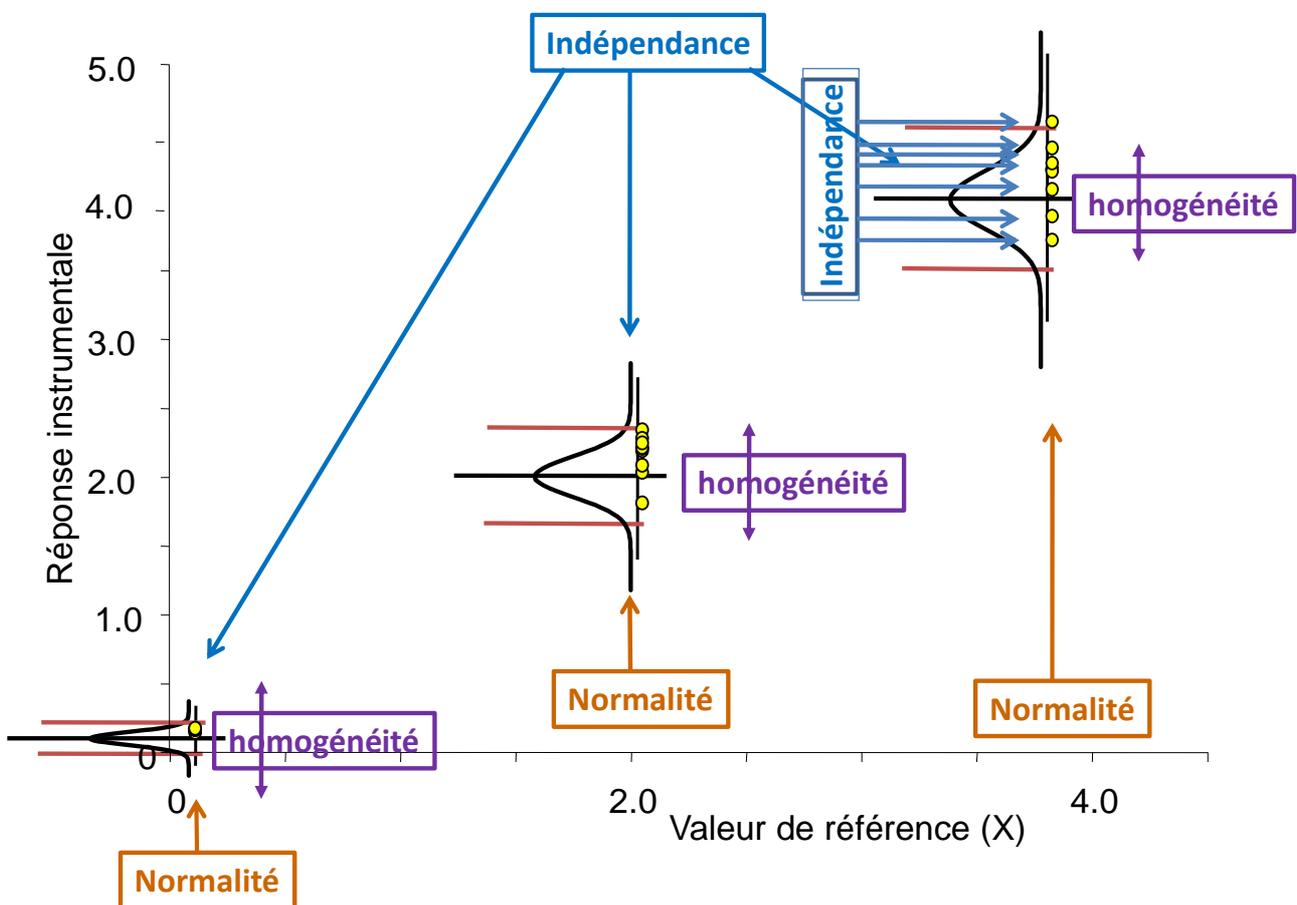


Figure 12 - Hypothèses statistiques requises pour les variables quantitatives.

Annexe « Spécificité »

Pour les méthodes quantitatives, la spécificité est la capacité d'une méthode à mesurer un analyte dans un échantillon sans que cette mesure soit faussée par d'autres composants de l'échantillon.

Un manque de spécificité peut être dû, par exemple, à des effets matrices, des interférences spectrales ou des recouvrements de pics en chromatographie.

Pratiquement, les interférences ont deux conséquences néfastes : soit elles entraînent une surestimation de la concentration de l'échantillon, car la réponse est plus élevée que ce qu'elle devrait être (ex : effet additif ou multiplicatif) ; soit elles causent une sous-estimation de la concentration, car le signal est partiellement masqué. Dans les deux cas, elles occasionnent donc un biais de justesse (erreur systématique). La caractéristique « justesse » permet de caractériser cette performance.

Pour les méthodes qualitatives, la spécificité (S_p) se calcule après avoir établi le tableau de contingence des résultats (tableau 4).

Tableau 4 - Synthèse des résultats obtenus lors de l'étude de caractérisation des performances.

		Statut de l'échantillon	
		Positif	Négatif
Résultat donné par le test à caractériser	Positif	PA = accords positifs = vrais positifs	PD = déviations positives = faux positifs
	Négatif	ND = déviations négatives = faux négatifs	NA = accords négatifs = vrais négatifs
Sommes		N+	N-

La spécificité est le pourcentage (%) de résultats négatifs trouvés parmi les négatifs attendus ; elle se calcule par la formule suivante :

$$S_p = \frac{NA}{NA+PD}$$

Il est possible de déterminer l'erreur commise (intervalle de confiance) sur ce paramètre en appliquant la formule suivante :

$$eS_p = \pm 1.96 \sqrt{\frac{Sp(1 - Sp)}{n}}$$

avec (eS_p) l'erreur au niveau de confiance de 95% estimée en fonction du nombre d'échantillons analysés. Cependant, ce calcul n'est valide que dans le cas où les deux conditions suivantes sont vérifiées :

le produit $n \times S_p > 5$ et $n \times (1 - S_p) > 5$



Dans le cas où les conditions ne sont pas vérifiées, il faut se reporter au tableau de correspondance de Schwartz, 1993 (intervalle de confiance à 95%) présent dans la norme XP U47-600-2.

Il est également possible d'utiliser des logiciels statistiques appropriés (R, SAS...) pour calculer ces intervalles de confiance, en particulier un intervalle de confiance binomial exact.

Exemple : Pour une spécificité mesurée de 95% en analysant 101 échantillons connus comme négatifs, l'erreur (e_{Sp}), au niveau de confiance de $1-P(\alpha) = 95\%$, est estimée à plus ou moins 4% (correspondant à un intervalle de confiance de S_p compris entre 91% à 99%), ceci en respectant les conditions d'application de la formule : $101 \times 0.95 = 95.95$ et $101 \times (1-0.95) = 5.05$ (valeurs 5).

Pour les petits échantillons, il est possible d'utiliser la table suivante (Tableau 5) extraite de (Schwartz, 1993) afin de calculer l'intervalle de confiance à 95% d'un pourcentage de S_p .

Tableau 5 - Intervalle de confiance à 95% d'un pourcentage de S_p .

Effectif de l'échantillon	Pourcentage observé									
	5	10	15	20	25	30	35	40	45	50
10	-	0-45	1-50	3-56	5-60	7-65	9-70	12-74	15-78	19-81
20	0-25	1-32	3-38	6-44	9-49	12-54	15-59	19-64	23-68	27-73
30	0-20	2-27	5-33	8-39	11-44	15-49	19-54	23-59	27-64	31-69
40	1-17	3-24	6-30	9-36	13-41	17-47	21-52	25-57	29-62	34-66
50	1-15	3-22	6-28	10-34	14-39	18-45	22-50	26-55	31-60	36-64
60	1-14	4-21	7-27	11-32	15-38	19-43	23-48	28-53	32-58	37-63
70	1-13	4-20	8-26	11-31	15-37	20-42	24-47	28-52	33-57	38-62
80	1-12	4-19	8-25	12-30	16-36	20-41	25-46	29-52	34-57	39-61
90	2-12	5-18	8-24	12-30	16-35	21-41	25-46	30-51	34-56	39-61
100	2-11	5-18	9-24	13-29	17-35	21-40	26-45	30-50	35-55	40-60
150	2-10	6-16	10-22	14-27	18-33	23-38	27-43	32-48	37-53	42-58
200	2-9	6-15	10-21	15-26	19-32	24-37	28-42	33-47	38-52	43-57
500	3-7	8-13	12-18	17-24	21-29	26-34	31-39	36-44	41-49	46-54
1 000	4-7	8-12	13-17	18-23	22-28	27-33	32-38	37-43	42-48	47-53
2 000	4-6	9-11	13-17	18-22	23-27	28-32	33-37	38-42	43-47	48-52



Annexe « Sensibilité »

Pour les méthodes qualitatives, la sensibilité (S_e) se calcule après avoir établi le tableau de contingence des résultats (tableau 6).

Tableau 6 Synthèse des résultats obtenus lors de l'étude de caractérisation des performances.

		Statut de l'échantillon	
		Positif	Négatif
Résultat donné par le test à caractériser	Positif	PA = accords positifs = vrais positifs	PD = déviations positives = faux positifs
	Négatif	ND = déviations négatives = faux négatifs	NA = accords négatifs = vrais négatifs
Sommes		N+	N-

La sensibilité est le pourcentage (%) de résultats positifs trouvés parmi les positifs attendus et se calcule par la formule :

$$S_e = \frac{PA}{PA+ND}$$

Il est possible de déterminer l'erreur commise (intervalle de confiance) sur ce paramètre en appliquant la formule suivante :

$$eS_e = \pm 1.96 \sqrt{\frac{S_e(1 - S_e)}{n}}$$

avec (eS_e) l'erreur au niveau de confiance de 95% estimée en fonction du nombre d'échantillons analysés. Cependant, ce calcul n'est valide que dans le cas où les deux conditions suivantes sont vérifiées :

le produit $n \times S_e > 5$ et $n \times (1 - S_e) > 5$

Il est également possible d'utiliser des logiciels statistiques appropriés (R, SAS...) pour calculer ces intervalles de confiance, en particulier un intervalle de confiance binomial exact.

