

**AVIS D'EXPERTS | Du déterminisme au datamining****La quête du "Graal"  
se tiendra dans l'hyperespace***Et les métrologues auront enfin leur mot à dire...*

Depuis son émergence, le but de l'industrie a été de produire (reproduire) des objets de façon à les vendre. La création de valeur passe dès lors par les clients. Ils doivent avoir envie d'acheter le produit proposé (compétitivité hors coût), au juste prix (compétitivité coût) et en être satisfait (qualité). Comment y parvenir ?

Le plus souvent, chaque produit, une fois imaginé, doit être dessiné (plan projet), prototypé, validé<sup>(1)</sup> puis industrialisé. Au final, nous obtenons un ensemble de "recettes" qui fixe les conditions d'obtention du produit final. Il s'agit formellement de l'ensemble des plans et spécifications qui permettent à la production de réaliser sa mission : produire et surtout, produire conforme. En effet, toute "non-conformité", à l'usine comme chez le client, est génératrice de coûts et d'insatisfactions, ce qui pénalise le produit, voire l'entreprise qui peut, si elle ne sait pas se corriger, finir par disparaître...

Dans ce contexte, le « Graal » se présente comme LA recette qui permet l'obtention du produit dans les meilleures conditions de coûts, de délais et de qualité. Nous datons à 1790 la première révolution industrielle, issue de l'utilisation du charbon comme source d'énergie. La deuxième révolution, autour de 1850,



**Jean-Michel POU,**  
Président fondateur de la société Delta Mu, membre des commissions "Métrologie" et "Méthodes statistiques" de l'Afnor et président du cluster d'excellence "Auvergne Efficience Industrielle"...

naît avec l'électricité. Depuis, l'activité industrielle n'a pas cessé de se développer et de progresser, passant des ateliers de quelques dizaines de personnes à des firmes mondiales de plusieurs milliers de salariés. Avec l'ouverture des frontières, le terrain de la compétition est aujourd'hui planétaire et la course à la compétitivité devient plus que jamais vitale. La 4<sup>e</sup> révolution industrielle<sup>(2)</sup>, tel qu'elle commence à se dessiner, s'appuiera sur le numérique : interconnexions en temps réel de tous les acteurs de la *supply chain*, capa-

cité de stockage massif de données (Big data), et augmentation considérable des puissances de calcul.

Si l'énergie, la technique et les outils de production ont considérablement évolué depuis 1790, nous pouvons constater (regretter) que le mode de pensée qui conditionne nos recettes n'a, quant à lui, pas encore vraiment évolué. Or, il existe un important potentiel de compétitivité à développer en remettant en cause nos recettes d'antan. Nous venons d'un monde déterministe dans lequel deux et deux font quatre (exemple des chaînes de côtes) alors que la réalité est plus imprévisible. Les théories statistiques qui s'attachent, pour certaines, à décrire le caractère probabiliste des choses<sup>(3)</sup>, ont connu un essor important à partir du

Malgré des propriétés intéressantes, l'industrie est restée dans la pratique la plus courante à l'idée qu'il fallait que tous les bouchons soient plus gros que tous les stylos pour qu'ils fonctionnent alors qu'en réalité, il faut qu'un bouchon particulier soit plus gros que le stylo particulier qu'il devra fermer.



XIX<sup>e</sup> siècle, mais elles ont du mal à pénétrer le quotidien du monde industriel, même si la M.S.P (maîtrise statistique des procédés) et le Six Sigma ont connu certains succès. Bertrand Russell<sup>(4)</sup> résume assez bien, à mon sens, la difficulté quasi existentielle d'accepter le monde probabiliste en écrivant « *Ce que les hommes veulent, en fait, ce n'est pas la connaissance, c'est la certitude.* » La Statistique, et le monde probabiliste qui lui est consubstantiel ne proposent finalement que des doutes ... Aussi, et malgré des propriétés intéressantes, l'industrie est restée dans la pratique la plus courante à l'idée qu'il fallait que tous les bouchons soient plus gros que tous les stylos pour qu'ils fonctionnent alors qu'en réalité, il faut qu'un bouchon particulier soit plus gros que le stylo particulier qu'il devra fermer. Les conséquences de cette pensée sont coûteuses, très coûteuses...

« Si l'inférence statistique est née du besoin de quantifier une population totale sur la base de quelques échantillons, pour contourner la question des coûts importants liés à la mesure, la technologie permet souvent aujourd'hui de sortir de ce contexte. »

Dans la compétition mondiale actuelle, il devient pourtant indispensable de changer, non pas par plaisir, mais plus prosaïquement pour survivre. L'utilisation des nouveaux outils et des nouvelles technologies finira par imposer cette évolution ...

#### Notre histoire : le déterminisme

En 1836, l'État crée le service des Poids et mesures dont la mission était alors de garantir la loyauté des mesures.

