

# Introduction au procédé d'échantillonnage

M. Deleuil

Rhône-Poulenc-Rorer, 13, Quai Jules Guesde, BP. 114, 94403 Vitry sur Seine, France

People is more and more affected by heterogeneous materials, with complex spatial distribution (biology, ecology, quality for use). Qualitative as well as quantitative measurement efficacy depends on both the relevance of the measured dimension, and the representativeness of the analysed fraction.

Sampling and sample adaptation to on-line analyser, or to laboratory analysis, concerning application testing or sample storage, must agree with correction rules, in order to reduce bias risk, to decrease non-systematic errors, and to allow total error estimation.

Every case needs a particular treatment. Critical factors are the actual objective, the heterogeneity of both constitution and distribution scales, the lot size, and the spatial disposition of the lot to be handled.

Particulate materials are discussed and compared with molecular materials, to enlighten some fundamental notions such as:

- Sampling by number or by weight. Liberation parameter.
- Weight to be sampled ( $M_E$ ). Increments to be collected ( $N$ ).
- Sampling design. Static or dynamic lot presentation.

Analytical result conversion into lot estimation. The eminent multidisciplinary character of the sampling process is highlighted. The Analyst has to certify correct methods for the entire process and to assume responsibility. Sampling is easier to realize when early designed, during the engineering conception.

The sampling practice is nowadays impossible to be appreciated as entirely satisfactory. Improvements are necessary, not only with gases and natural materials, but specially with particulate materials, concerning theoretical bases as well as reglementation and apparatus performance.

## L'échantillonnage dans le processus d'acquisition de l'information

La victoire appartient à celui qui est le mieux informé. Les domaines militaire, bancaire ou médical en livrent chaque

jour la preuve. La révolution informatique y puise sa justification.

La chimie, en tant qu'activité de transformation de la matière, n'échappe pas à cette règle. Les informations dont elle a besoin sont de nature *chimique*, mais aussi biochimique, physicochimique et physique, que nous groupons ici sous l'appellation *physique*.

Les informations chimiques sont *qualitatives* (identification des espèces présentes), ou *quantitatives* (dosage d'une concentration). Elles relèvent alors de l'échelle moléculaire (ou ionique), et utilisent assez librement des opérations intermédiaires, comme l'extraction, le broyage ou la dissolution.

Les informations physiques portent sur les *phases* en présence (polymorphisme), sur les *interfaces* (catalyse hétérogène, parois), et sur *l'élaboration finale* (transformation du produit chimique en *produit d'usage*) : poudre à laver, matière plastique, peinture, engrais, additif nutritionnel, médicament. Les informations relèvent alors de dimensions texturales, micro, méso ou macroscopiques, en milieux polyphasiques formulés, qui sont difficiles à échantillonner et à caractériser. De plus, l'adaptation à l'analyse est souvent délicate, une dissolution ou un broyage étant généralement prohibés.

Tous les milieux sont concernés : les liquides homogènes, les liquides hétérogènes, les gaz, les solides divisés et les matériaux. Limiter la métrologie à l'analyse chimique de produits chimiques desservirait la notion de service rendu, et appauvrirait la compétitivité de cette industrie. Que les analyses soient chimiques ou physiques, en ligne ou au laboratoire, qu'elles visent à des mises au point ou à des contrôles, par des mesures ou par des tests, elles exigent simultanément (Fig. 1) :

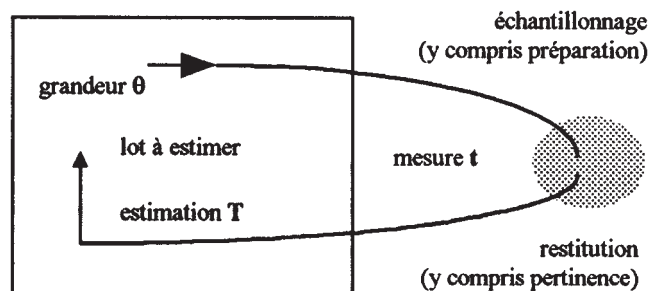


Figure 1. Estimation = pertinence + échantillonnage + mesure.

- *Pertinence* : le caractère mesuré doit être significatif vis-à-vis du but recherché (valeur d'usage) ;
- *Représentativité* : la grandeur mesurée (concentration, surface, porosité, humidité) doit être, dans la prise d'essai, la plus semblable possible à ce qu'elle est, en moyenne, dans le lot ;
- *Justesse et précision* : l'erreur totale doit être acceptable.

Le lot est l'ensemble dont on veut connaître la grandeur  $\theta$ . Il est délimité de façon précise, dans l'espace, dans le temps, ou par des nombres. Il peut être le *contenu* d'un réacteur, le rejet d'une cheminée pendant 24 h, cinquante fûts, ou encore le sous-sol d'une région. L'estimation porte sur tout le lot, rien que le lot. Si l'on pouvait analyser le lot entier et sans traitement, les erreurs de pertinence et d'analyse disparaîtraient. Mais celles de la représentativité disparaîtraient.