Dans le cas où les conditions ne sont pas vérifiées, il faut se rapporter au tableau 5 de correspondance fourni dans l'annexe « Spécificité ».



Annexe « Fonction d'étalonnage/Efficacité (PCR)»

Le modèle d'étalonnage permet de relier la réponse instrumentale (Y) à la variable explicative (X). Le modèle établi lors de la validation est celui qui sera utilisé en routine. En termes statistiques, le modèle sera estimé par des paramètres et ce modèle doit être choisi au mieux pour que la relation entre X et Y soit cohérente et forte.

Pour établir une fonction d'étalonnage, il faut plusieurs niveaux de concentration de solutions étalons. Ces niveaux peuvent être établis avec ou sans matrice. Ils sont communément appelés « standard d'étalonnage ». Dans la mesure où, généralement, les paramètres du modèle sont estimés par la méthode des moindres carrés, il est préférable de prendre un niveau de plus que de coefficients à estimer (voir tableau 7). Par exemple, si la relation d'étalonnage est un modèle linéaire, défini par deux coefficients, il est judicieux de prendre 3 niveaux. Pour une fonction d'étalonnage d'équation logistique à 4 coefficients, il faut au minimum avoir 5 niveaux (voir tableau 7).

La répartition des points dépend du type de modèle. Une répartition équidistante n'est pas forcément une répartition optimale. Dans le cas d'un domaine étendu, des points peu espacés, près des LOQ inférieures et supérieures, sont souhaitables. Une étude pilote sur la fonction d'étalonnage permettra d'optimiser le plan expérimental final. Dans le cas où une étude pilote ne pourra être réalisée, le nombre de niveaux sera augmenté.

Pour chaque niveau, il sera réalisé des répétitions (répétitions intra jour) et chaque série sera également répétée. En effet, ces conditions de répétabilité et de fidélité intermédiaire permettront d'identifier les sources de variations et ainsi permettront d'optimiser le modèle. Le nombre minimal de répétitions en conditions de répétabilité est fixé à deux et le nombre minimal de séries en conditions de fidélité intermédiaire est fixé à 3.

La fonction d'étalonnage peut être établie avec ou sans la matrice. Si un effet matrice est étudié, un plan équilibré est souhaitable pour les niveaux et pour les répétitions. En conséquence, il y aura au minimum 3 niveaux avec 2 répétitions par niveau et 3 séries réalisées avec ou sans matrice ce qui fait un total de 36 mesures à réaliser.

Dans le tableau 7 sont indiquées les différentes combinaisons possibles de niveaux, de séries et de répétitions en fonction de quelques modèles classiques de fonction d'étalonnage.

Tableau 7 - Modèles usuels de fonctions d'étalonnages, calcul du modèle inverse, nombre minimal de niveau nécessaires, de répétitions par niveaux et nombre minimal de séries pour établir ce modèle.

Fonction	Étalonnage direct	Étalonnage inverse	Nombre de coefficients	Nombre minimal de niveaux	Nombre de répétitions par niveau	Nombre minimal de répétitions par série
Linéaire	$y = a + bx$	$x = \frac{y - a}{b}$	2	3	2	3
Quadratique	$y = a + bx + cx^2$	$x = \frac{-b + \sqrt{b^2 - 4c(a - y)}}{2c}$	3	4	2	3
Puissance	$y = a \times x^b$	$x = 10^{\frac{\log(y) - \log(a)}{b}}$	2	3	2	3
Logistique à 4 paramètres	$y = b + \frac{a - b}{\left(1 + \frac{x}{c}\right)^d}$	$x = c \left[\frac{a - b}{y - b} - 1 \right]^{\frac{1}{d}}$	4	5	2	3



Généralement, sur le plan statistique, la qualité de la relation entre la variable X et la variable Y est décrite par le calcul du coefficient de corrélation (r) ou le coefficient de détermination (r^2). Ce ne sont pas des paramètres pertinents pour apprécier cette relation et ce ne sont pas de bons indicateurs d'une relation de proportionnalité entre les variables X et Y.

Une bonne approche pour sélectionner un modèle d'étalonnage est l'étude de la distribution des résidus standardisés entre les valeurs obtenues et les valeurs prédites par le modèle. Par définition, les résidus standardisés sont les résidus « normalisés » par leur écart-type. Si l'écart-type est déterminé de façon empirique, le résidu est appelé résidu « studentisé ». La figure 13 présente un exemple de distribution de résidus. La distribution de ces résidus studentisés doit être dans l'intervalle ± 2 et ne doit pas être structurée.

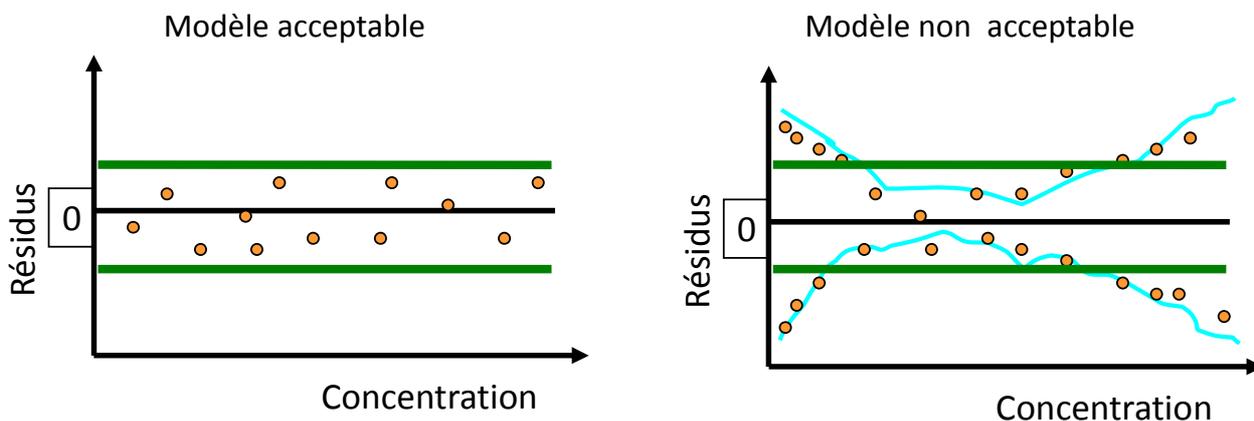


Figure 13 - Exemple de distributions de résidus (valeurs trouvées - valeurs attendues) obtenues pour différents modèles d'étalonnage.

Il peut être également vérifié que l'erreur générée par le modèle ne soit pas supérieure à l'erreur expérimentale car cela démontrerait que l'erreur vient essentiellement du modèle et non de la méthode. Cette analyse peut se faire par une analyse de la variance (ANOVA).

Les tests statistiques utilisés pour vérifier l'existence d'une pente et d'une ordonnée à l'origine ne sont pas conseillés car les tests utilisés sont perfectibles et conduisent parfois à rejeter à tort l'existence d'une pente.

Annexe « Justesse »

Elle s'étudie par l'expression de la différence entre l'ensemble des valeurs mesurées à l'infini et une valeur de référence. La vraie valeur de la justesse est déterminée par rapport à un matériau de référence certifié (MRC). Dans le cas d'absence de MRC, une valeur qui est acceptée comme une valeur conventionnellement vraie peut être utilisée.

La justesse est étudiée pour chaque niveau de concentration avec ou sans la matrice. Ils sont communément appelés « standards de validation ». Par définition, la justesse nécessite la réalisation de répétitions. Le nombre minimal est de 2, si la justesse est étudiée sur 3 niveaux. Cependant, ce nombre peut être augmenté si cela est une exigence réglementaire ou si la variabilité est importante.

Il existe plusieurs modes d'expression de la justesse (tableau 8) car le paramètre de validation qui la caractérise est le biais et il peut être exprimé de façon absolue ou relative. Ils sont tous équivalents.

Tableau 8 - Différentes expressions de la justesse.

Paramètre de validation de la justesse	Expression	Équation
Biais de justesse (absolu)	B	$\bar{y} - V$
Biais de justesse (relatif)	$B (\%)$	$b = \frac{\bar{y} - V}{V} \times 100$
Taux de récupération (relatif) (taux de recouvrement/récupération)	$TR(\%)$	$TR = \frac{\bar{y}}{V} \times 100$

avec \bar{y} la moyenne générale de toutes les mesures effectuées sur le même matériau de référence de valeur V .

L'intervalle de confiance de la moyenne générale est obtenu par la formule suivante :

$$\left[\bar{y} - t_{1-\frac{\alpha}{2}, I-1} \times \sqrt{\frac{J(s_{FI}^2 - s_r^2) + s_r^2}{IJ}}; \bar{y} + t_{1-\frac{\alpha}{2}, I-1} \times \sqrt{\frac{J(s_{FI}^2 - s_r^2) + s_r^2}{IJ}} \right]$$

avec $t_{1-\frac{\alpha}{2}, I-1}$ le percentile d'une distribution de Student pour un nombre de degrés de liberté $I - 1$ et un niveau de confiance $(1 - \alpha)$, s_r^2 , s_{FI}^2 , sont des variances de répétabilité et de fidélité intermédiaire.

Si cet intervalle de confiance est noté $[L; U]$, l'intervalle de confiance du biais de justesse est déduit $[L - V; U - V]$.

Lorsque la justesse est étudiée sur plusieurs niveaux ($n \geq 3$), il est conseillé de tracer la droite de justesse qui permet de vérifier :

- s'il existe une proportionnalité entre les valeurs de référence et les quantités retrouvées estimées par les moyennes générales. Cette proportionnalité est par définition une estimation de la linéarité de la justesse, souvent requise pour conclure à la validité d'une méthode de mesure ;
- si les quantités retrouvées sont proches des valeurs de référence et si la méthode présente, en général, un défaut de justesse.



La comparaison se fait par rapport à la première bissectrice qui est un indicateur de l'absence d'un défaut de justesse.

La droite de justesse est calculée selon les principes décrits dans « Labo-Stat : Guide de validation des méthodes d'analyse » (M. Feinberg, 2009).

