

Garantir la qualité des mesures et de l'interprétation des données, une exigence scientifique et expertale

Cet article met en avant la nécessité d'une approche rigoureuse dans le cadre des propres expérimentations de l'expert, qu'elles soient ponctuelles pour une expertise ou inscrites dans un processus plus large d'élaboration d'une méthode. Il recommande l'usage de plans d'expériences structurés, adaptés aux objectifs visés, ainsi qu'une analyse statistique des données. Cette approche vise à renforcer la solidité des conclusions et à guider une prise de décision éclairée.



Guillaume Boudarham
Docteur en sciences physiques
Dirigeant d'un laboratoire d'analyses de résidus de tir
Expert judiciaire près la cour d'appel de Versailles dans la spécialité G-18.02 Chimie des résidus de tir

INTRODUCTION

L'analyse de nombreux rapports d'expertise révèle que certaines conclusions formulées par les experts reposent soit sur des expérimentations qu'ils ont eux-mêmes menées pour valider une hypothèse, soit sur des méthodes qu'ils ont développées en interne, ces pratiques étant souvent fondées sur un nombre restreint d'essais réalisés sans véritable rigueur scientifique. Malgré ces lacunes méthodologiques, ces travaux personnels influencent fortement leurs conclusions et potentiellement les décisions judiciaires qui en découlent. Le présent article propose quelques concepts pour améliorer la qualité des propres expérimentations de l'expert et renforcer la solidité des conclusions qui en sont issues. Il s'agit d'un article à visée qualitative : les aspects statistiques ne sont pas abordés et feront l'objet d'un développement ultérieur. L'objectif est avant tout d'encourager les experts – y compris ceux ayant une formation scientifique limitée – à adopter une démarche expérimentale normalisée telle que la mise en œuvre de plans d'expériences et l'analyse statistique des résultats.

VERS UNE DÉMARCHÉ EXPÉRIMENTALE NORMALISÉE

Malgré la pluralité des méthodes scientifiques, la démarche scientifique s'articule en plusieurs grandes étapes : après l'observation d'un phénomène et la définition d'une problématique, des hypothèses sont formulées puis testées par des expériences contrôlées, suivies de la collecte de données et de l'analyse des résultats pour valider ou rejeter les hypothèses. La validation d'une méthode quant à elle est la procédure par laquelle on démontre, preuves expérimentales à l'appui, que les performances de la méthode permettent de répondre aux exigences de l'usage auquel elle est destinée. La réussite d'une expérimentation repose à la fois sur la rigueur méthodologique, la pertinence de la conception des essais, et la qualité des mesures réalisées.

La rigueur méthodologique désigne ici la manière dont l'expérimentation est conduite : respect strict des protocoles, contrôle des biais, recours à des outils fiables et analyse objective des résultats. La pertinence de la conception, quant à elle, renvoie à la qualité de la

réflexion en amont : choix judicieux des variables à mesurer, clarté des objectifs, et structuration du protocole pour répondre efficacement à la problématique. Dans le cadre d'une expérimentation, certaines erreurs peuvent influencer les résultats de manière systématique sans que cela soit immédiatement visible. Il peut s'agir, par exemple, d'un instrument mal réglé, d'une méthode de mesure inadaptée, ou encore d'un choix de données ou d'échantillons qui oriente les conclusions dans un sens particulier. Parfois, c'est aussi l'interprétation elle-même qui est influencée par des attentes ou des hypothèses préalables. Ces dérives, si elles ne sont pas contrôlées, risquent de conduire à des conclusions erronées. De nombreux mécanismes dépendent de multiples facteurs.

L'étude de la dépendance d'une variable de sortie d'un tel mécanisme soulève plusieurs questions [1] : quels sont les facteurs qui exercent le plus d'influence sur la réponse ? Des interactions ou corrélations existent-elles entre certains facteurs ? Est-il possible de réduire le nombre de mesures tout en conservant un maximum d'informations

utiles ? Les résultats obtenus sont-ils affectés par des erreurs systématiques ? Par exemple, il est légitime de s'interroger sur les facteurs influençant le vieillissement d'une encre et sur l'élaboration d'une méthode permettant de dater une encre sur un document. C'est dans cette logique que s'inscrivent les plans d'expériences, qui constituent un outil fondamental de structuration des essais permettant d'organiser et d'optimiser les expérimentations.