La plupart du temps, on doit se résoudre à analyser une faible partie du lot. Cette prise d'essai doit être traitée, pour qu'elle convienne à l'analyse, en quantité et en qualité. Il en résulte une série d'opérations, de prélèvements et de traitements, qui font passer du lot à l'analyse, et qui constituent le procédé d'échantillonnage :  $\theta$  est *simulé* par  $T$ , lui-même *estimé* par le résultat  $t$  (Fig. 1).

Les grandeurs d'estimation sont diverses : concentration, surface, densité, ou bien couleur, goût, dureté, vitesse de dissolution... Mais chaque chaîne d'acquisition est spécifique par sa pertinence, ses prélèvements, ses préparations, sa mesure et sa restitution. La mesure peut porter sur  $\theta$ , la *grandeur* même à estimer : dosage de l'eau pour connaître l'humidité du lot. La pertinence n'est alors pas problématique.

Mais elle peut porter sur une variable  $T$  de *substitution* à  $\theta$  : une granulométrie peut vouloir indiquer une coulabilité. Un dosage d'eau peut s'interpréter en risque de mottage, ou une concentration en antioxydant en terme de stabilité. La pertinence peut alors être critique.

La relation entre  $\theta$  et  $T$  sort du cadre de l'échantillonnage. Mais l'analyste qui s'apprête à mettre au point les prélèvements et la mesure, doit savoir, impérativement, en quoi elle consiste.

Le résultat analytique  $t$  sert à estimer la teneur  $T$  du lot.

L'erreur relative sur  $T$  est  $e_R = \frac{|T-t|}{T}$  ( $t$  = moyenne des mesures)

L'analyse est dite :

- *Exacte*, si  $e_R < \varepsilon$  (erreurs systématique et aléatoire négligeables) ;
- *Juste*, si  $|T - t_m| < \varepsilon$ , (erreur systématique négligeable) ;
- *Fidèle*, si  $s_t^2 < \varepsilon$ , avec  $s_t^2$  = variance de l'erreur totale (erreur aléatoire négligeable) ;
- *Biaisée*, si  $|T - t_m| > B_0$ , avec  $B_0$  = erreur systématique tolérée ;

– *Irreproductible*, si  $s_t^2 > S_0^2$ , avec  $S_0^2$  = erreur aléatoire tolérée ;

– *Représentative*, si  $r^2 < B_0^2 + S_0^2$  (degré suffisant pour la somme justesse et fidélité). Ceci revient à dire que tous les éléments du lot ont la même chance de se trouver dans l'échantillon (probabilisme).

La situation de  $T$  par rapport à  $t$  peut être diverse.

Si  $t_i$  et  $t_s$  sont les limites inférieure et supérieure de l'*intervalle de confiance*, cinq cas principaux peuvent se présenter :

- *Pertinence inconnue* :  $T$  ne préjuge pas de  $\theta$ , la seule *bonne* information. La proximité de  $T$  avec  $t$  n'a pas d'importance ;
- *Prélèvement non maîtrisé* :  $T = \theta$ , mais l'erreur sur  $T$  est inconnue. On ne peut donc estimer  $T$  à partir de  $t$  ;
- *Biais* :  $T = \theta$ , et  $T$  est estimé être dans l'intervalle de confiance, mais les valeurs de  $t$ ,  $t_i$  et  $t_s$ , sont fausses ;
- *Irreproductibilité* :  $T = \theta$ , et  $T$  est dans l'intervalle de confiance. Cela est juste, mais peu précis, car l'intervalle de confiance est grand ;
- *Représentativité* :  $T = \theta$ , et  $T$  estimable de façon juste et précise.

Une information de qualité relève donc de trois démarches :

- ① S'assurer de la signification de la grandeur réellement mesurée ;
- ② Éviter un biais important ;
- ③ Minimiser les erreurs aléatoires.

Nous aborderons ici les points ② et ③.

## Les erreurs systématiques

Les erreurs systématiques de mesure se jugent en comparant à un étalon, et (ou) en confrontant plusieurs types de mesure. Les erreurs systématiques de prélèvement relèvent en principe de la même méthodologie, mais il est fort rare que l'on confectionne des lots étalon ou des lots d'étude, pour tester les prélèvements, ne serait-ce que parce que chaque matière est spécifique.

Il en résulte une situation particulière : la seule garantie contre le biais est le respect du principe de correction.

Si les prélèvements suivent la théorie, s'ils bénéficient de modes opératoires et de matériel correct, ces prélèvements sont des *échantillons*, car on assume alors qu'ils sont sans biais. Si les prélèvements s'écartent de l'un des critères de correction, ou bien si certaines conditions sont inconnues, ce sont des *specimen*, ce qui signale que le résultat  $t$  ne peut être utilisé pour estimer  $T$ .

Les principes de correction sont :

- La relation (chimique ou physique) entre  $t$  et  $T$  est avérée, et  $T \approx \theta$  ;
- les prélèvements sont *probabilistes* : les éléments du lot ont tous la même chance d'être prélevés, et les éléments hors du lot ont tous une chance nulle d'être prélevés (pas de contamination) ;
- les traitements conservent le probabilisme et l'intégrité du produit vis-à-vis de  $T \approx \theta$ . Par exemple, l'humidité est conservée si l'on veut doser l'eau, apprécier une coulabilité ou une comprimabilité, etc. ;
- la mesure et son étalonnage donnent bien  $t$ , quantitativement.