Compte tenu de l'erreur associée, il peut exister un décalage entre la droite de justesse et la première bissectrice : cela traduit un défaut de justesse. Afin de savoir si statistiquement la droite de justesse est différente de cette bissectrice, il convient de calculer l'intervalle de confiance de cette droite.

La figure 14 illustre une méthode avec et sans biais de justesse.

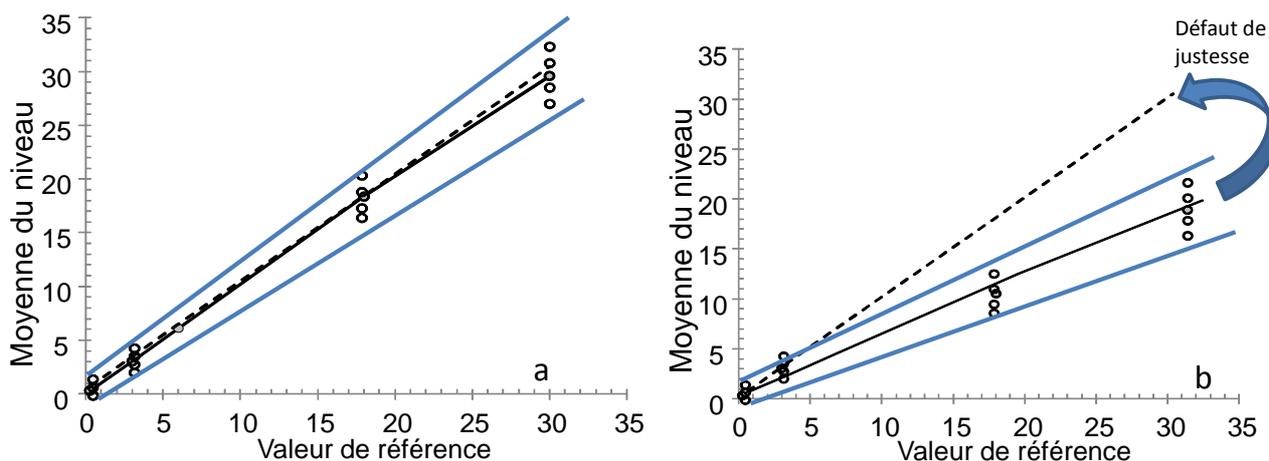


Figure 14 - Exemple de droite de justesse et son intervalle de confiance.

Le cartouche (a) montre une méthode sans biais (l'intervalle de confiance couvre la 1^{ère} bissectrice (trait pointillé)).

La méthode représentée en (b) présente une erreur de justesse.

En présence d'un défaut de justesse, il est possible de calculer sur la droite de justesse un facteur de correction. Ce facteur de correction est égal à l'inverse de la pente de la droite, pente qui peut être estimée par une régression linéaire classique (voir exemple acrylamide dans l'annexe « Exactitude »). Si ce facteur est constant sur l'ensemble des niveaux, il est applicable sur tout le domaine de validité.

Dans le cas de l'utilisation d'un facteur de correction, il est conseillé de préparer à nouveau des standards de validation et de refaire la validation de la méthode. En effet, l'utilisation d'un facteur de correction impacte le calcul de l'incertitude et ce calcul ne peut être réalisé sur les standards de validation qui ont servi à estimer ce facteur de correction. L'avantage de refaire cette validation est de valider, dans le processus global, le facteur de correction qui sera utilisé en routine.

Annexe « Fidélité »

La fidélité exprime l'étroitesse de l'accord entre une série de mesures provenant de multiples prises d'un même échantillon homogène dans des conditions prescrites. La fidélité fournit une indication sur les erreurs liées au hasard. La fidélité peut être évaluée à trois niveaux : la répétabilité, la fidélité intermédiaire (intra laboratoire) et la reproductibilité (inter laboratoires).

La fidélité traduit uniquement la distribution des erreurs aléatoires et n'a aucune relation avec la valeur vraie ou spécifiée. La mesure de fidélité est calculée à partir de l'écart-type des résultats d'essais et peut être exprimée sous forme de coefficients de variation (CV). Les mesures quantitatives de la fidélité dépendent des conditions stipulées. Il est classiquement distingué les conditions de :

- répétabilité : conditions où les résultats obtenus par la même méthode sur des échantillons d'essai identiques dans le même laboratoire, par le même opérateur, utilisant le même équipement et pendant un court intervalle de temps ;
- fidélité intermédiaire : conditions où les résultats sont obtenus par la même méthode sur des échantillons d'essai identiques dans le même laboratoire, avec différents opérateurs utilisant des équipements différents et pendant un intervalle de temps donné ;
- reproductibilité : conditions où les résultats sont obtenus par la même méthode sur des échantillons d'essais identiques dans différents laboratoires, avec différents opérateurs et utilisant des équipements différents.

Les conditions de répétabilité et de reproductibilité sont des ensembles particuliers de conditions extrêmes. La reproductibilité est établie au moyen d'une étude inter-laboratoires. La manière dont la fidélité intermédiaire doit être établie dépend néanmoins des circonstances dans lesquelles la procédure d'analyse est utilisée en routine. Les variations typiques à étudier sont les effets jour, opérateur, équipement, etc. Ces effets ne doivent pas être étudiés individuellement. L'utilisation d'un plan d'expérience est recommandée.

Il convient d'étudier simultanément la répétabilité et la fidélité intermédiaire. Par définition, il est nécessaire d'avoir des répétitions. Le nombre de répétitions minimal par niveau est de 2. Dans le cas où il y a une forte variabilité, le nombre de répétitions peut être porté à 3 ou plus, notamment si le nombre de niveaux étudiés est de 3. Le nombre de série minimal est de 3, quel que soit le nombre de niveaux étudiés.

Certaines réglementations, référentiels ou normes définissent, selon les domaines, le nombre de répétitions et le nombre de séries. Le nombre de séries et le nombre de répétitions par niveau sont liés et dépendent de la variabilité de la méthode. De plus, si la variabilité issue de la répétabilité est faible par rapport à celle générée par la fidélité intermédiaire, le plan d'expérience devra par exemple inclure plus de séries et simplement un nombre minimal de 2 répétitions par niveau. Une étude pilote ou de pré-validation est conseillée pour pouvoir optimiser le plan expérimental en fonction des sources de variations.

Répétabilité

L'étude de la répétabilité permet de déterminer la variabilité intra-série dans des conditions de répétabilité bien définies (même opérateur, même matériel, même laboratoire, même série etc.).

L'étude de la répétabilité ne peut être faite seule mais doit prendre en compte la fidélité intermédiaire et / ou la reproductibilité. Une étude de répétabilité sans prendre en compte ces sources d'erreur conduit à une estimation biaisée de la répétabilité.



Les calculs sont basés sur les fascicules de la norme ISO/CEN 5725. Dans cette norme, le paramètre de répétabilité estimé est exprimé sous forme :

- d'écart-type de répétabilité σ_r et sa variance σ_r^2
- de coefficient de variation
- de limite de fidélité.

Dans la norme ISO 5725-3, le modèle statistique simplifié est :

$$Y_{ij} = m + L_i + e_{ij}$$

Avec m qui représente la moyenne des Y_{ij} , L_i représente les effets facteurs qui suivent une loi $N(0, \sigma_L^2)$. En ce qui concerne la validation de méthode analytique, le facteur L est le facteur série. Les e_{ij} représentent les erreurs aléatoires qui suivent une loi $N(0, \sigma_e^2)$.

Les calculs sont explicités dans la norme ISO 5725-3 et sont résumés dans l'annexe « Reproductibilité ».

Fidélité intermédiaire

L'étude de la fidélité intermédiaire (FI) (appelée reproductibilité intra-laboratoire en microbiologie des aliments) consiste à évaluer, dans une caractérisation interne, la variabilité inter-séries. Les sources de variations sont nombreuses, comme par exemple l'appareillage, l'opérateur etc. Classiquement, l'effet jour est principalement étudié.

Les calculs sont basés sur la norme ISO 5725-3. Dans cette norme, le paramètre de fidélité intermédiaire estimé est exprimé sous la forme :

- d'écart-type de fidélité intermédiaire σ_{FI} et sa variance σ_{FI}^2
- de coefficient de variation
- de limite de fidélité.

Dans la norme ISO 5725-3, le modèle statistique simplifié est :

$$Y_{ij} = m + L_i + e_{ij}$$

avec m qui représente la moyenne des Y_{ij} , L_i représente les effets facteurs qui suivent une loi $N(0, \sigma_L^2)$. En ce qui concerne la validation de méthode analytique, le facteur L est le facteur série. Les e_{ij} représentent les erreurs aléatoires qui suivent une loi $N(0, \sigma_e^2)$.

Les calculs sont explicités dans la norme ISO 5725-3. Les calculs seront résumés dans l'annexe « Reproductibilité ».

Reproductibilité

L'étude de la reproductibilité permet de caractériser la variabilité totale incluant la variabilité liée aux laboratoires utilisant la méthode. Cette caractérisation se fait au moyen d'un essai interlaboratoires de validation. Les six fascicules de la norme ISO 5725 décrivent les processus et les calculs statistiques à mener. Ces calculs sont également repris et complétés dans le chapitre 6 de « Labo-Stat : Guide de validation des méthodes d'analyse » (M. Feinberg, 2009).

Le modèle défini est le modèle théorique issu de la norme ISO 5725. Le modèle statistique simplifié est :

$$Y_{ij} = m + L_i + e_{ij}$$



avec m qui représente la moyenne des Y_{ij} , L_i représente les effets facteurs (dans ce cas le laboratoire) qui suivent une loi $N(0, \sigma_L^2)$. Les e_{ij} représentent les erreurs aléatoires qui suivent une loi $N(0, \sigma_e^2)$.

Le plan d'expérience, basé sur des échantillons identiques reçus par les laboratoires, doit comporter au moins huit laboratoires ayant réalisé 2 à 3 répétitions (IUPAC).

L'analyse statistique repose sur une analyse de variance à un facteur. Il convient d'imposer le nombre de répétitions aux laboratoires afin d'avoir un plan d'expérience équilibré.