Pour que les données issues de ces plans soient interprétables, il est essentiel d'assurer leur qualité métrologique. C'est là qu'intervient la métrologie, qui fournit les outils et méthodes pour garantir la validité des résultats. Elle s'appuie sur plusieurs concepts, quelque peu abstraits, définis dans le Vocabulaire international de métrologie (VIM), qui sont repris ci-après dans leur version originale¹ :

- La métrologie est la science des mesures et ses applications. La métrologie comprend tous les aspects théoriques et pratiques des mesurages, quels que soient l'incertitude de mesure et le domaine d'application.
- Une grandeur est une propriété d'un phénomène, d'un corps ou d'une substance, que l'on peut exprimer quantitativement sous forme d'un nombre et d'une référence.
- Un mesurande est la grandeur que l'on veut mesurer.
- Une valeur de référence est la valeur d'une grandeur servant de base de comparaison pour les valeurs de grandeurs de même nature. La valeur de référence peut être une valeur vraie d'un mesurande, et est alors inconnue, ou une valeur conventionnelle, et est alors connue.
- La valeur conventionnelle d'une grandeur est la valeur attribuée à une grandeur par un accord pour un usage donné.
- Un mesurage est un processus consistant à obtenir expérimentalement une ou plusieurs valeurs que l'on peut raisonnablement attribuer à une grandeur.
- Un résultat de mesure est l'ensemble de valeurs attribuées à un mesurande, complété par toute autre information pertinente disponible.
- La valeur mesurée est la valeur d'une grandeur représentant un résultat de mesure.
- La valeur vraie est la valeur d'une grandeur compatible avec la définition de la grandeur.
Note : la valeur vraie d'une grandeur correspond à la valeur qu'elle aurait dans des conditions idéales, sans aucune erreur de mesure. En pratique, cette valeur est inatteignable, car toute mesure comporte une incertitude. On cherche donc à s'en approcher au mieux, ce qui conduit à utiliser une valeur de référence : une estimation reconnue comme suffisamment proche de la valeur vraie pour être utilisée comme base de comparaison.
- Un instrument de mesure est un dispositif utilisé pour faire des mesurages, seul ou associé à un ou plusieurs dispositifs annexes.
- Un système de mesure est un ensemble d'un ou plusieurs instruments de mesure et souvent d'autres dispositifs, comprenant si nécessaire réactifs et alimentations, assemblés et adaptés pour fournir des informations destinées à obtenir des valeurs mesurées dans des intervalles spécifiés pour des grandeurs de natures spécifiées.
- Une indication est la valeur fournie par un instrument de mesure ou un système de mesure.
- L'exactitude de mesure est l'étroitesse de l'accord entre une valeur mesurée et une valeur vraie d'un mesurande.
- La justesse de mesure est l'étroitesse de l'accord entre la moyenne d'un nombre infini de valeurs mesurées répétées et une valeur de référence.
Note : en pratique, on ne peut pas mesurer la justesse de mesure au sens strict de la définition du VIM (moyenne d'un nombre infini de mesures), mais on peut l'estimer à partir d'un nombre fini et raisonnable de mesures répétées, en suivant une démarche normalisée définie dans l'ISO 5725.
- La fidélité de mesure est l'étroitesse de l'accord entre les indications ou les valeurs mesurées obtenues par des mesurages répétés du même objet ou d'objets similaires dans des conditions spécifiées.
- Une condition de répétabilité est une condition de mesurage dans un ensemble de conditions qui comprennent la même procédure de mesure, les mêmes opérateurs, le même système de mesure, les mêmes conditions de fonctionnement et le même lieu, ainsi que des mesurages répétés sur le même objet ou des objets similaires pendant une courte période.
- Une condition de reproductibilité est une condition de mesurage dans un ensemble de conditions qui comprennent des lieux, des opérateurs et des systèmes de mesure différents, ainsi que des mesurages répétés sur le même objet ou des objets similaires.
- La répétabilité de mesure est la fidélité de mesure selon un ensemble de conditions de répétabilité.
- La reproductibilité de mesure est la fidélité de mesure selon un ensemble de conditions de reproductibilité.
- Une erreur de mesure est la différence entre la valeur mesurée d'une grandeur et une valeur de référence.
- Une erreur aléatoire est la composante de l'erreur de mesure qui, dans des mesurages répétés, varie de façon imprévisible.
- Une erreur systématique est la composante de l'erreur de mesure qui, dans des mesurages répétés, demeure constante ou varie de façon prévisible.
- Une incertitude de mesure est un paramètre non négatif qui caractérise la dispersion des valeurs attribuées à un mesurande, à partir des informations utilisées.
- Une vérification est la fourniture de preuves tangibles qu'une entité donnée satisfait à des exigences spécifiées.