La pertinence et l'échantillonnage posent peu de problèmes de biais avec les solutions (dosage chimique, milieu homogène). Ils sont au contraire critiques dans l'estimation de la fonctionnalité d'un produit, ou dans la caractérisation d'un milieu hétérogène.

Pourtant, les efforts prodigués pour atteindre l'exactitude sont souvent plus grands au niveau de la mesure et sur les milieux homogènes. D'où la pensée que l'échantillonnage est négligé, ou mal enseigné, pour des raisons plus ou moins mystérieuses [1].

La cause de cette situation nous semble de nature technique. Dans l'analyse chimique des solutions, l'arsenal des validations vérifie la justesse et la précision de la mesure qui est le seul problème. Dans nombre d'analyses physiques d'un milieu complexe, tout compte, du choix de la mesure à la représentativité, mais :

- les méthodes et les appareils dont on dispose sont souvent insuffisants [2-5]. Par exemple, les sondes biaisent, les préleveurs s'obstruent...
- les recommandations officielles peuvent être ambiguës [2]. Les lois publiées sont empiriques. Elles donnent parfois des résultats éloignés (comparer [1] et [6]), ne serait-ce que pour des problèmes d'unité...
- Une vaste pluridisciplinarité, alliant la connaissance des milieux complexes, l'application des produits, les procédés, les échantillonnages les analyses chimique et physique, les technologies et les règlements est nécessaire, mais très difficile à mettre en place.

Il n'y a pas négligence, mais pénurie.

### Les erreurs aléatoires

Le biais assumé par le principe de correction reste inconnu, alors que les erreurs aléatoires sont quantifiées par le nombre de répétitions. Avec une distribution normale des mesures,  $T - t$  à 95 % de chances de tomber dans l'intervalle de confiance  $t_s - t_i$  de la moyenne :

$$T = t \pm k \cdot s \quad t_s = t + k \cdot s \quad t_i = t - k \cdot s \quad \frac{RSD^1}{100} = \frac{s}{t} = \frac{\sigma}{T}$$

<sup>1</sup>RSD/100 = relative standard deviation = écart-type relatif.

Ceci signifie que :

- l'on admet l'absence de biais ;
- $t$  est la moyenne de  $N$  mesures,  $s$  est l'écart-type de la moyenne ;
- Une  $N + 1$ ème mesure aurait 95 % de chances d'être dans l'intervalle de confiance  $t \pm 1,96 \cdot s$  avec  $N > 100$  ;
- Si  $N \rightarrow \infty$  (analyse de tout le lot !), alors  $s \rightarrow 0$  ;  $t_s - t_i \rightarrow 0$  ;  $t \rightarrow T$  ;
- $N$  étant limité, la valeur réelle  $T$  du lot reste inconnue. Elle est seulement estimée par  $t \pm k \cdot s$  ;
- $k$  est le paramètre de Student, lu au niveau  $N - 1$  et probabilité  $P(0,05)$ . Par exemple (2), pour juger de l'homogénéité d'un mélange, la FDA demande que l'écart-type relatif ( $s/t$ ) soit donné pour  $N = 10$  (mélange intermédiaire) ou  $N = 30$  (mélange final) ;
- La variance ( $s_t^2$ ) est la somme des variances : les erreurs aléatoires s'ajoutent, pour donner l'erreur totale :

$$s_t^2 = s_E^2 + s_A^2$$

$s_E = \text{échantillonnage} = \text{prélèvements} + \text{traitements}$

$s_A = \text{analyse} = \text{mesure} (+ \text{pertinence})$ .

L'erreur sur la pertinence n'est pas de nature aléatoire. Du moins, les auteurs négligent-ils cet aspect, comme nous le ferons ici.

L'erreur aléatoire de la mesure est connue. Son écart-type relatif  $s_A/t$  est souvent faible :  $\sim 2\%$ , et parfois moins de 1 %. L'erreur aléatoire d'échantillonnage peut varier fortement :  $s_E/t$  peut être  $< 1\%$  pour les milieux homogènes, mais dépasser 20 voire 50 % [3] pour les milieux ségrégués.

La performance de la mesure en matière d'écart-type relatif n'a d'intérêt que si  $s_A/t$  et  $s_E/t$  sont du même ordre de grandeur.

Pour illustrer cette affirmation, considérons le cas  $t = 100$  ;  $s_A = 2$  ;  $s_E = 5$  ;  $N = 16$  ;  $k = 2,13$  ( $N - 1 = 15$ , probabilité 0,05). On obtient  $t_i = 5,39$  ;  $T = 100 \pm 16,9$ . Ce RSD de 5,39 provient pour 4,65 de l'échantillonnage, et pour 0,74 de l'analyse : réduire  $s_A$  est totalement inutile.

Il est difficile d'amener les erreurs de pertinence, de prélèvement, de traitement et de mesure, toutes au même niveau. Chacune a sa nature (test ou mesure, unités, importance relative), chacune est pratiquée par des métiers différents, chacune s'attache à un objectif particulier (compétitivité, qualité, enregistrement).