La décomposition de la variance totale en variance intra-laboratoire et en variance interlaboratoires permet d'identifier les sources de variabilité. En effet, la variabilité totale est la somme de la variabilité intra-laboratoire et de la variabilité interlaboratoires.

$$SCE_T = SCE_r + SCE_L$$

avec SCE_T : somme totale des carrés des écarts ; SCE_r : somme des carrés intra-laboratoires ; SCE_L : somme des carrés inter laboratoires.



Annexe « Exactitude »

Par définition, l'exactitude exprime l'étroitesse de l'accord entre le résultat et la valeur de référence acceptée, aussi appelée « valeur vraie ». L'étroitesse de l'accord ainsi observée est la résultante de la somme des erreurs systématiques et aléatoires, en d'autres termes l'erreur totale liée au résultat. Par conséquent, l'exactitude est l'expression de la somme de la justesse et de la fidélité. De ce fait, c'est la caractéristique de performance majeure à étudier dans une approche globale.

Un outil graphique utilisant une approche statistique basée sur le calcul d'un intervalle statistique de dispersion a été développé. Cet outil a été appelé « profil d'exactitude » et il permet de statuer sur la validation de la performance de la méthode. Proposé en 2003 par une commission de la société française des sciences et techniques pharmaceutiques (SFSTP), il est utilisé dans de nombreux domaines tels que la physico-chimie, mais également le dénombrement bactérien par méthode classique, l'ELISA et la PCR quantitative. Il est reconnu au niveau des instances de normalisation ISO et AFNOR et a été introduit dans des normes AFNOR telles que les normes NF V03-110, NF 47-600-2, et la nouvelle version de la norme ISO 16140-2.

L'exactitude est la combinaison des caractéristiques de justesse (biais ou erreur systématique) et de fidélité (erreur aléatoire). Ce sont des critères de performance d'une méthode dont l'analyse séparée ne garantit pas nécessairement que la méthode soit apte pour ce qu'elle doit faire en routine. La validation de la méthode doit permettre de vérifier que les critères de performance sont en adéquation avec l'objectif de la méthode. De plus, si à l'issue de l'étude, la méthode n'est pas validée, il serait intéressant d'identifier les sources d'erreur.

Pour l'analyste et le statisticien, ainsi que pour le client, une bonne méthode signifie que le résultat fourni ne s'écarte pas trop de la vraie valeur, ce qui peut s'écrire (éq 1.) :

$$x - \mu_T \quad \text{Équation 1}$$

avec μ_T la valeur vraie.

Pour accepter que cette différence soit faible, il faut que cette différence soit bornée par des limites d'acceptation notées $\pm\lambda$, ce qui peut s'écrire (éq. 2) :

$$-\lambda < x - \mu_T < \lambda \Leftrightarrow |x - \mu_T| < \lambda \quad \text{Équation 2}$$

Deux notions distinctes apparaissent : (i) une limite d'acceptation de performance et (ii) une décision par l'analyste d'accepter ou de rejeter une méthode en fonction de ses performances.

Garantir les résultats signifie de fixer le risque attendu de mesures en dehors des limites d'acceptation. Ce risque doit être fixé *a priori* et se traduira par une probabilité (éq. 3) :

$$(P|x - \mu_T| < \lambda) \geq \beta \quad \text{Équation 3}$$

avec β la proportion minimale fixée *a priori* de mesures dans les limites d'acceptation $\pm\lambda$.

La réalisation du plan d'expérience pour évaluer les critères de performance implique généralement des répétitions et plusieurs niveaux de concentrations. De ce fait, la garantie des résultats doit se faire pour un intervalle de concentration. Pour satisfaire cette condition, il convient de construire un intervalle de tolérance de mesures attendues sur le domaine couvert par les concentrations. Cet intervalle de confiance de mesures attendues au niveau β s'exprime par (éq 4) :

$$E_{\hat{\mu}_M, \hat{\sigma}_M} \{P[|x - \mu_T| < \lambda] / \hat{\mu}_M, \hat{\sigma}_M\} \geq \beta \quad \text{Équation 4}$$

avec E signifiant « valeur attendue » calculée à l'instant des mesures en fonction des estimateurs du biais ($\hat{\mu}_M$) et de la variance ($\hat{\sigma}_M$) à disposition de l'analyste. La figure 15 illustre un profil d'exactitude.



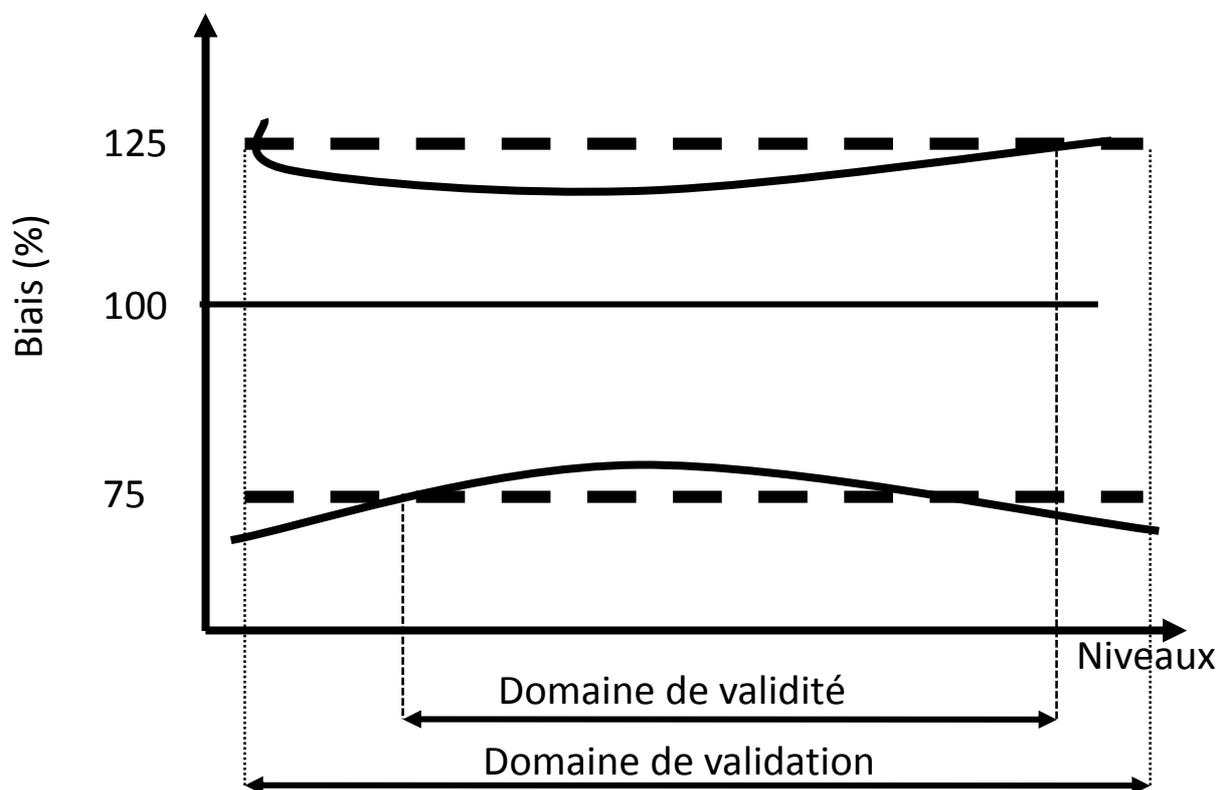


Figure 15 – Illustration d'un profil d'exactitude.

Les lignes pointillées délimitent l'intervalle d'acceptabilité à $\pm \lambda$, ici $\pm 25\%$, c'est-à-dire où le biais peut varier entre 75 et 125 %. Cet intervalle représente le domaine de validation. Les lignes continues représentent l'intervalle de tolérance ou profil d'exactitude. L'intersection entre les lignes d'acceptabilité et les bornes de l'intervalle de tolérance délimitent le domaine de validité. À l'intérieur de ce domaine, la méthode est capable de produire une proportion élevée et connue de résultats acceptables

Choix des limites d'acceptabilité et de la proportion β :

Les limites d'acceptabilité vont être définies dans la mesure du possible, en se référant à un référentiel, à une expérience professionnelle ou une exigence scientifique.

– **Référentiel.** Plusieurs textes réglementaires fournissent des valeurs qui peuvent servir de base pour fixer λ . Par exemple, la décision de la commission 2002/657 indique dans son tableau 2 la justesse minimale des méthodes quantitatives pour les médicaments et résidus de médicaments vétérinaires dans les produits d'origine animale. D'autres documents publiés par la commission du Codex Alimentarius définissent une « approche critère » pour différents analytes et matrices qui devrait permettre une clarification de la situation au niveau international.

– **Expérience professionnelle.** Certaines professions ont défini des limites d'acceptabilité sous la forme d'écart maximum acceptable (EMA), comme cela se pratique pour le contrôle de qualité ou la métrologie légale. C'est une approche classique en biologie médicale pour les paramètres du sang ou dans l'industrie pharmaceutique pour le contrôle des spécialités pharmaceutiques.

– **Modèle empirique.** Il existe des modèles mathématiques qui relient des performances analytiques à la concentration. Le plus connu est celui développé par Horwitz, modifié par Thompson (2000). Il est repris dans de nombreux textes et permet de prédire la valeur attendue de l'écart-type de reproductibilité S_R en fonction d'un niveau de concentration C .

En fonction du niveau de concentration, il y a plusieurs modèles pour l'estimer :

- pour une concentration inférieure à $1.2 \cdot 10^{-7} \text{ kg.kg}^{-1}$, il s'exprime sous la forme, $S_R = 0.22 C$
- pour une concentration comprise entre $1.2 \cdot 10^{-7} \text{ kg.kg}^{-1}$ et 0.138 kg.kg^{-1} , il s'exprime sous la forme, $S_R = 0.22 C^{0.8495}$
- pour une concentration supérieure à 0.138 kg.kg^{-1} , il s'exprime sous la forme, $S_R = 0.01 C^{0.5}$.

– **À partir de la limite de quantification.** Lorsque l'objectif de la méthode est l'application d'une norme sanitaire, il est possible de déduire des limites d'acceptabilité, à partir de la limite de quantification calculée avec le profil d'exactitude. La décision de validation se fera alors en fonction du risque encouru, obtenu *a posteriori*.