Ce service existe encore aujourd'hui et c'est grâce à son action que nous ne doutons pas des quantités échangées lors des transactions commerciales. Cela ne signifie pas pour autant que les quantités mesurées sont justes, car et contrairement à un sentiment inconscient, les mesures ne peuvent pas être justes.

En baignant dans ce monde des mesures apparemment justes depuis si longtemps, nos recettes industrielles en ont les stigmates.

En effet, et de nos jours encore, une valeur mesurée dans une tolérance permet de déclarer la conformité d'un objet. La valeur mesurée n'étant pas la valeur vraie, les tolérances ne représentent pas les valeurs vraies admissibles pour l'objet. Nos recettes, mises au point dans un monde de valeurs mesurées, sont par conséquent des recettes pour valeurs

mesurées, non pour des valeurs vraies ! Ces pratiques conduisent à des tolérances nettement plus serrées que le besoin réel, ceci malgré un besoin impérieux de compétitivité. Dans un monde où les ressources semblaient infinies et la compétition industrielle équitable (en termes de coûts de production), il était possible d'imaginer d'en rester là. Ce n'est plus possible de nos jours et il est plus que temps de réviser nos

croyances pour progresser et retrouver la compétitivité...

#### Le monde probabiliste

Le développement de la science statistique a permis de relever un défi important : celui de la production en grande série. En effet, lorsque les objets sont nombreux et les mesures coûteuses, il n'est pas envisageable de tous les mesurer. De ce fait, il a fallu rechercher des solutions pour pouvoir, à partir de quelques échantillons, se forger une idée

globale de la production. Ce passage de l'échantillon à la totalité s'appelle « l'inférence statistique ». Pour faire simple, une moyenne et un écart-type, obtenus sur la base d'une dizaine d'échantillons, permettent de se donner une idée sous certaines hypothèses de la distribution de tous les objets produits. Dans ce cas, les erreurs de mesure, inévitables, se cumulent avec les incertitudes liées à la qualité et l'effectif de l'échantillonnage. Ces paramètres (moyenne et écart-type) ne sont donc pas les vrais paramètres de la population, mais des images inexorablement imparfaites...

Néanmoins, toute une théorie visant à améliorer le pilotage des process industriels s'est développée autour de l'approche probabiliste : la M.S.P. Elle a investi de nombreuses usines, mais pas toutes ! Il existe encore de nombreux sites industriels qui n'ont pas franchi ce cap, la complexité, réelle, de la théorie mathématique sous-jacente n'y étant probablement pas étrangère...

Au-delà de la M.S.P, la théorie statistique est aussi appliquée dans le domaine de "l'établissement des recettes". Les plans d'expérience permettent en effet de quantifier l'impact de tel ou tel facteur suspecté d'agir sur le process afin d'en donner les limites de variation (tolérances) acceptables eu égard au résultat final attendu. Cette approche est souvent compliquée, inévitablement longue à mettre en œuvre et n'offre aucune garantie de succès, car des facteurs peuvent être ignorés alors qu'ils ne sont peut être pas négligeables. De plus, et là encore, nous sommes dans un monde de mesures qui ne reflète pas la juste réalité, mais, faute de mieux, on fait avec...

Quelques outils de la statistique ont été "packagés" dans une théorie que nos amis anglo-saxons ont nommée Six Sigma. Si cette approche a connu, et connaît encore, de larges succès industriels (de Motorola à Nissan), elle a du mal à pénétrer le monde industriel national et surtout celui des petites entreprises. Le coût et le

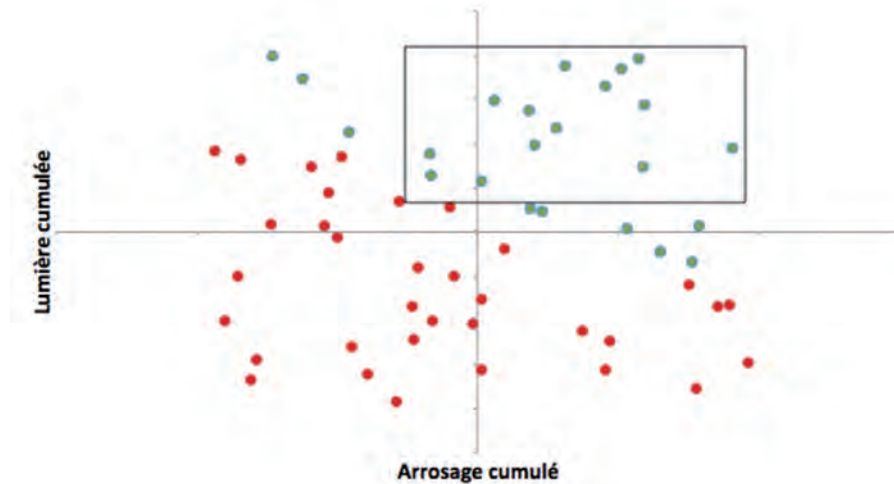


Figure 1. Exemple de représentation graphique en deux dimensions de facteurs à considérer pour modéliser la culture de plantes.

temps de la formation des experts du domaine (Green belt, black belt...) sont souvent jugés rétrogrades pour des résultats dont de nombreuses personnes doutent, souvent d'ailleurs faute d'avoir réellement essayé !