### L'hétérogénéité de la matière

L'analyse porte sur une infime fraction du lot, que celle-ci soit prélevée ou observée autour d'un capteur. Il faut donc

composer avec la *divisibilité* du lot, ou faire circuler celui-ci *parfaitement agité* devant le capteur, pour conserver la correction du probabilisme.

Quand le but est chimique, et sur un liquide homogène, les entités agissantes sont indéniablement les molécules. L'homogénéité est atteinte dès la prise en compte de quelques milliers d'entre elles, c'est-à-dire  $10^{-20}$  mole. Un tel lot est divisible à une micro échelle.

Des ordures ménagères, des poudres pharmaceutiques, ou des préparations alimentaires, n'ont pas la divisibilité d'un liquide. Le niveau d'hétérogénéité est multiple. Le cas du dosage des aliments vitaminés illustre tout à fait ce problème [7] :

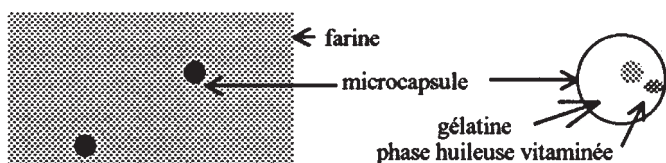


Figure 2. Aliment enrichi de microcapsules vitaminées.

Une farine est additionnée de capsules de taille 5–500  $\mu\text{m}$ , pour posséder une activité vitaminique de  $10^4$  UI/kg (Fig. 2). Le critère officiel d'homogénéité consiste à pratiquer deux déterminations sur le même échantillon de 30 g, et à vérifier que les résultats ne diffèrent pas de plus de 20 %. Les grains ont en moyenne une activité de 10 UI/grain, ce qui donne une moyenne de 30 grains dans 30 g. Mais un grain de 400  $\mu\text{m}$  a une activité 1 000 fois supérieure à un grain de 40  $\mu\text{m}$ , car la phase huileuse vitaminique est proportionnelle au poids des capsules.

La farine étant très pauvre en capsules, la distribution de Poisson s'applique (événements rares). L'écart-type relatif est trouvé égal à 18,4 %. Comme deux mesures peuvent avoir jusqu'à quatre écart-types de distance, elles peuvent différer de  $4 \times 18,4 = 73,6$  %, du seul fait d'une prise d'essai limitée à 30 g. La définition du mélange est donc inopérante, du seul fait de la quantité prélevée. Il faut substituer un échantillon à la prise d'essai.

Il est impossible d'aborder ici tous les cas. Il nous semble cependant inévitable de développer le cas des solides divisés, tant il est fréquent en finition des produits chimiques (cristallisation, filtration, séchage, broyage, mélange, conditionnement, etc.).

Un lot de 1  $\text{m}^3$  de particules de masse volumique apparente 730  $\text{kg}/\text{m}^3$ , de diamètre moyen en poids 100  $\mu\text{m}$ , de densité 1,4 a une équivalence de  $10^{12}$  particules sphériques. Mais les particules réelles sont toutes différentes entre elles, par leur poids, leur teneur en impureté, leur forme, leur

taille. Le lot comporte plus de  $10^{16}$  individualités, non interchangeables.

Aucun fractionnement ne peut tout à fait représenter le lot. Nous touchons là à l'hétérogénéité de *constitution*. Dans les lots moléculaires, l'hétérogénéité est due à la non-interchangeabilité des molécules de nature différente (éthanol  $\neq$  eau). Mais les molécules d'éthanol sont identiques entre elles, et interchangeables (ce qui donne un sens à la notion de concentration).

Comment opérer sur des particules individuelles, pour respecter le probabilisme, en donnant à chaque particule la même chance d'être prélevée ?

Une façon d'opérer consiste à définir dix familles de particules en fonction de leur poids, les particules d'une même famille étant supposées interchangeables.

La loi de Bernouilli est alors applicable :

$$n = \frac{k^2}{(\sigma/T)^2} \left( \frac{1}{T} - 1 \right)$$

- $n$  = nombre de particules à prélever pour estimer la teneur en nombre  $T$  d'une famille.
- $k = 1,96$  quand  $n$  est grand. Le niveau de probabilité est 0,05.
- $\sigma^2$  = variance de  $T$  (assimilable à  $s_E^2$ , variance de l'erreur aléatoire de constitution faite sur  $t$ , car  $n$  est grand).