Les limites d'acceptabilité sont souvent exprimées sous la forme d'un pourcentage mais ce n'est pas une obligation. De même, pour simplifier, il est classiquement choisi une valeur constante pour tout le domaine d'application. Mais, si le domaine d'application est large, des valeurs variant en fonction du niveau de concentration peuvent être utilisées.

Dans le cas où les valeurs de référence sont connues avec une incertitude élevée, le choix des limites d'acceptabilité devra être compatible avec cette incertitude.

La proportion β . Elle représente la proportion de futurs résultats qui seront en moyenne compris dans les intervalles de tolérance. La valeur choisie pour β dépend largement du champ d'application. Il est évident que plus β est petit, par exemple 70%, plus la méthode risque de produire des résultats qui ne correspondent pas aux spécifications annoncées. C'est pourquoi il est préférable de fixer cette proportion au moins à 80%.

Cet outil permet de décider de la validité ou non de la méthode en superposant le domaine de validation souhaité et le domaine de validité acceptable en fonction des critères de performance de la méthode.

Elle permet également d'être un outil de diagnostic comme le montre l'exemple suivant.

Exemple d'application : dosage de l'acrylamide dans le plasma de porc

L'acrylamide est un produit neurotoxique néoformé lors de la cuisson de certains aliments. Il est issu d'une réaction de Maillard par combinaison de sucre (i.e. glucose) et de certains acides aminés comme l'asparagine. Si de nombreuses études documentent la teneur en acrylamide dans les aliments, peu d'études ont permis d'en quantifier l'absorption après ingestion. Une étude de pharmacocinétique, réalisée chez le porc doit permettre de déterminer la biodisponibilité de l'acrylamide. Au préalable une méthode analytique est nécessaire pour quantifier dans le plasma de porc les concentrations en acrylamide. La méthode retenue pour effectuer ce dosage est une méthode HPLC associée à une détection par spectrométrie de masse (MS). La gamme étudiée varie de 10 à 5000 ng/mL, compte tenu de l'absence d'information sur les taux plasmatiques circulants après ingestion d'aliment contenant de l'acrylamide.

Deux gammes sont préparées : une gamme standard d'étalonnage et une gamme dans du plasma de porc (standard de validation). A 200 μL de plasma est ajouté 100 μL de solution saturée de ZnSO_4 puis 1000 μL d'acétonitrile et 100 μL de standard interne (acrylamide D5). Après agitation et centrifugation, le surnageant est évaporé. L'éluat est repris avec 200 μL d'acétate d'ammonium 0,01M pH 6. Le volume d'injection est de 50 μL . Les conditions analytiques utilisées sont un débit chromatographique de 0,2 mL/min à travers une colonne Hypercarb (5 μ 50-



2 mm) et une détection MS sur l'ion moléculaire 72 de l'acrylamide. La gamme standard d'étalonnage en acrylamide varie de 10 à 5000 ng/mL et est réalisée dans une solution d'acétate d'ammonium 0,01M ramenée à pH 6 avec de l'acide formique.

Le plan d'expérience utilisé consiste en 3 jours, 6 niveaux et 2 répétitions (3x6x2) soit 36 essais pour les standards d'étalonnage et pour les standards de validation.

La figure 16 présente les réponses obtenues en fonction des concentrations introduites pour la gamme d'étalonnage et pour la gamme de standards de validation.

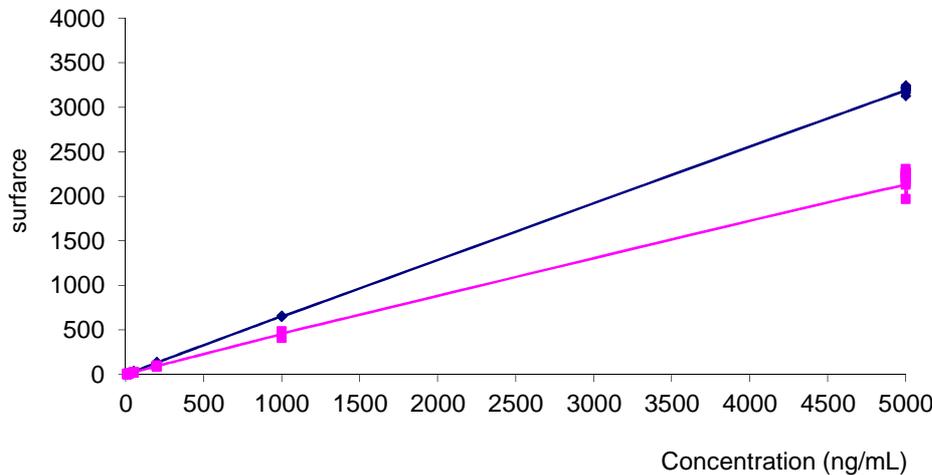


Figure 16 – Dosage d'acrylamide dans le plasma de porc : données brutes observées.

Les résultats sont obtenus sur 3 jours et pour deux répétitions par niveau. A noter l'effet matrice entre les standards d'étalonnage (courbe et point bleu) et les standards de validation (courbe et carré fuchsia).

Les données ont été traitées en trois étapes :

1. une analyse des données brutes et tracé du profil d'exactitude ;
2. une analyse pour déterminer le facteur de correction de l'effet de matrice (soit l'inverse du taux de recouvrement) ;
3. le calcul des profils d'exactitude après correction des concentrations.

Analyse des données brutes : La fonction de réponse utilisée pour décrire la relation entre les concentrations et la réponse est une régression quadratique pondérée (facteur de pondération de type 1/X avec X = concentration introduite). La figure 17 présente le profil d'exactitude obtenu.

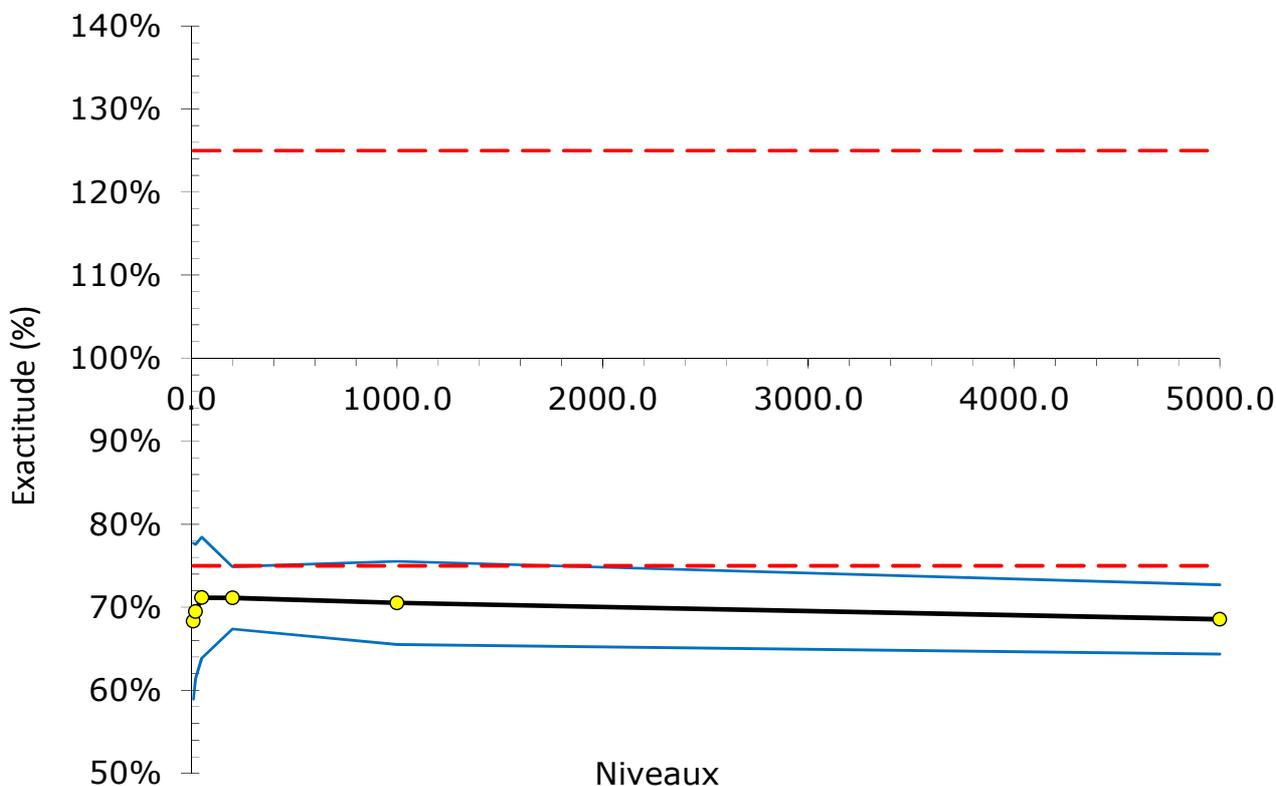


Figure 17 - Profil d'exactitude réalisé avec les données brutes.

Les résultats sont obtenus avec une régression quadratique pondérée pour modéliser la fonction de réponse.

Un décalage est observé entre le profil d'exactitude et les limites d'acceptabilité qui ont été fixées à 25%. Ce décalage est dû à l'effet matrice. Il faut donc estimer un facteur de correction correspondant au biais de la méthode.

Pour faire ce calcul, on utilise une équation de type $ax+b$ entre les valeurs théoriques et les valeurs calculées en retour. La pente de l'équation entre les concentrations en retour et les concentrations théoriques correspond à l'inverse du facteur de correction considéré. La figure 18 illustre ces calculs.