La fidélité se rapporte à la constance des mesures d'une même grandeur dans des conditions similaires. Elle est liée aux erreurs aléatoires, qui correspondent aux variations imprévisibles pouvant survenir d'une mesure à l'autre. Par exemple, lors de la mesure du poids de détente d'une arme à feu, des variations peuvent survenir d'une mesure à l'autre en raison d'erreurs aléatoires.

Cependant, il sera important de vérifier que ces fluctuations restent dans des limites acceptables en utilisant des indicateurs de dispersion tels que l'écart-type, la variance ou le coefficient de variation dans les conditions spécifiées. La justesse, quant à elle, mesure l'écart entre les résultats obtenus et la valeur attendue, et elle est influencée par les erreurs systématiques, qui sont des écarts

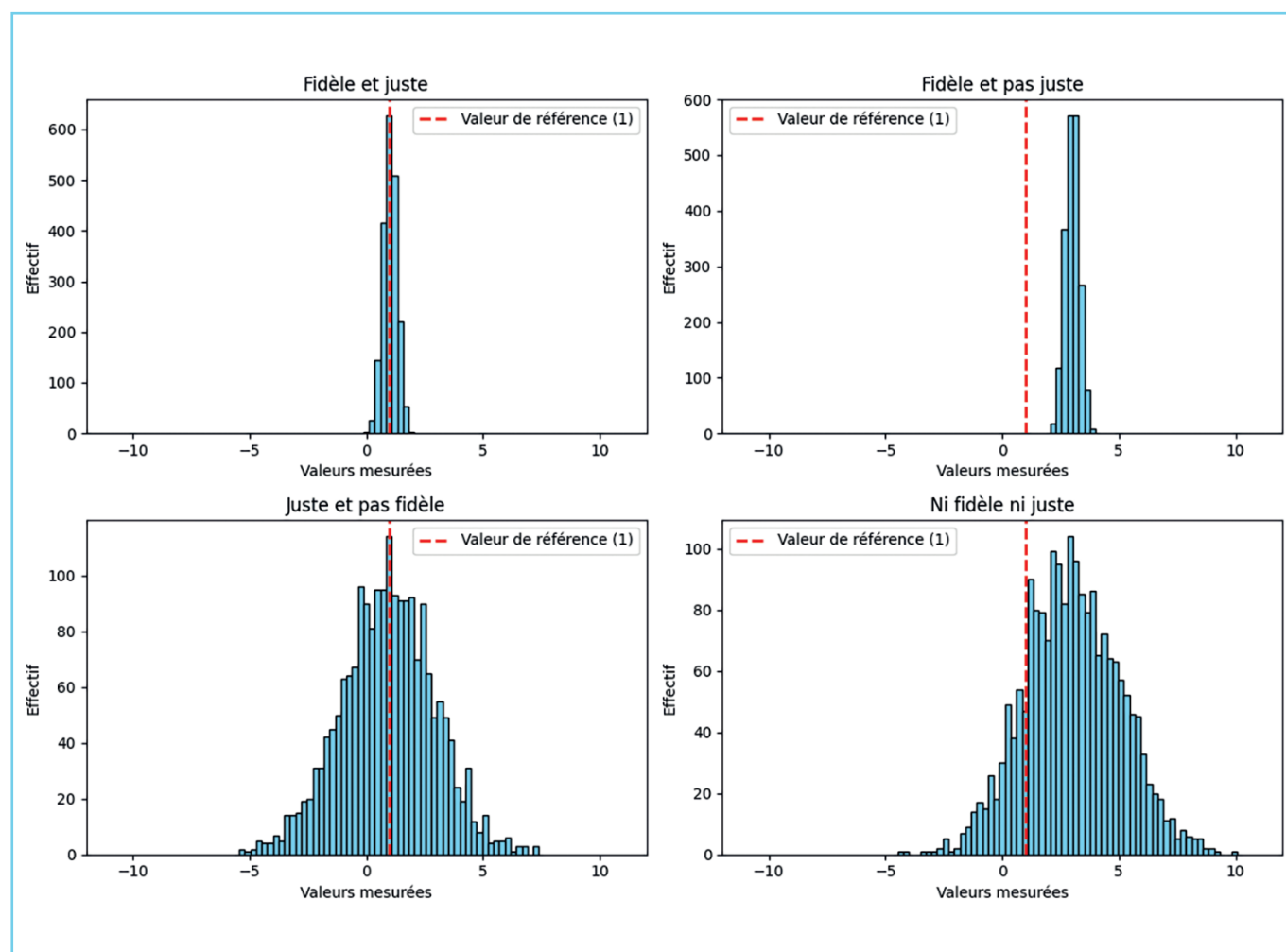
constants pouvant être causés par un instrument mal réglé ou une mauvaise utilisation de celui-ci. Par exemple, si une balance affiche systématiquement une masse inférieure à la valeur réelle, les mesures seront faussées, manquant de justesse. La répétabilité quantifie la constance des mesures lorsque la même expérience est réalisée à plusieurs reprises dans des conditions identiques, avec le même opérateur, au même endroit et en utilisant le même instrument pendant une courte période. Par exemple, lorsqu'un opérateur effectue plusieurs essais de dureté Vickers sur le canon d'un fusil de chasse, des variations significatives des résultats indiquent que les mesures ne sont pas répétables. Cela pourrait se produire, par exemple, si les surfaces mesurées ne sont pas parfaitement planes et polies. Enfin, la reproductibilité permet de vérifier si les résultats peuvent être reproduits ailleurs avec d'autres opérateurs et des équipements différents. Par exemple, le dosage des métabolites

