



Le crépuscule des données

*Brochure à caractère commercial où la SCM démontre,
en un chapitre unique, sa parfaite connaissance
des méthodes anciennes, présentes et futures,
pour l'aide à la décision*

Ex ungue leonem !

Des difficultés nouvelles

Les entreprises rencontrent souvent des difficultés, même dans leur cœur de métier. Ici, des équipements critiques ne fonctionnent pas correctement, la sûreté est mise en cause, l'impact sur l'environnement est mal mesuré. Là, on s'aperçoit que les attentes des clients ou consommateurs sont mal prises en compte : le produit phare est délaissé et la politique commerciale est à revoir. Les règles ont changé, les méthodes de conception, de calcul et de qualification aussi. Mais n'accusons pas trop vite le ciel : les entreprises elles-mêmes ont leur part de responsabilité ; elles n'ont pas su prendre les devants ; elles attendent tout des Pouvoirs Publics. Au mieux, elles ont confié des analyses prospectives à des débutants. Le résultat de leur travail, promu "outil informatique d'aide à la décision" par le miracle de la technologie, permet toutes sortes de visualisations en 2d, 3d, en relief et en couleurs, mais c'est un logiciel mal conçu, mal maîtrisé, mal exploité. Un bédouin n'en voudrait pas pour sa tente en plein désert : "Produits avariés nés d'un siècle vaurien", disait déjà Baudelaire ("Les Fleurs du Mal").

Des approches anciennes

Toutes ces questions, vitales pour les entreprises, ont fait l'objet d'études approfondies depuis longtemps et sérieusement à une époque où cela se pratiquait encore. Mais les approches retenues il y a 40 ou 50 ans, prolongées depuis, sont maintenant obsolètes. A l'époque, faute de données spécifiques, on adoptait des lois académiques pour qualifier la durée de vie d'un équipement ou la propagation d'une pollution ; elles résultaient d'un consensus entre experts. De même, les démonstrations de sûreté étaient réalisées de manière empirique, en prenant de confortables "marges" à toutes les étapes de la conception et de la réalisation, ce qui aujourd'hui n'est plus possible.

Par principe, faute de connaissances précises et spécifiques, faute de retour d'expérience exploitable, les approches du passé, aussi bien pour la construction que pour le dimensionnement d'équipements, étaient empiriques et grossières. Le résultat était généralement satisfaisant : bon nombre de ponts romains tiennent encore. Pour s'assurer de la parfaite compétence des experts, les Romains avaient adopté les "lex de maiestate" (lois de majesté) : quiconque commettait une erreur de calcul avait "offensé la majesté du peuple romain" et était condamné à la peine capitale.

Un nouveau souci de précision

Aujourd'hui, à l'inverse, la tendance est à la précision factice dans les modèles. On prétend comprendre toutes les lois de la physique : "l'humanité bavarde, ivre de son génie", disait encore Baudelaire. Ces lois, après les avoir comprises, assimilées, dominées même !, les professionnels vont les coder sous forme de "boîtes noires", de logiciels exploitables par tous, qui vont maintenant servir à construire un bâtiment, un pont, une aile d'avion, etc.

Malheureusement, il y a plus de choses sur la terre et dans le ciel que n'en rêve cette philosophie, dont la prétention est souvent excessive : on a simplifié les contraintes à l'excès, négligé quantité de facteurs, comme par exemple le vieillissement des composants. La modélisation 3d par éléments finis a été faite de manière simplifiée, et les lois de la physique sont linéarisées pour être traitées par la machine ; tout ceci est invisible de l'extérieur. On dispose d'un très beau logiciel, à l'interface bien réglée, qui retourne 15 chiffres après la virgule, quelle que soit la question posée. Les journalistes s'extasient, les politiques applaudissent, les financements pleuvent.

Mais la Nature reprend vite ses droits : Satan conduit le bal. Réaliser un outil informatique ne permet jamais la prise en compte des incertitudes : un logiciel retourne un résultat précis à partir de données précises. Or les incertitudes sont inhérentes aux lois de la Nature, qui comportent toujours une certaine variabilité. Chacun a à l'esprit des exemples de toits de supermarché qui s'effondrent sous le poids d'une couche de neige pourtant bien mince et bien prévisible, des exemples d'appareils sophistiqués qui tombent en panne pendant leur premier mois d'utilisation, de tramways qui sortent de leurs rails, etc. Les journalistes cessent de s'extasier, les politiques d'applaudir, les financements de pleuvoir.