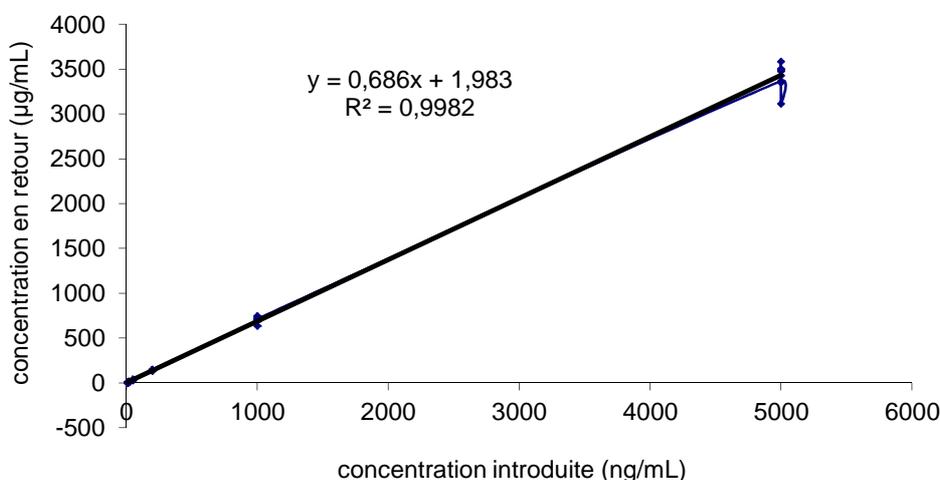


Figure 18 - Droite de régression entre les concentrations introduites et les concentrations en retour.



L'équation de la droite est :

(Concentration en retour) = 0,686 *(Concentration introduite) + 1,983.

Le coefficient de détermination (R^2) est de 0,9982.

Le facteur de correction (F_c) à appliquer est donc de :

$$F_c = \frac{1}{\text{pente}} = \frac{1}{0,686} = 1,457$$

Un nouveau calcul du profil d'exactitude est réalisé en tenant compte du facteur de correction. La correction est effectuée sur les réponses. La figure 19 montre le profil d'exactitude obtenu.

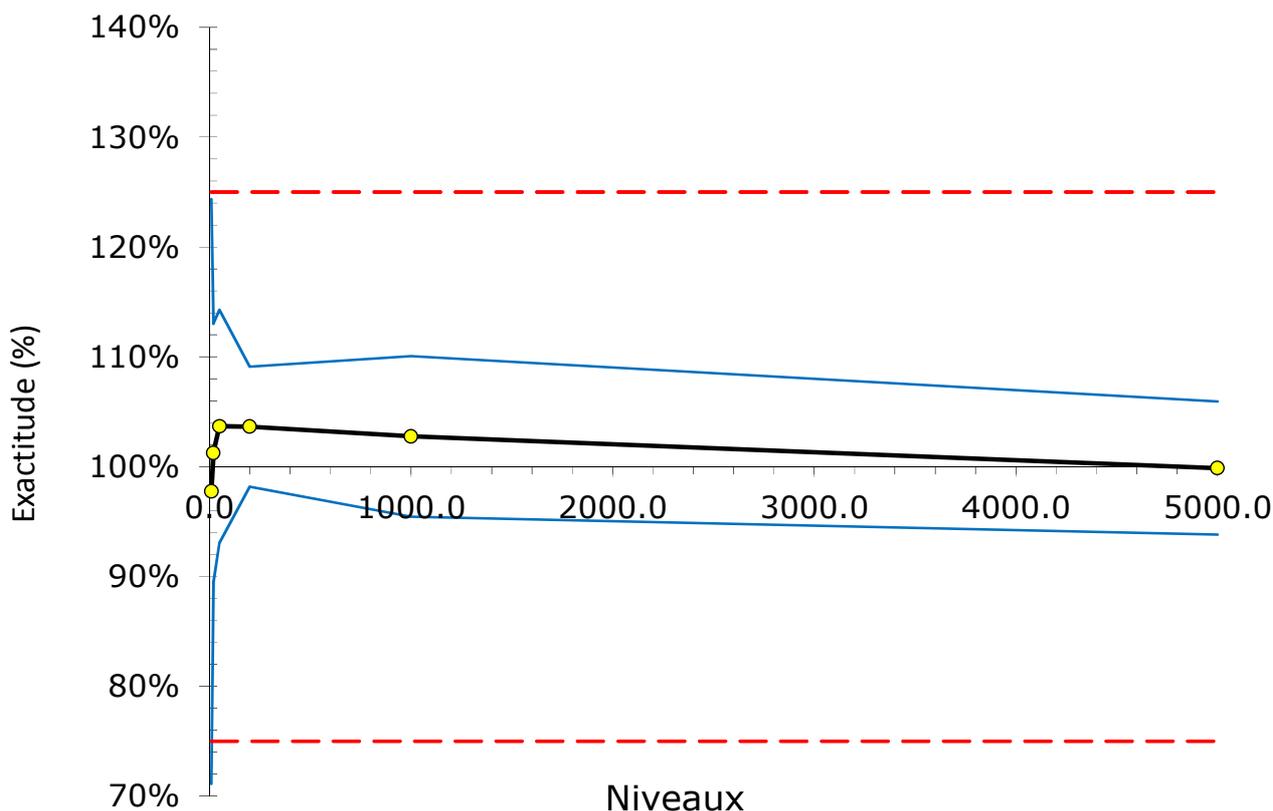


Figure 19 - Profil d'exactitude réalisé avec les données corrigées.

Les résultats sont obtenus avec une régression quadratique pondérée (type 1/X) pour modéliser la fonction de réponse.

La méthode n'est pas validée sur tout le domaine étudié. En effet, la borne inférieure du profil d'exactitude est au-delà de la limite d'acceptation fixée (25%) pour le niveau de concentration égal à 10 ng/mL.

Le calcul de la limite de quantification inférieure et supérieure donne respectivement comme valeur 14,05 ng/mL et 5000 ng/mL. La limite de détection est estimée à 1,02 ng/mL.

Le tableau 9 résume pour l'ensemble des niveaux de concentrations les résultats obtenus.



Tableau 9 - Résultats des calculs du profil d'exactitude après correction des données brutes.

Niveaux	1	2	3	4	5	6
Valeur cible	10	20	50	200	1000	5000
Moyenne niveau	9,95	20,25	51,84	207,28	1027,53	4993,50
Ecart-type de fidélité (sFI)	0,88	1,42	3,32	7,28	47,54	196,84
Coefficient intermédiaire	1,08	1,09	1,08	1,07	1,07	1,08
Nombre de degrés liberté	6,09	4,35	5,05	7,34	6,35	6,26
Valeur basse tolérance	8,59	17,90	46,53	196,34	954,47	4690,31
Valeur haute tolérance	11,32	22,60	57,14	218,22	1100,59	5296,69
Biais absolu	-0,04	0,25	1,84	7,28	27,53	-6,50
Recouvrement (justesse) (%)	99,54	101,25	103,67	103,64	102,75	99,87
Limite basse tolérance (%)	85,86	89,50	93,07	98,17	95,45	93,81
Limite haute tolérance (%)	113,21	113,00	114,28	109,11	110,06	105,93
Limite d'acceptabilité basse	75,00	75,00	75,00	75,00	75,00	75,00
Limite d'acceptabilité haute	125,00	125,00	125,00	125,00	125,00	125,00

La validation montre, sur la base d'un modèle de régression quadratique pondérée, que l'intervalle de dosage est établi entre 14,05 et 5000 ng/mL. Il est également montré que l'effet matrice est systématique et que pour le corriger, il faut appliquer un facteur de correction de 1,457.

Le profil d'exactitude est un outil de diagnostic permettant de situer le ou les points critiques de la méthode. En effet, la méthode peut être biaisée et dans ce cas, le profil d'exactitude permet de valider un facteur de correction. La méthode peut présenter une forte variabilité, ce qui se traduira sur le profil par un intervalle de tolérance plus large que le domaine de validation. La décomposition en répétabilité et fidélité intermédiaire permettra d'identifier la source qui provoque cette variabilité.

L'ensemble de ces calculs peut être réalisé en utilisant une feuille Excel, disponible sur l'espace intranet de la plateforme PAS, ou par des logiciels spécifiques (e noval, ...).



Annexe « Limites »

Les limites de détection (LOD) et de quantification (LOQ) sont définies dans de nombreuses lignes directrices, réglementations ou normes.

Limite de détection

La limite de détection (LOD) est la limite en-dessous de laquelle l'analyte est considéré comme non détecté, avec une certaine probabilité associée. De nombreuses réglementations, lignes directrices ou normes proposent des méthodes d'estimation de la LOD.

Pour les méthodes quantitatives, il se dégage trois approches :

- (i) une approche visuelle : la LOD est déterminée en utilisant des échantillons de concentration connue.
- (ii) une approche à partir du rapport signal sur bruit (S/B) : le rapport S/B est établi en utilisant le signal obtenu sur un échantillon contenant l'analyte recherché et le signal obtenu sur un échantillon ne contenant pas l'analyte (échantillon appelé communément « blanc ». Un rapport S/B compris entre 3/1 et 2/1 est généralement considéré comme acceptable. D'autres approches comme S-B sont également possibles.
- (iii) il est possible de calculer la LOD comme un multiple de l'écart-type de la moyenne des blancs. Généralement, un facteur 2 ou 3 est appliqué selon le niveau de confiance désiré (95 % ou 99 %).
- (iv) $LD = 3 \times S_B$, avec S_B l'écart-type de la moyenne des blancs.
- (v) une approche à partir de la fonction d'étalonnage : cette approche implique d'avoir un modèle linéaire. Le principe de calcul est le suivant : la droite d'étalonnage est prolongée pour trouver l'intersection avec l'axe des Y notée a_0 . La répétition des droites d'étalonnage permet de calculer un écart-type de cette valeur notée S_{a_0} . Il peut être calculé un niveau critique Y_{NC} :

$$Y_{NC} = a_0 + u_{1-\alpha} \times S_{a_0}$$

où $u_{1-\alpha}$ est le quantile d'une loi normale centré réduite. Ce quantile est fixé à 5 %, ce qui donne $u_{1-\alpha} = 1.645$.

Pour obtenir la LOD, il faut introduire le risque β d'avoir des échantillons contenant l'analyte mais qui seraient identifiés comme échantillons blancs. La formule devient :

$$Y_{LD} = a_0 + (u_{1-\alpha} + u_{1-\beta}) \times S_{a_0}$$

où $u_{1-\alpha}$ est le quantile d'une loi normale centré réduite. Ce quantile est fixé à 5 %, ce qui donne $u_{1-\alpha} = 1,645$.