de la cocaïne dans le sang doit donner des résultats similaires lorsqu'il est réalisé dans différents laboratoires, les écarts observés devant être inférieurs à un seuil spécifié. Qu'il s'agisse de répétabilité ou de reproductibilité, les variations entre les résultats de mesure doivent être suffisamment faibles pour rester dans des limites acceptables, garantissant ainsi la qualité des résultats.

Les histogrammes ci-dessous offrent une représentation visuelle des notions de justesse et de fidélité. Ces graphiques sont composés de barres correspondant chacune à une classe, c'est-à-dire un intervalle de valeurs mesurées. La hauteur de chaque barre montre combien de valeurs se trouvent dans cet intervalle. Plus la barre est haute, plus il y a de valeurs dans cette plage. La ligne rouge verticale indique la valeur vraie (choisie ici égale à 1). Dans le premier cas (« Fidèle et juste »), les valeurs sont centrées sur la valeur vraie, avec une faible dispersion. Dans le deuxième cas (« Fidèle et pas juste »),

les valeurs sont centrées sur une valeur éloignée de la valeur vraie, mais avec peu de dispersion. Dans le troisième cas (« Juste et pas fidèle »), les valeurs sont centrées autour de la valeur vraie, mais présentent une forte dispersion. Enfin, dans le quatrième cas (« Ni fidèle ni juste »), les valeurs sont centrées autour d'une valeur éloignée de la valeur vraie, avec une forte dispersion.

Sans entrer dans les détails, il est important de souligner qu'il existe plusieurs outils statistiques reposant sur des tests d'hypothèse et des critères de décision permettant, par exemple, d'évaluer si la moyenne d'un ensemble de mesures diffère de manière significative d'une valeur de référence ou si l'écart observé peut être attribué au hasard, de comparer les moyennes de deux séries de mesures, de détecter des valeurs aberrantes, de définir les limites de répétabilité et de reproductibilité (c'est-à-dire l'écart maximal admissible entre deux mesures). Examinées sans l'utilisation de ces outils statistiques,



Histogrammes offrant une représentation visuelle des notions de justesse et de fidélité.

les données peuvent être interprétées de manière subjective, entraînant des conclusions erronées ou orientées. Prenons l'exemple d'un expert qui effectue trente mesures du poids de détente d'une arme à feu, qu'il souhaite comparer à une valeur de référence fixée à 1,60 kg. La moyenne des mesures est de 1,55 kg, avec un écart-type de 0,10 kg. La question est alors de savoir si cette moyenne est conforme à la valeur de référence. À première vue, l'écart semble minime, ce qui incite à penser que la moyenne observée est conforme à la valeur de référence. Cependant, une telle interprétation superficielle ne tient pas compte de la variabilité des données et ne prend pas en considération le niveau de confiance que l'on peut accorder à la conclusion. Un test statistique, en l'occurrence le test de Student, permet d'analyser les résultats de façon plus rigoureuse. En fixant le risque d'erreur à 5%, il établit que la moyenne observée s'écarte significativement de la valeur de référence, ce qui conduit au rejet de l'hypothèse nulle de conformité.