La taille  $n$  du prélèvement ne dépend pas de la sensibilité de l'analyse, mais seulement du niveau imposé au *RSD* ( $100\sigma/T$ ). Trop d'analystes ont l'habitude de demander « une faible prise d'essai, vue la sensibilité de leur mesure ». L'assertion est acceptable pour les milieux moléculaires, mais incorrecte en général, car elle ne tient pas compte de la représentativité. Dans notre exemple :

$$T = 0,1 \text{ (10 \%)} ; \sigma/T = 0,06 \text{ (RSD = 6)} ; \\ n = 9\,600 \text{ particules.}$$

Quatre remarques immédiates sont à faire :

- *RSD* correspond à l'erreur aléatoire maximale, alors que la loi de Bernouilli ne donne que l'erreur due à l'hétérogénéité de la matière. Le prélèvement de 9 600 particules suppose que la prise, les traitements et la mesure se font *sans erreur*. L'hypothèse peut s'avérer acceptable quand  $T$  est faible ( $T < 0,01$ ), mais souvent insuffisante pour  $T > 0,1$ .
- La loi est statistique, en *nombres*, alors que la plupart des échantillons sont en *poids*. Des particules de diamètre respectif 10, 100 et 200  $\mu\text{m}$  pèsent  $7,3 \cdot 10^{-13}$ ,  $7,3 \cdot 10^{-10}$ , et  $5,8 \cdot 10^{-9}$  kg.

Comment passer des nombres aux poids ? À nombre égal, les grosses contribuent plus au poids. Elles servent donc de base des nombres, car elles demandent plus de précision numérique :  $\rho (\pi/6) d_{95}^3$ . Le diamètre  $d_{95}$  est de l'ordre du diamètre donnant 95 % de passant.



Soit ici  $d_{95} = 200 \mu\text{m}$ .  $M_E = n \cdot \rho (\pi/6)d_{95}^3 = 56 \text{ mg}$ .

$M_E$  est la masse d'échantillon primaire (Fig. 3).

En prélevant 56 mg on assure la correction du modèle vis-à-vis des particules  $d_{95}$ . Les particules moyennes et fines sont connues avec une précision supérieure.

- Si l'on prélève non plus 56 mais 100 mg,  $n$  passe à 17 200 et  $\sigma/T$  est réduit à 0,045. Ainsi, la demande  $RSD < 6$  est mieux assurée, et laisse un peu de place à quelques erreurs aléatoires pratiques. La loi de Bernouilli s'exploite des deux façons : si  $RSD$  est imposé, on calcule  $M_E$ . Si  $M_E$  a été pratiqué à une valeur fixée par d'autres critères, on calcule le  $RSD$  qui en résulte.
- La métrologie ne génère parfois que de faibles erreurs ( $RSD = 3$  pour  $T = 0,01$ ). Recherchons le même niveau avec l'échantillonnage : la loi de Bernouilli donne  $n = 4,2 \cdot 10^6$ ,  $M_E = 24,5 \text{ g}$ , ce qui est déjà contraignant dans certaines industries. Mais si l'échelle d'hétérogénéité passe de 200 à 2000  $\mu\text{m}$ , le poids à prélever atteint  $M_E = 24,5 \text{ kg}$  ! La précision sur les faibles teneurs, et, surtout, sur les hétérogénéités macroscopiques (en mm), se paye très cher en poids d'échantillon à prélever.

Considérons un lot de microcapsules de vitamines, prêtes à être mélangées à la farine (Fig. 2). La loi de Bernouilli ne s'applique plus. Ce ne sont plus des grains vitaminés, d'autres non, mais des grains tous vitaminés, plus ou moins. La teneur  $T$  n'est plus *libérée* au niveau de la taille des grains, mais au niveau des microgouttes huileuses internes. L'hétérogénéité de constitution ne coïncide plus avec l'échelle granulaire.

Basée sur le modèle des minerais, une solution [1] consiste à introduire le paramètre  $l$  de libération ( $0 < l < 1$ ), rapport des diamètres goutte/capsule. De façon générale, il faut utiliser une loi de distribution quantitative. C'est la solution adoptée par Bourgeois dans le cas des farines vitaminées [7]. Sans développer ici ces modèles, on pressent qu'ils appellent à connaître mieux la composition locale et les agencements. Les études de stabilité, de compréhension de la qualité, de reproductibilité et d'échantillonnage vont de pair. Un modèle donnant satisfaction est porteur d'une grandeur plus pertinente et d'une signification physique sur la texture.

Les machines de division en petit conditionnement (sachets, gélules, comprimés, doseur, etc.) fonctionnent en échantillonnage. Elles sont à la fois source et destination des progrès de ce procédé. L'information contenue dans les fiches de contrôles peut être valorisée par les variogrammes [8]. Le rôle limitant, mais aussi révélateur, de l'échantillonnage y est souligné. Connaissant la constitution du lot, à l'échelle des grains et à l'échelle fine de la libération, on ne peut toujours pas bâtir le plan d'échantillonnage. Il faut aussi en connaître la *distribution*.

La position des grains dans le lot peut être aléatoire (lot *homogène* à l'échelle macroscopique), ou au contraire déterminée par un critère, qui peut être leur poids, leur rugosité, leur humidité, leur charge électrostatique (lot *ségréatif*).

Les capsules de vitamine les plus grosses peuvent se rassembler à la périphérie des trémies, tandis que les capsules les plus fines peuvent se concentrer au milieu...

Sans ségrégation, le probabilisme est respecté en prélevant un nombre  $n$  suffisant de grains, n'importe où, comme dans un liquide homogène.  $M_E$  est pris en un, deux, ou trois prélèvements, par exemple en bas, au milieu et en haut d'un fût. Mais avec un lot ségréatif, il faut *parcourir* le lot. Quel nombre  $N$  de prélèvements doit-on opérer, pour obtenir  $M_E$  ? Il tombe sous le sens que cela dépend de la taille du lot (un train n'est pas un fût), et du degré de ségrégation (une poudre free-flowing n'est pas une poudre cohésive).