Un tramway sorti de ses rails, Paris, mai 2018

Du bon usage de l'ordinateur

Maintenant qu'elle dispose d'outils informatiques, l'humanité se croit l'égale des Dieux, et mieux encore : l'arrogance des chercheurs n'a d'égale que la sottise des journalistes. Archimède n'avait pas d'ordinateur, mais il savait concevoir et calculer ; l'homme est allé sur la Lune sans ordinateur, mais n'a pas été capable d'y retourner depuis qu'il s'en est équipé.

L'ordinateur est un cache-misère : il va donner une apparence de respectabilité à la pensée la plus pauvre, à la démonstration la plus étriquée, à la conception la plus fumeuse. Les négligences, les insuffisances, les simplifications, les faussetés mêmes, vont être dissimulées dans les replis des instructions informatiques, d'où plus personne ne saura les débusquer. Le résultat, rebaptisé "intelligence artificielle", sera sanctifié par toutes les communautés d'utilisateurs. Les médecins épidémiologistes sauront, en appuyant sur deux ou trois touches, démontrer que les lignes à haute tension sont nocives pour la santé, que les téléphones portables réduisent l'espérance de vie, que les fraises des bois favorisent le saut à skis.

Quiconque approche une voiture est promis à une fin prochaine, immédiate s'il s'agit d'un diesel. La gendarmerie nationale y voit un nouveau moyen de désigner des coupables.

Tous ces programmes informatiques auxquels personne ne comprend rien, que c'est beau ! Et lorsqu'une génération aura passé, personne ne saura plus qui les a conçus, ni pourquoi. Version moderne de "Oceano Nox" : "Rien ne sait plus vos noms, pas même une humble pierre".

Pour conserver l'intelligence et la mémoire, il faudrait par principe ne confier à l'ordinateur que le traitement de problèmes que l'homme sait résoudre sous forme simplifiée : nous savons faire à la main s'il y a 10 variables ; la machine le fera pour 100 000. Il est absurde, toujours par principe, de lui confier des problèmes que nous ne saurions pas résoudre.

La démonstration de sûreté

La voici, telle que la conçoivent les apôtres de la logique artificielle, de la démonstration formelle à la Turing :

1. Le Grenelle de la Sûreté

Un groupe de jeunes gens bien mis, élevés au bio depuis l'enfance, introduit des données dans la machine. Celle-ci ronronne de satisfaction, mais sans ostentation ni hésitation : on sent bien qu'elle sait ce qu'elle fait, qu'elle domine la situation. Au bout d'un moment, un message apparaît sur l'écran : "tout est correct", suivi des logos des entreprises, institutions, centres de recherche, qui ont participé à cette brillante démonstration. Un voyant vert s'allume, avec ce libellé : "merci de votre attention".



Les participants se mettent autour d'une table ; la réunion est maintenant dirigée par un psychologue. A l'unanimité, le vote est favorable : les conditions de sûreté sont respectées et l'exploitation va pouvoir commencer.

Retentit alors l'hymne de la transition énergétique : "Les Tableaux d'une Sous-Exposition", par Nicolas-Modeste Moussorgsky-Bulot. La séance est levée et chacun s'épanouit.

2. La revanche d'Archimède

Eh bien, pour le moment, ce n'est pas cela du tout, une démonstration de sûreté. Elle doit se faire "à l'ancienne" : il faut découper le problème entier en un très grand nombre de petits cas spécifiques, il faut montrer que tous les cas possibles ont été recensés et que, dans chaque cas, on dispose d'une réponse appropriée. C'est ce qu'aurait fait Archimède. Une démonstration de sûreté est une démonstration, au sens mathématique du mot ; le mathématicien seul peut en évaluer l'ampleur. Le résultat est un énorme dossier, de plusieurs centaines de pages, que chacun doit pouvoir analyser à sa guise. Il ne saurait y avoir d'argument d'autorité.

Dans la plupart des domaines, les Autorités de Sûreté, derniers gardiens du Temple de la Qualité, se montrent de plus en plus exigeantes, non seulement sur les résultats attendus, mais sur la rigueur des démonstrations. Si celles-ci sont obtenues au moyen d'outils logiciels, quelle est la valeur de ces logiciels ? Comment ont-ils été validés ? Sont-ils représentatifs ? Les Autorités de Sûreté ne se contentent plus de réponse du type "le logiciel représente un consensus entre experts".

Une idée très répandue chez les ingénieurs, surtout depuis l'apparition de l'informatique et du calcul scientifique, est que les calculs doivent être les plus précis possible : un logiciel d'investigation permettrait ainsi une compréhension fine de chaque situation. Les machines doivent tourner plus vite, faire moins d'erreur d'arrondi, etc. On parviendrait ainsi, avec des boîtes noires mises dans des machines sophistiquées, à avoir une bonne vision des lois de la Nature.