Le document de référence intitulé « Évaluation des données de mesure - Guide pour l'expression de l'incertitude de mesure » [2] a pour objectif de normaliser la façon d'exprimer l'incertitude liée à une mesure. Dès l'introduction, il est précisé que « *Lorsqu'on rend compte du résultat d'un mesurage d'une grandeur physique, il faut obligatoirement donner une indication quantitative sur la qualité du résultat pour que ceux qui l'utiliseront puissent estimer sa fiabilité. En l'absence d'une telle indication, les résultats de mesure ne peuvent pas être comparés, soit entre eux, soit par rapport à ces valeurs de référence données dans une spécification ou une norme. Aussi est-il nécessaire qu'il existe une procédure facilement applicable, aisément compréhensible et largement acceptée pour caractériser la qualité du résultat d'un mesurage, c'est-à-dire pour évaluer et exprimer son incertitude.* »

CONCLUSION

La fidélité, la justesse, la répétabilité et la reproductibilité de mesure sont des concepts importants qui permettent

d'évaluer la performance d'une méthode. La fidélité fait référence à la constance des mesures, tandis que la justesse désigne la proximité des mesures par rapport à une valeur de référence. Les histogrammes présentés dans cet article permettent d'apprécier visuellement ces deux concepts. Par ailleurs, la répétabilité et la reproductibilité sont également des critères importants à considérer. La répétabilité garantit la constance des mesures dans des conditions identiques tandis que la reproductibilité assure que les résultats sont cohérents dans des conditions différentes. Des outils statistiques reposant sur des tests d'hypothèse et des règles de décision permettent, par exemple, d'évaluer si la moyenne d'une série de mesures diffère significativement d'une valeur de référence ou si l'écart observé peut être attribué au hasard, de comparer les moyennes de deux séries de mesures, de détecter des valeurs aberrantes, de définir les limites de répétabilité et de reproductibilité. Le lecteur désireux d'enrichir sa compréhension de cet article ou de mettre en application les concepts abordés pourra consulter en ligne les documents de référence du Bureau International des Poids et Mesures (BIPM), à savoir : le « Vocabulaire international de métrologie - Concepts fondamentaux et généraux et termes associés » (VIM) ainsi que le « Guide pour l'évaluation des données de mesure et l'expression de l'incertitude de mesure » (GUM). Il est également important de rappeler que, dans le cadre du développement d'une méthode, la publication de ses résultats permet un examen critique par ses pairs, validant ainsi la rigueur de l'approche adoptée et garantissant la transparence du processus. Cela ouvre la voie à des retours constructifs, susceptibles d'affiner la méthode et d'en renforcer la qualité. En outre, rendre ces résultats publics favorise le partage des connaissances au sein de la communauté scientifique, participant activement à l'avancement du domaine et à la reconnaissance du travail accompli. Nous déconseillons fortement d'utiliser des méthodes ou des résultats internes qui n'ont pas fait l'objet d'une validation scientifique rigoureuse.

Des logiciels libres et gratuits comme R sont efficaces pour analyser les résultats expérimentaux en traitant de grandes quantités de données. L'intelligence artificielle (IA), comme ChatGPT basé actuellement sur le modèle GPT-4o d'OpenAI, peut également aider à l'interprétation des données en soutenant le travail de l'expert. Elle permet de détecter des anomalies, de proposer des analyses pertinentes et d'élaborer des plans d'expérience structurés.

Cependant, malgré ces avancées technologiques, les compétences de l'expert restent indispensables. L'IA ne peut pas se substituer à l'expertise humaine, qui est essentielle pour comprendre pleinement le contexte, interpréter correctement les résultats et prendre des décisions éclairées. Comme ailleurs, l'IA doit donc être vue comme un outil complémentaire, et non comme un substitut aux compétences de l'expert.

L'auteur remercie chaleureusement le docteur Michael Stock, directeur du département de métrologie physique au Bureau International des Poids et Mesures (BIPM), pour sa relecture attentive de cet article ainsi que pour ses conseils avisés qui ont grandement contribué à l'amélioration de celui-ci.

NOTE

1. Pour les définitions des termes non expliqués ici, le lecteur est invité à se référer au document de référence disponible en ligne [3].

BIBLIOGRAPHIE

- [1] Wikipédia, l'encyclopédie libre, *Plan d'expériences*, 12 mai 2024, <https://fr.wikipedia.org/wiki/Plan_d%27exp%C3%A9riences>, page consultée le 11 avril 2025.
- [2] Bureau International des Poids et Mesures (BIPM), *Évaluation des données de mesure - Guide pour l'expression de l'incertitude de mesure (GUM)*, JCGM 100:2008(F), <<https://www.bipm.org/fr/committees/jc/jcgm/publications>>.
- [3] Bureau International des Poids et Mesures (BIPM), *Vocabulaire international de métrologie. Concepts fondamentaux et généraux et termes associés (VIM)*, 3^e édition, version 2008 avec corrections mineures, JCGM 200:2012, <https://www.bipm.org/documents/20126/2071204/JCGM_200_2012.pdf>.