De nos jours, qu'elles soient déterministes ou probabilistes, nos recettes sont robustes aux erreurs de mesure, c'est-à-dire qu'elles n'y sont pas vraiment sensibles. De ce fait, elles n'expriment pas le besoin fonctionnel réel, laissant place à un potentiel d'amélioration important. L'ère qui s'ouvre maintenant, l'industrie 4.0, offre une perspective alternative d'ores et déjà expérimentée avec succès en différents domaines : le datamining.

### Big Data et Datamining : la recherche du "Graal" dans l'hyperespace

Si l'inférence statistique est née du besoin de quantifier une population totale sur la base de quelques échantillons, pour contourner la question des

coûts importants liés à la mesure, la technologie permet souvent aujourd'hui de sortir de ce contexte. En effet, avec les capteurs actuels de toute sorte, filaires, bluetooth ou wifi, les automates connectés, les systèmes numériques de contrôle commande (DCS) qui délivrent de l'information en mode quasi continu sur tout type de paramètres, on entre dans l'ère du "100 % mesuré", en tout cas du "100 % mesurable". Nous voyons par exemple se développer des machines 3D capables de mesurer les pièces en ligne, à 100 %<sup>(5)</sup> !

Toutes ces données qui deviennent disponibles sont également aisément stockables, à un coût quasiment dérisoire, grâce aux évolutions extraordinaires en ce domaine. Elles sont donc exploitables et archivables et la révolution pour la métrologie de demain me semble bien se dessiner dans ce contexte...

Pour comprendre les enjeux, nous pouvons considérer un exemple simple.

Imaginons que nous souhaitions trouver la recette optimale pour obtenir des plantes d'une longueur supérieure à une limite. Nous pourrions par exemple identifier les facteurs "arrosage" et "lumière" comme facteurs à considérer. En

acquérant chaque jour, et tout au long de la production, la quantité de lumière et la quantité d'arrosage reçues par chaque plante, il est possible, au moment de la récolte, d'associer le paramètre "longueur" (grandeur d'intérêt) aux paramètres influents "arrosage cumulé" et "lumière cumulée". Chaque plante peut alors être représentée par un point de coordonnées ("arrosage cumulé" ; "lumière cumulée") dans un simple plan (X ; Y). Le point est par exemple vert si la longueur est supérieure à la limite, rouge dans le cas contraire. En déplaçant un carré (ou plutôt un rectangle) dans le plan, nous pouvons rechercher un endroit dans lequel nous trouvons une surdensité de points verts<sup>(6)</sup>. Les cotés du rectangle dans lequel se trouvent les points verts donnent les quantités cumulées de lumière et d'arrosage qui permettent d'obtenir des plantes de longueur au moins égale à la longueur souhaitée.

Imaginons maintenant que nous considérions un troisième facteur influent, la quantité d'engrais distribuée. La démarche est rigoureusement la même si ce n'est que nous ne sommes plus dans le plan, mais dans l'espace.

Chaque point a en effet trois coordonnées (lumière, arrosage, engrais) et il est rouge ou vert suivant que la plante est inférieure ou supérieure à la limite. Cette nouvelle coordonnée doit elle aussi être explorée. Il nous faut donc maintenant un cube (en réalité un parallélépipède) pour parcourir l'espace et trouver, comme ci-avant, une zone ne contenant que des points verts. Les longueurs en X, Y et Z du parallélépipède donnent alors "la recette" pour obtenir les plantes voulues en tenant compte des 3 facteurs identifiés.

« Pour réussir, il faut également que les données disponibles soient de qualité, c'est essentiel : voilà la mission du métrologue de demain... »