En tenant compte du fait que la LOD doit être exprimée dans la même unité que la concentration, c'est-à-dire prendre en compte la pente de la droite d'étalonnage (sensibilité), la LOD est calculée par la formule :

$$LD = 3,3 \times \frac{S_{a_0}}{a_1}$$

où a_1 est la sensibilité.

Le GT privilégie cette approche. Le facteur 3,3 peut être modifié et cette modification doit être justifiée.



Pour les méthodes qualitatives, le principe général est de procéder soit à la contamination artificielle des échantillons à différents niveaux, soit à des dilutions d'un ou plusieurs échantillons positifs et de les analyser avec la méthode à valider, avec plusieurs répétitions par niveau.

La première approche pour interpréter les résultats consiste à compter le nombre de répétitions positives par niveau sur le nombre total de répétitions pour en déduire, avec le critère de décision fixé, la limite de détection de la méthode.

Un exemple est présenté dans le tableau 10.

Tableau 10 - Nombre de répétitions positives par niveau.

Concentration de l'analyte (unité non précisée et valeurs arbitraires)	Résultats
20	10+/10
15	10+/10
10	8+/10
5	5+/10
2	1+/10

La valeur de :

- la LOD100% sera de 15 (niveau le plus bas pour lequel toutes les répétitions sont positives) ;
- la LOD50% sera de 5 (niveau le plus bas pour lequel 50% des répétitions sont positives).

La 2^{ème} approche pour interpréter les résultats, décrite dans la nouvelle version de la Norme NF EN ISO 16140-2 relative à la validation de méthodes en microbiologie des aliments, est fondée sur la probabilité de détection (POD) modélisée par un modèle linéaire généralisé.

Note : Cette approche appliquée dans la norme aux méthodes microbiologiques est applicable aux autres types de méthodes qualitatives.

Dans la norme NF EN ISO 16140-2, la limite de détection (équivalente au niveau de détection dans la norme), la LOD₅₀ est utilisée, à savoir le niveau de contamination donnant une probabilité de 50 % d'obtenir un résultat d'analyse positif. Cette probabilité de 50 % a été choisie car, se trouvant au point d'inflexion de la courbe sigmoïde de sensibilité d'une méthode qualitative (POD en fonction du niveau de contamination), la LOD₅₀ est estimée avec la meilleure précision à ce niveau.

Plan d'expérience : au moins 3 niveaux de contamination, avec répétitions :

- Témoin négatif : au moins 5 répétitions ;
- Niveau de contamination bas, aux environs de la LOD théorique, donnant des résultats positifs et négatifs lors de répétitions : au moins 20 répétitions ;
- Niveau de contamination élevé : au moins 5 répétitions.



La LOD est calculée à partir des résultats obtenus, en établissant la relation entre les observations, que l'on attend être distribuées selon la loi binomiale, et le nombre de bactéries, connu par inoculation artificielle et distribué selon la loi de Poisson dans les prises d'essai.

La LOD est calculée en appliquant un modèle linéaire généralisé. Le modèle choisi est le modèle log-log complémentaire (LLC).

Note : Le modèle LLC n'est pas seulement applicable aux bactéries, dont la distribution dans les aliments suit une loi log-normale pour les niveaux élevés de contamination.

Modèle log-log complémentaire (LLC)

Soit :

- d le niveau de contamination, exprimé en unités formant colonies (UFC) par unité de masse ou de volume (par exemple, UFC/g) ;
- n_+ est le nombre de répétitions donnant un résultat positif au niveau d ;
- n_{tot} est le nombre total de répétitions au niveau d ;

p est la probabilité d'obtention de résultats positifs :

$$p = E\left(\frac{n_+}{n_{tot}}\right).$$

Par application de la fonction de relation non linéaire suivante, p est transformée en un paramètre η :

$$\eta = \ln(-\ln(1-p))$$

logarithme népérien du niveau de contamination ($\ln d$) :

$$\eta = a_0 + f_i + \ln d$$

où :

- $a_0 = \ln A_0$ (A_0 étant la taille de la prise d'essai, en g ou ml) ;
- $f_i = \ln F_i$ (F_i désignant l'effet matrice de la matrice i).

Les détails du modèle sont décrits dans la publication de Wilrich (2009).

La LOD est estimée comme suit :

$$LOD = \frac{\ln(1-p)}{\hat{F} A_0}$$

avec :

- $p = 0,5$ (correspondant au choix de la LOD_{50}) ;
- \hat{F} obtenu par résolution de l'équation suivante :

$$\sum_{j=1}^q \left(\frac{y_j d_j}{\exp(A_0 \hat{F} d_j) - 1} - (n_j - y_j) d_j \right) = 0$$

où j est la répétition considérée.

Pour le calcul de la LOD, un programme Excel spécifique (Excel « PODLOD_ver6.1) a été développé et est disponible sur le site internet :

<http://www.wiwiss.fu-berlin.de/fachbereich/vwl/iso/ehemalige/wilrich/index.html>



Limite de quantification

Il n'existe pas de définition précise de la limite de quantification. C'est une valeur empirique à partir de laquelle il est possible de quantifier. Il a été proposé que la LOQ soit la plus faible concentration dans un échantillon qui puisse être quantifiée avec une fidélité et une exactitude acceptable dans des conditions expérimentales indiquées.

La LOQ peut être calculée par plusieurs approches :

(i) à partir du rapport signal sur bruit (S/B) : le rapport S/B est établi en utilisant le signal obtenu sur un échantillon contenant l'analyte recherché et le signal obtenu sur un échantillon ne contenant pas l'analyte (échantillon appelé communément « blanc »). Un rapport S/B compris entre 10/1 est généralement considéré comme acceptable. D'autres approches sont possibles comme la différence S-B.

(ii) à partir d'un écart-type, des mesurages répétés sont effectués sur un échantillon dont la teneur en analyte est proche de la LOQ estimée. L'écart-type de ces répétitions, noté S_{bl} , est ensuite calculé. On a alors :

$$LOQ = 10 \times S_{bl}$$

(iii) comme pour la LOD, la LOQ est souvent calculée à partir du rapport signal sur bruit et/ou exprimée comme un multiple d'un écart-type. La LOQ est généralement dérivée de la LOD en appliquant un facteur 2 ou 3 :

$$LOQ = 6,6 \times \frac{S_{a0}}{a_1} \text{ ou } LOQD = 9,9 \times \frac{S_{a0}}{a_1}$$

(iv) à partir du premier point de gamme de la gamme utilisée pour établir le modèle d'étalonnage. Dans ce cas, la LQ n'est pas optimisée mais est acceptable si elle a des performances acceptables.

(v) à partir de l'approche globale basée sur le profil d'exactitude, il est possible de calculer la LOQ de façon à garantir une exactitude, c'est-à-dire des performances (justesse + fidélité) acceptables. Dans ce cas, la LOQ est optimisée et il est possible de définir une LOQ inférieure et une LOQ supérieure.

En dehors des calculs proposés par les diverses réglementations, ou normes, le GT Val 2 recommande de calculer la LOQ par une approche globale (profil d'exactitude).



Annexe « Incertitude »

Pour un mesurande X, l'incertitude est exprimée généralement sous forme d'incertitude élargie absolue accompagnant le résultat rendu. L'incertitude élargie $U(X)$ est dérivée de l'incertitude composée $U_c(X)$ par l'expression :

$$U(X) = k \times U_c(X)$$

avec k le facteur d'élargissement. Ce facteur est fixé à 2 ou 3 selon que 95 % ou 99 % des valeurs possibles soient dans l'intervalle défini par l'incertitude.

Par exemple, si une valeur C a été mesurée à 25 unités arbitraires (u.a.) et que l'estimation de son incertitude composée $U_c(C)$ est de 3.75 u.a., l'incertitude élargie à 95 % sera de :

$$U_c(C) = 2 \times 3,75 = 7,5 \text{ u.a.}$$

Le résultat rendu avec son incertitude sera :

$$C = 25,0 (\pm 7,5) \text{ u. a.}$$

En d'autre terme, 95 % des valeurs de C seront comprises dans l'intervalle [17,5 ; 32,5].

L'incertitude élargie peut-être exprimée sous forme de pourcentage : le GT Val 2 recommande l'utilisation de l'incertitude élargie absolue.

Utilisation de l'incertitude élargie dans la prise de décision

La prise de décision doit se faire sur la base de l'utilisation de l'incertitude élargie. Le graphique 20 montre un exemple d'utilisation de l'incertitude élargie dans la prise de décision.

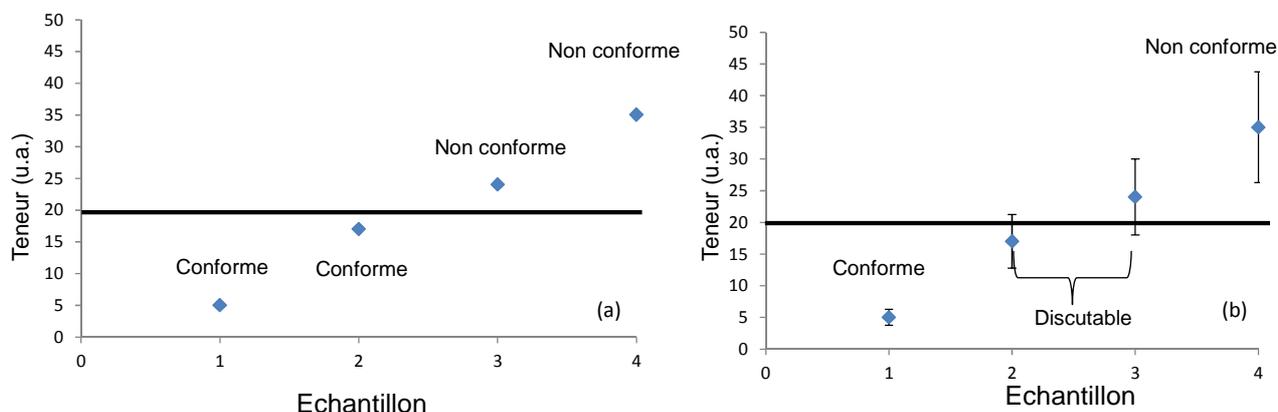


Figure 20 - Importance de l'incertitude dans la prise de décision.