Les statistiques nous enseignent que l'intervalle de confiance est inversement proportionnel à la racine carrée du nombre de mesures. À valeur égale de l'écart-type, l'intervalle de confiance diminue d'un facteur 2 si l'on prélève non pas 4 mais 10 incréments, d'un nouveau facteur 2 entre 10 et 30 incréments, et d'un autre facteur 2 entre 30 et 120 incréments (ce que les appareils automatiques réalisent couramment). Les prélèvements manuels sont fastidieux. Il est important de distinguer les cas simples ( $N = 1$  à 5) des cas critiques ( $N = 30$ ), sauf si les prélèvements sont automatiques ( $N = 30$  à 300). La façon d'organiser  $N$  compte tout autant que la valeur  $N$ .

Considérons un lot de 730 kg,  $M_E = 30 \text{ g}$ ,  $N = 30$ . Il faudrait découper, par la pensée, le lot en 730 000 fractions, tirer au sort 30 d'entre elles, et les prélever ! Les appareils de prélèvement sont incapables de découper la fraction n° 62933 sans prendre un peu de 62932, et sans abandonner un peu de 62933. Ceci est inévitable, mais peut-être acceptable, si aucun biais n'est introduit. Il serait par contre grave de ne pas avoir accès à l'ensemble du lot (les  $n^{\text{os}} > 600.000$  seraient par exemple exclus). Les recoins des wagons, les vis de prélèvement en position fixe, les bras trop courts des manipulateurs, etc. induisent malheureusement cette erreur.

La meilleure réponse est le prélèvement en *lot dynamique*. Ce n'est plus l'espace qui est découpé, mais le temps, pendant la vidange ou le remplissage. Le flux *total* est pris pendant quelques secondes, un grand nombre de fois. On divise ensuite l'échantillon obtenu par 10 ou par 100, pour obtenir les 30 g souhaités (Fig. 3).

Les procédés continus sont plus favorables que les discontinus. Ils fonctionnent en régime *stationnaire* (les régimes transitoires étant éminemment irréproductibles). Ils délivrent plus simplement une information qui est de meilleure qualité. De façon contradictoire, ils restent rares dans des industries où la notion de lot est pourtant importante, en pharmacie par exemple...

## La pratique de l'échantillonnage

L'échantillonnage est un *procédé*. Il en possède tous les grands caractères :

- enchaînement de plusieurs opérations (prélèvements, mélange, division, dissolution, etc.) ;
- appel aux lois de transformation de la matière (extraire, conditionner, sans dénaturer) ;
- maîtrise des flux (ne pas contaminer, ne pas perdre, ne pas dégrader) ;
- base technologique (type de lots, préleveurs, diviseurs, appareillage pour la préparation à l'analyse) ;
- utilisation de modèles physiques, recours aux méthodes statistiques de traitement des données, validation des modes opératoires (importance de la chimie et de la physique).

Cet ensemble se concrétise par **le plan** d'échantillonnage.

Chaque plan est spécifique, mais le cas suivant est assez démonstratif :

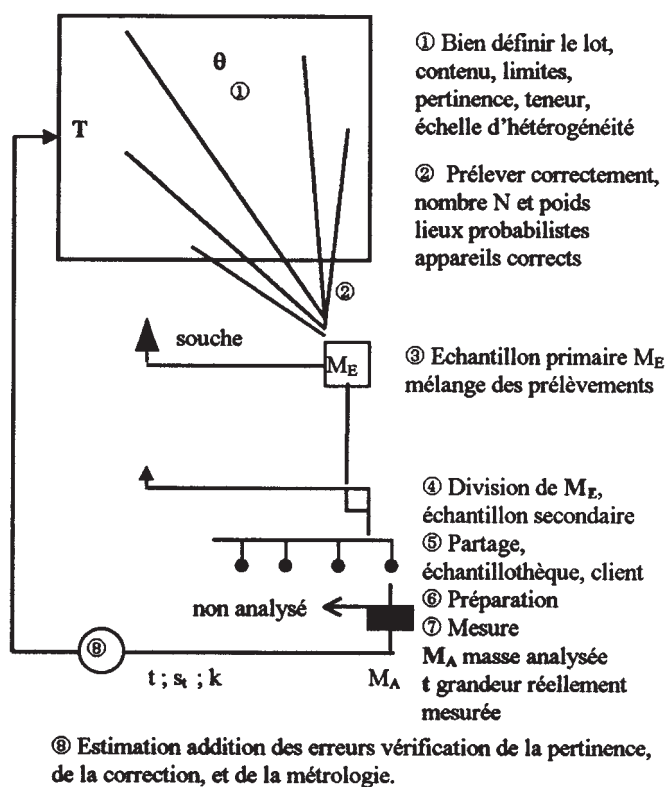


Figure 3. Le plan d'échantillonnage.