Cette approche, complètement erronée, est systématiquement rejetée par les Autorités de Sûreté, car elle relève de la naïveté et de l'arrogance. Prenons l'exemple d'un processus qui dépend de 50 paramètres (un processus de fabrication industrielle en compte souvent beaucoup plus) et imaginons pour simplifier que chaque paramètre ne peut prendre que 10 valeurs (en pratique, il y en a une infinité). Nous devons alors explorer 10^{50} possibilités, ce qui est hors de portée d'un ordinateur quel qu'il soit. Imaginons que chaque run prenne une heure (dans la pratique plusieurs jours) ; il faudrait alors de l'ordre de 10^{44} siècles pour tout explorer. Certains experts diront qu'il suffit de choisir les runs au hasard ; un hasard bienveillant guiderait ces explorations précisément vers les situations qui le méritent. Cette forme de mysticisme ne résiste pas à l'analyse ; elle est condamnée par les Autorités de Sûreté. On a moins de chance, procédant ainsi, de détecter des situations à risque que n'en aurait un ivrogne aveugle de franchir une porte légèrement entrebâillée en lançant une fléchette au hasard.

Il faut donc concevoir des outils grossiers, qui s'apparentent au survol d'un territoire à haute altitude par un avion, plutôt que des outils fins, qui s'apparentent au creusement d'un puits à intervalles réguliers. Pour une première analyse, une compagnie pétrolière commence par un survol du relief, et non par des forages tous les km.

Ces outils grossiers sont nécessairement de caractère probabiliste, parce que les lois de la Nature sont très complexes et que nous les connaissons mal. On est donc amené à extraire, pour chaque situation, les paramètres qui auront la plus grande importance et à surveiller ceux-là en premier lieu ; cette méthode s'appelle "hiérarchisation de paramètres".

Notre expérience très ancienne de ces questions nous permet de dire ceci très clairement : ni la science en conserve, sous forme de modèles programmés dans des boîtes noires, ni les dires d'experts, ne peuvent apporter de réponse complètement satisfaisante aux préoccupations actuelles de sûreté et d'impact sur l'environnement.

Intérêt commercial : argument de vente

Les entreprises pourraient peut-être cesser de prendre les consommateurs pour des andouilles, tout juste capables de réagir à une "communication", de plus en plus dépourvue de contenu. On voit des jeunes gens autour d'une table, nourris au bio (les mêmes que précédemment) en extase parce que leur savon de toilette contient du "tétramisophisol au glutamate de pyridium", comme disait Jean Yanne. On sent qu'ils s'épanouissent, objectif premier de la nouvelle génération.

Une argumentation commerciale peut parfaitement reposer sur la qualité des études qui ont été réalisées : les produits sont plus robustes ; on les a testés dans des conditions difficiles (par exemple intempéries, température très haute ou très basse, résistance au vieillissement, etc.). Ces arguments ont toujours été ceux de Volvo et de Mercedes, constructeurs automobiles, et ils trouvent un écho favorable auprès des consommateurs.



On achète peut-être un dentifrice parce que l'on voit une jolie fille avec les dents blanches, mais ce type d'argument ne fonctionne pas pour un produit industriel. Il faut justifier la solidité au moyen d'arguments factuels : l'affirmation ne suffit pas. C'est là que la R&D de l'entreprise intervient, pour fournir des arguments qui, en définitive, seront de nature commerciale. Ils reposeront sur des faits, sur des expériences, sur des tests. Insistons-y bien : montrer que les process de l'entreprise sont entièrement sous contrôle est, en définitive, un excellent argument commercial.

Le crépuscule des données

On a bien cru, à un moment, voir la dernière donnée à son dernier soupir, sous la double pression des modèles académiques et des consensus éclairés. Pour une bouchée de pain, pour une poignée de seigle, n'importe quelle équipe académique, dotée d'une blouse blanche et d'un ordinateur, vous fera un modèle de propagation de la diphtérie en zone rurale au Dahomey. Le Grenelle de l'Environnement, en 2007, a pour sa part consacré l'abandon de toute expérimentation au profit de consensus : plus de faits, plus de données, plus de mesures ; il suffit de mettre les gens autour d'une table et qu'ils soient d'accord : on en déduira que le produit est nocif et on prendra les dispositions législatives qui s'imposent.