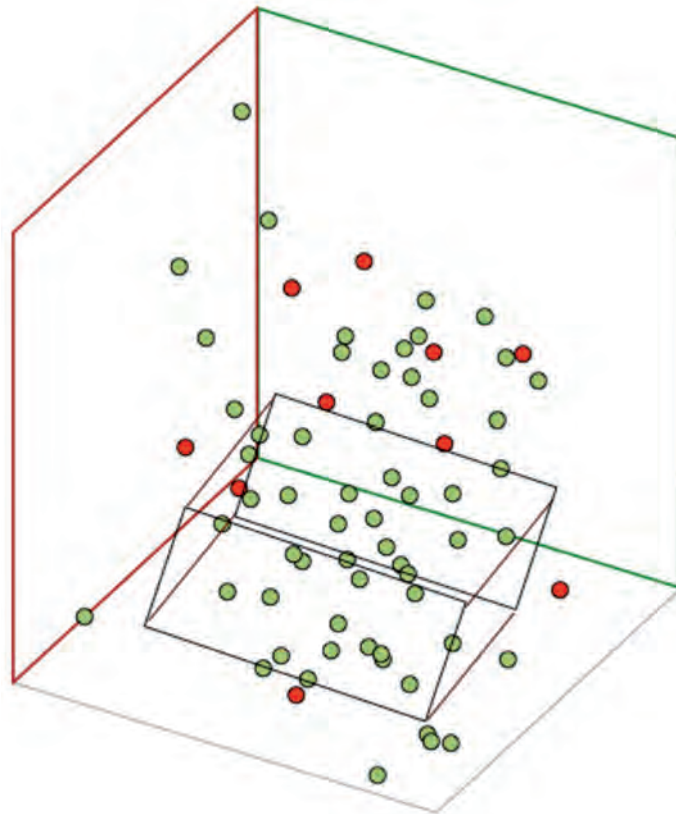


Figure 2. Représentation graphique en trois dimensions d'un troisième facteur influent pour la culture de plantes. Nous ne sommes plus dans le plan, mais dans l'espace. En ajoutant d'autres facteurs, on passe dans l'hyperespace.

En ajoutant un, deux, dix, cent, mille autres facteurs d'influence, on ajoute autant de dimensions à notre espace précédent qui devient alors un hyperespace<sup>(7)</sup>. Pour retrouver les zones de densité maximale en points verts, nous devons scruter les  $n$  dimensions de cet hyperespace et il nous faut donc passer d'un cube à un hypercube (de cotés de longueurs variables évidemment, donc pas réellement "hypercubique").

Dans cette nouvelle approche, les données expérimentales sont primordiales, car ce sont sur elles que repose la recette optimale, le « Graal ». Ainsi, elles doivent être métrologiquement de qualité et rester comparables dans le temps pour continuer à être exploitables. Contrairement aux approches précédentes, la recette n'est pas robuste aux erreurs de mesure. Le métrologue de demain devra donc s'investir, non plus dans la "qualité" des instruments de mesure (étalonnage), mais bel et bien dans la qualité des mesures pour que la technologie (capteurs, stockage et capacité de calculs) serve de façon optimale la performance de son entreprise. En datamining, chaque

nouvelle expérience (production) améliore la connaissance. Elle doit donc être considérée, et il faut s'en donner les moyens, comme un pas de plus en direction du "Graal".

Pour conclure en rassurant les industriels, et particulièrement les métrologues, la difficulté du datamining ne réside plus dans l'application de formules incompréhensibles. Il s'agit, et la difficulté est bien réelle, de savoir piloter efficacement l'hypercube dans l'hyperespace afin de trouver les conditions optimales de réalisation, le fameux « Graal » !

Pour réussir, il faut également que les données disponibles soient de qualité, c'est essentiel : voilà la mission du métrologue de demain...

Toujours soucieuse de préparer l'avenir, Delta Mu s'intéresse à cette technologie. Les pilotes actuels d'hypercubes sont peu nombreux et font office de véritables pionniers. Nous avons la chance, chez Delta Mu, d'avoir fondé l'an passé avec l'un d'entre eux, IP Leanware, l'association Auvergne Efficience Industrielle<sup>(8)</sup>, labélisée par la région Auvergne Cluster d'Excellence ●

(1) La phase de validation peut conduire à revoir la conception initiale, et ceci autant de fois que nécessaire à l'obtention du résultat attendu. Cette mise au point récursive est consommatrice de temps. Elle impacte non seulement le prix de revient de l'objet, mais aussi le *time to market*, ce dernier pouvant remettre en cause la pertinence du produit lui-même.

(2) Le concept de 3<sup>e</sup> révolution industrielle a été popularisé par M. Jeremy Rifkin. Il s'agit du développement des nouvelles technologies de l'information et de la communication, à partir du milieu du XX<sup>e</sup> siècle.

(3) Les productions industrielles répondent à ce caractère probabiliste.

(4) Bertrand Russell (1872, 1970) est un mathématicien, logicien, philosophe, épistémologue, homme politique et moraliste britannique.

(5) Par exemple la société MS3D.

(6) Nous pouvons aussi rechercher une zone dans laquelle nous trouvons un maximum de densité de points verts.

(7) L'hyperespace est un « espace » qui possède plus de trois dimensions.

(8) Auvergne Efficience Industrielle est une association fondée par trois entreprises auvergnates : Phiméca, IP Leanware et Delta Mu, dont l'objectif est de développer des outils/services innovants tournés vers l'amélioration factuelle de la performance industrielle.