Dans le graphique (a), les résultats sont rendus sans incertitude. Dans le graphique (b), les résultats sont rendus avec l'incertitude élargie U 95%.

Lorsque le résultat est rendu sans incertitude, la prise de décision peut être erronée. Ainsi dans le graphique 20(a), il peut être conclu que l'échantillon 2 est conforme et l'échantillon 3 est non conforme. En réalité, il ne peut pas être conclu à la conformité ou non de ces échantillons si l'incertitude élargie est prise en compte (graphique 20(b)), sauf si des règles de gestion ont été préalablement établies.

Calcul d'incertitude

Basé sur la loi de la propagation de l'erreur et en accord avec les recommandations du GUM, l'incertitude type composée peut s'écrire :

$$u^2(Y) = u^2(\delta) + S_B^2 \sum c_i u^2(X') + S_r^2$$

avec Y, le mesurande, δ représente le biais estimé, S_B^2 est la variance de fidélité intermédiaire, C_i est le coefficient de sensibilité, et S_r^2 est la variance de répétabilité.

Le terme $\sum c_i u^2(X')$ n'est pas pris en compte dans la mesure où les incertitudes de préparations de standard sont négligeables. L'équation peut être simplifiée :

$$u^2(Y) = u^2(\delta) + S_B^2 + S_r^2$$

Or la variance de fidélité intermédiaire et la variance de répétabilité constituent la variance totale S_p^2 :

$$S_p^2 = S_B^2 + S_r^2$$

L'équation s'écrit :

$$u^2(Y) = u^2(\delta) + S_p^2$$

Il a été démontré par Feinberg *et al.* que l'incertitude du biais peut s'estimer par :

$$u^2(\delta) = \frac{S_p^2(1 - \gamma + \gamma/J)}{I}$$

avec J le nombre de répétitions dans les conditions définies, I le nombre de conditions différentes et γ le rapport de la variance de répétabilité sur la variance totale (S_r^2/S_p^2).

En remplaçant $u^2(\delta)$ dans l'équation de $u^2(Y)$, nous pouvons écrire :

$$u^2(Y) = S_p^2 \left[1 + \frac{(1 - \gamma + \gamma/J)}{I} \right] = S_{TI}^2$$

où S_{TI}^2 est la variance de l'intervalle de tolérance attendu, c'est-à-dire la variance observée sur le profil d'exactitude.

Le carré de l'incertitude type composée est égal à la variance totale observée dans le profil d'exactitude, plus la variance du biais.

En reprenant les données de l'étude de validation de l'acrylamide dans le plasma de porc fourni en annexe « Exactitude », le profil d'incertitude a été déterminé. Le tableau 11 résume l'ensemble des valeurs obtenues.



**Tableau 11 - Ensemble des performances obtenues
pour déterminer les teneurs d'acrylamide dans le plasma de porc.**

(cf. annexe « Exactitude »). L'incertitude a été estimée à chaque niveau

Niveaux	1	2	3	4	5	6
Valeur cible	10	20	50	200	1000	5000
Moyenne niveau	9,95	20,25	51,84	207,28	1027,53	4993,50
Ecart-type de fidélité (sFI)	0,88	1,42	3,32	7,28	47,54	196,84
Coefficient intermédiaire	1,08	1,09	1,08	1,07	1,07	1,08
Nombre de degrés liberté	6,09	4,35	5,05	7,34	6,35	6,26
Valeur basse tolérance	8,59	17,90	46,53	196,34	954,47	4690,31
Valeur haute tolérance	11,32	22,60	57,14	218,22	1100,59	5296,69
Biais absolu	-0,04	0,25	1,84	7,28	27,53	-6,50
Recouvrement (justesse) (%)	99,54	101,25	103,67	103,64	102,75	99,87
Limite basse tolérance (%)	85,86	89,50	93,07	98,17	95,45	93,81
Limite haute tolérance (%)	113,21	113,00	114,28	109,11	110,06	105,93
Limite d'acceptabilité basse	75,00	75,00	75,00	75,00	75,00	75,00
Limite d'acceptabilité haute	125,00	125,00	125,00	125,00	125,00	125,00
Incertitude relative (%)	19,03	15,54	14,39	7,76	10,21	8,46
Incertitude basse (%)	81,0	84,5	85,6	92,2	89,8	91,5
Incertitude haute (%)	119,0	115,5	114,4	107,8	110,2	108,5

Approche globale pour le calcul de l'incertitude :

Dans un souci de simplification, il est possible d'utiliser l'incertitude élargie maximale observée sur le domaine de validité et l'appliquer à tous les niveaux d'un même analyte.

À titre d'exemple sur le jeu de données présenté ci-dessus, l'incertitude relative appliquée à tout le domaine serait de 19 %.

Modélisation de l'incertitude

Il est possible de dériver un profil d'incertitude à partir du profil d'exactitude, Gassner *et al* (2014) ont présenté un exemple.

Il est possible d'établir une relation entre les niveaux des concentrations et l'incertitude type estimée à chaque niveau, Cette relation peut prendre différentes formes, sans toutefois expliquer les causes de l'incertitude, Il s'agit d'un modèle empirique qui va permettre de prédire l'incertitude entre deux niveaux étudiés.

Un modèle classique pour décrire la relation entre le niveau et l'incertitude est un modèle de type « fonction puissance » :

$$u(Y) = bY^c$$

avec b et c les coefficients de l'équation estimés par régression linéaire de la fonction puissance.

Cette fonction peut être également appelée fonction d'incertitude, Le tracé de la fonction d'incertitude conduit à obtenir un profil qui est dénommé profil d'incertitude.



À partir des paramètres b et c et de la fonction $u(Y)$, il est possible d'estimer l'incertitude élargie :

$$Y = 2 \times bY^{c-1}$$

L'ensemble de ces calculs peut être réalisé en utilisant une feuille Excel, disponible sur l'espace intranet de la plateforme PAS.

La modélisation de l'incertitude est réalisée par une fonction puissance d'équation :

$$u(Y) = 1,241Y^{-0,02}$$

La modélisation est fournie sur la figure 21.

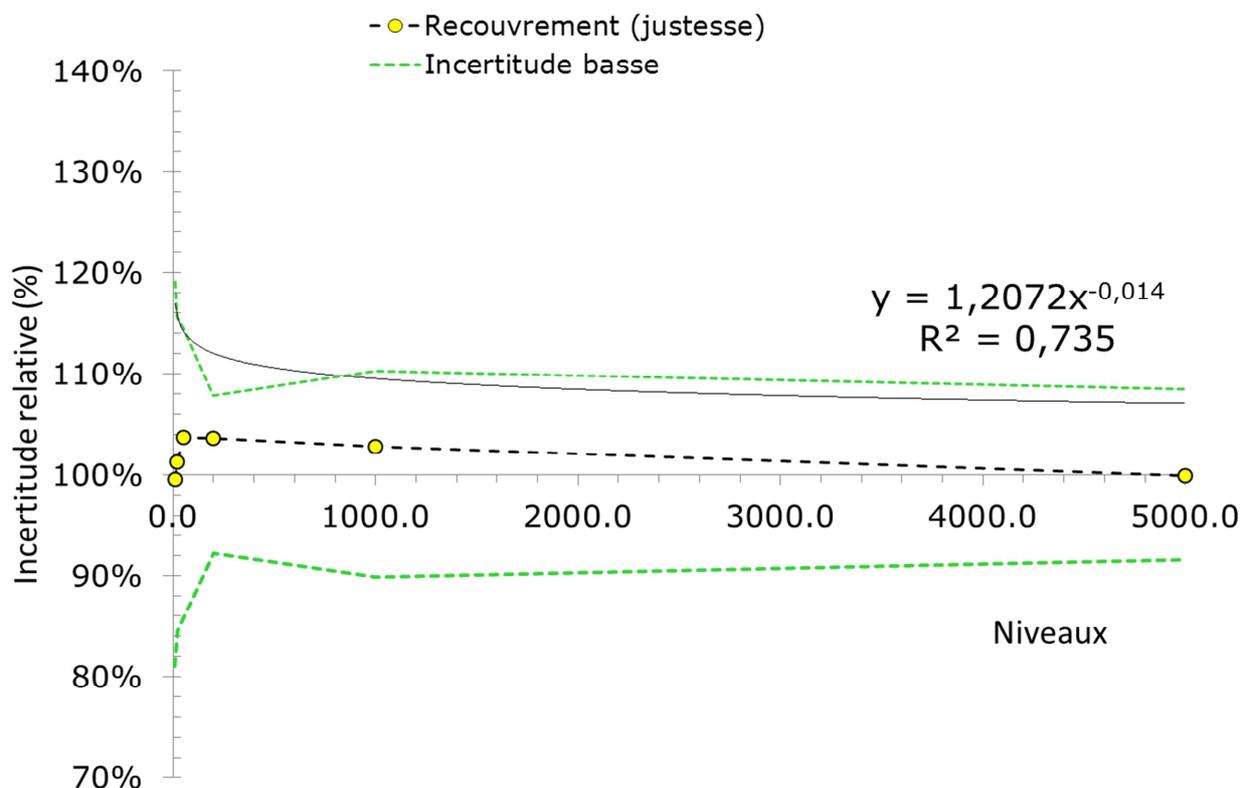


Figure 21 - Modélisations par une fonction puissance du profil d'incertitude.

L'équation de l'incertitude relative nous permet de connaître, en tout point du domaine, l'incertitude relative.

Par exemple l'incertitude pour un niveau de 750 $\mu\text{g/mL}$ sera de :

$$u(Y) = 1,2072(750)^{-0,014} = 1,1003$$

soit 110,03 %, autrement dit 10,03 %, c'est-à-dire comprise entre l'incertitude relative observée à 200 $\mu\text{g/mL}$ et l'incertitude relative observée à 1000 $\mu\text{g/mL}$ respectivement de 7,8 % et 10,2 % (cf, tableau 11).

L'incertitude élargie sera alors de 2X 10,03 % soit 20,06 %.