Les facteurs monodisciplinaires rendent le plan incorrect. Nous citerons quatre attitudes erronées fréquentes :

① Absence de plan. L'analyste demande ce dont il a besoin (*sic*), c'est-à-dire ⑤, et adapte le *specimen* ⑥ à son analyse, sans réfléchir à adapter l'analyse au problème. Il

restitue  $t$  et  $s_A^2$ , avec les précautions d'usage : prudence sur l'échantillonnage (*sic*).

② Incorrection des appareils. Par exemple, des sondes à logettes sont utilisées [2] pour vérifier l'homogénéité d'un mélange de poudres pour lequel les Autorités imposent  $RSD < 6$ . Les sondes provoquent un biais de plus de 6 %. Les sondes se révèlent être incorrectes [3-5]. Mais les autres solutions sont tout aussi suspectes : tantôt incorrectes (lot non parcouru), tantôt inaptes à la faisabilité (voûtes, pertes, poussières, mauvais nettoyage). Les constructeurs font des efforts louables, mais qui demandent à être poursuivis.

③ Incorrection de l'état d'esprit technique. Découragés d'apprendre qu'aucun appareil n'est tout à fait satisfaisant, certains n'attachent plus d'attention à ce sujet douteux. D'autres se plaignent de l'isolement dans lequel l'opération d'échantillonnage est plongée.

Il faut examiner les problèmes au cas par cas, sans pessimisme ni légèreté. Dans l'industrie chimique, de bonnes solutions sont souvent possibles, si la chaîne du procédé d'information est pensée en termes de procédé, et non en opérations de spécialistes. La solution dépasse le fabricant, le contrôle, l'assurance qualité, l'ingénieur, le chimiste, l'analyste, le spécialiste de l'échantillonnage, et le vendeur d'appareils.

④ Incorrection de l'état d'esprit gestionnaire. On veut que le mieux informé gagne, mais sans nourrir la compétence nécessaire pour assurer la pertinence, le plan, la mesure, et leur exécution. La formation est parfois négligée, en particulier sur les points clé ① (prélever) et ⑥ (préparer).

La connotation de l'échantillonnage et de la métrologie doit être information, adaptation fondée et rapide, souplesse, rigueur, et qualité, et non pas erreur, temps perdu, nombri-lisme de spécialistes. Le pire coût est celui de la non qualité.

Les facteurs pluridisciplinaires sont le gage de la correction du plan, et de l'efficacité du renseignement. Nous en citerons quatre :

- Échantillonnage et métrologie sont du *contrôle*. Il faut connaître le lot et l'objectif pour mesurer pertinemment, exactement, et pour tirer profit de l'information. En développement, pendant qu'on apprivoise le produit, sa fonction, son procédé, durant l'*analyse de compréhension*, la pertinence prime. Les difficultés sont majeures : que représente le spécimen actuel ? Que faut-il mesurer ? Que faut-il simplifier pour rendre le contrôle jouable ? L'échantillonnage n'est pas opérationnel. L'ordre de grandeur prévaut sur le détail.
- Échantillonnage et métrologie sont l'affaire de tous, mais ils sont, en définitive, sous la responsabilité de l'*analyste*. Le cahier des charges est celui de l'analyste : représentativité, justesse, précision, robustesse, cadences, coût. Une part importante du plan est réalisée par l'analyste lui-même : division, partage, préparation, mesure, et, surtout, restitution du résultat. Du point de vue réglementaire, le

dossier d'échantillonnage entre dans la part analytique du dossier d'enregistrement.

- Ce dernier aspect n'est pas le moins incitatif. Les BPL-BPF, qui se contentaient, il y a peu de temps encore, de souhaiter que l'échantillon soit représentatif, demandent maintenant des explications scientifiques, et s'intéressent aux opérations intermédiaires. L'échantillonnage et la métrologie font partie de la compétitivité (s'ils sont d'un niveau insuffisant, le dossier est refusé).
- La chimie se préoccupe de plus en plus de produits formulés, de matériaux, et de milieux écologiques. Les liquides homogènes de la pétrochimie diminuent en importance relative, au profit des colloïdes, des pâteux, des polymères chargés, des formules pharmaceutiques, alimentaires, ou phytosanitaires. Les milieux biologiques, naturels, ou spatiaux (ne va-t-on pas échantillonner Mars !) prennent la première place. Par un juste retour des choses, la connaissance de l'échelle d'hétérogénéité est utile à la compréhension des propriétés du produit. Une taille de grains, une épaisseur de pellicule, un ordre d'introduction, peuvent devoir leur amélioration à l'échantillonnage.

Les publications portant sur la pratique [9-15] et sur l'enseignement de l'échantillonnage [16] sont peu nombreuses. Nous voyons là beaucoup moins un instinct de conservation du savoir-faire qu'un manque de savoir tout court. Dans certains cas (flux continu de poudres non ségrégative, en cimenterie par exemple), des chaînes entièrement automatiques de prélèvement, de préparation, et d'analyse par RX sont au point (Polsap, Pfaff...). Une analyse combinant RX et laser sur les particules fait son apparition dans l'industrie pharmaceutique, pour contrôler l'état d'avancement des mélanges (Camp).

Hormis ces systèmes clé en main, c'est généralement une solution particulière qu'il faut mettre en place. Si la pluridisciplinarité n'a pas parfaitement fonctionné, la solution est médiocre, sinon inapplicable, surtout si l'unité existe déjà, avec ses contraintes. On pratique alors sur spécimen...

C'est à la *conception du procédé* qu'il faut penser et prévoir l'échantillonnage :

- Au niveau de l'*analyse*. La cascade ③④⑤⑥ demande du temps et de la technicité, tout en multipliant le risque de biais et les occasions d'erreurs aléatoires, ceci d'autant plus que l'écart entre  $M_E$  et  $M_A$  est grand, et que la présentation de  $M_A$  est éloignée de l'état natif (exemple de la microscopie électronique). À niveau de pertinence égal, il faut toujours sélectionner l'analyse la plus compatible avec ③.
- Au niveau du *plan*. Analyser les prélèvements séparément, ou rassemblés, ne donne ni la même information, ni le même travail, ni le même temps de réponse. Noter que les prélèvements sont rassemblés pour l'échantillonnage, dont la mission est importante.
- Au niveau *technologique*. Il est difficile, coûteux, et risqué, d'opérer une implantation d'échantillonnage sur une

unité existante. Il est par contre simple et efficace de le faire sur le papier (conception), puis au moment du montage (réalisation). C'est souvent le manque de concertation analyste-ingénierie qui laisse passer ce moment privilégié.

Deux autres aspects sont, eux aussi, en pleine évolution :

- Exploitation des résultats : réseaux neuronaux.
- Gestion de l'échantillonnage. Échanges entre laboratoires agréés.

## Conclusion

La pertinence de la grandeur mesurée et l'échantillonnage donnent sa justification à la mesure (analyse ou test).

La retombée immédiate de l'échantillonnage consiste à éviter les erreurs grossières, et à réduire les imprécisions. Mais l'échantillonnage permet aussi de préciser la texture locale des mélanges (constitution) et les variations spatiales de cette texture (distribution). Des transferts de matière comme la diffusion, la ségrégation, ou la percolation peuvent être élucidés par la qualité des prélèvements et de leur interprétation.

De nombreuses technologies fonctionnent comme des échantillonneurs, et sont, de ce fait, redevables de leurs progrès : machine à comprimer, conditionnement en parts individuelles, etc.

L'échantillonnage demande une coopération poussée entre l'analyste, l'Assurance Qualité, le Bureau d'études, le concepteur du matériel, le fabricant, et les Instances officielles. Cette activité bénéficie des avantages de la pluridisciplinarité (compréhension mutuelle, solution réaliste, robuste, scientifique), mais en possède aussi les difficultés : le savoir et l'appareillage ne sont pas, aujourd'hui, à un niveau équivalent à celui de l'analyse même.

Il est prévisible que les aspects réglementaires deviendront plus exigeants au fur et à mesure que les moyens scientifiques et technologiques progresseront.

Certaines notions devraient être approfondies, et faire partie de la pensée quotidienne de toutes les personnes concernées : hétérogénéité de la matière, masse à prélever, procédé de prélèvement et de préparation, échantillonnage, aptitudes et inaptitudes des lots selon leur conformation (statique ou dynamique), cahier des charges du matériel (non contamination des milieux, probabilisme, intégrité du prélèvement, faisabilité des transferts).

Une collaboration Université/Concepteur/Utilisateur/Affaires Réglementaires semble être des plus utiles, pour le progrès des produits, des procédés, ainsi que pour la formation des personnels.

## Références

1. Gy, P. M. *Analisis* **1983**, *11* (9), 413-440.
2. Carstensen, J. T. *Drug Dev. Ind. Pharm.* **1996**, *22* (4), 285-290.
3. Muzzio, F. J. *Int. J. Pharm.* **1997**, *155*, 153-178.
4. Berman, J. *Drug Dev. Ind. Pharm.* **1996**, *22* (11), 1121-1132.
5. Garcia, T. P. *Pharm. Dev. Tech.* **1998**, *3* (1), 7-12.
6. Allen, T. Particle Size Measurement ; Elsevier: London, 1997. (Fifth ed). Vol. I, Chap. I.
7. Bourgeois, C. F. *Analisis* **1984**, *12*, (2), 84-89.
8. Gy, P. M. Sampling of particulate materials, Elsevier: Amsterdam, 1982.
9. Berman, J. *Drug Dev. Ind. Pharm.* **1995**, *21* (11), 1257-1283.
10. Delacroix, F. *Mesures* **1998**, *703* (mars), 77-79.
11. Le Blanc, D. A. *Pharm. Tech.* **1998**, may, 66-74.
12. Leschonski, K. Comparison of the errors of different instruments for particle size analysis, - Partec, Nürnberg.
13. Harris, C. C. *Powder Technol.* **1975**, *12*, 85-91.
14. Egermann, H. *Eu. J. Biopharm.* **1991**, *37* (4), 216-218.
15. Schofield, C. *Powder Tech.* **1976**, *15*, 169-180.
16. Leschonski, K. Probennahme und Probenteilung, Clausthal University post-graduate course, Germany.