Satan, de temps en temps, réagit à sa manière : d'un revers de fourche, il détruit une usine parfaitement certifiée ; il noie une zone, inondable ou non ; il fait ici et là dérailler un tramway parfaitement conçu, mais rien qui, dans l'ensemble, puisse émouvoir l'espèce humaine, ivre de son génie.

Concevoir un système d'information, mettre des capteurs en place, récupérer les données, tout cela a un coût, bien évidemment ; c'est difficile, c'est fastidieux. Le système ne fonctionne pas du premier coup ; il génère quantité de données aberrantes et contradictoires.

Quand on n'a pas de données, on est tranquille ; l'homme a bien évolué depuis Pascal, et maintenant il sait rester au repos dans une chambre. On peut tranquillement décider que les orbites des planètes sont circulaires : c'est harmonieux et satisfaisant. Mais, pour reconstituer l'orbite elliptique de Mars, il a fallu cinq ans de travail à Kepler, à partir des observations recueillies par Tycho Brahé.

Bien sûr, chacun voit qu'il y a quelque inconvénient à se passer complètement de données, mais le mieux est d'en avoir le moins possible : une seule est l'idéal, pour éviter toute contradiction. Si l'on en a deux, il faut qu'elles montrent que le phénomène évolue dans le sens désiré. Au-delà de deux, c'est l'horreur, c'est l'anarchie.

L'apport de l'intelligence artificielle

Tout à coup, quelqu'un a eu une idée de génie : des données, il y en a partout ; il suffit de savoir les récupérer et les traiter. C'est ce qu'on appelle le "big data". Plus besoin de se fatiguer à calibrer des instruments, à sortir le soir sous la pluie pour vérifier que les capteurs ne sont pas bouchés : on va utiliser des données "externes".

Nous aurions besoin de quantifier le nombre et l'impact des séismes, dans la région du sud de l'Italie, sur 2 000 ans : cela permettra d'étoffer les dossiers de sûreté pour les installations à risque dans la région. Mais les données n'existent pas, ou sont malaisées à recueillir : il faut dépouiller de vieux grimoires, voir s'ils mentionnent les dégâts au clocher de l'église, les fentes apparues dans les fortifications et convertir tout cela en magnitude de séisme. C'est long, fastidieux et imprécis : le résultat sera discutable.

Mais, ô merveille de la technologie, nous disposons des relevés des tournées des livreurs de sushis dans la région de Rennes, sur la période 2010-2015, avec quantité d'attributs très bien renseignés : heure de départ, heure d'arrivée, chemin suivi, type de scooter, nom du conducteur, etc. Voilà une mine d'informations de bonne qualité, que nous allons exploiter. En effet, le département d'astrophysique quantique de l'INSA de Rennes va nous montrer que la corrélation entre la livraison de sushis à Rennes et les séismes dans le sud de l'Italie est de 0.35. Rien ne résiste à l'INSA de Rennes, pas même les livraisons de sushis.

Vous vous interrogez sur l'organisation des villes de demain, d'où, comme tout bon politique, vous voulez bannir l'automobile ? Interrogez donc l'Ecole d'Ingénieurs de la Ville de Paris qui, pour quelques bouchées de pain rassis, saura les intégrer dans des "questionnements plus vastes sur la ville post-carbone". Tous ces modèles dont pas un ne mourra, que c'est beau !

Le big data se situe au confluent des deux caractéristiques principales de l'espèce humaine, qui sont la sottise et la paresse (la lésine, disait Baudelaire). Il permet de se doter de données à peu de frais, de les valider d'une manière en apparence scientifique et de sélectionner celles dont le responsable a besoin pour avoir raison, ce qui reste à l'évidence son but ultime.



Les méthodes de travail de la SCM

La SCM s'efforce de comprendre les lois de la Nature et de les restituer en langage approprié, de manière à permettre la prise de décision, après analyse des données disponibles. Nous ne sommes pas là pour donner raison à qui que ce soit, mais pour apporter un compte-rendu aussi objectif que possible. Prenons une comparaison : le propriétaire d'un champ fait appel à un géologue pour savoir si son champ recèle du pétrole, ou au contraire une pollution ; il paie le géologue, mais celui-ci doit fournir une réponse honnête, reflétant la réalité de ses investigations : il n'est pas là pour faire plaisir au propriétaire, même si celui-ci le paye.



Les mathématiques sont là pour comprendre les lois de la Nature ; c'est Satan qui les a conçues.

Nos compétences

Elles portent, par définition, sur tout ce qui peut être traité de manière quantitative. Ce peut être collectif (statistiques) ou individuel (optimisation d'un process particulier). Nous disposons de fiches détaillées, accessibles sur notre site web :

<http://scmsa.eu/fiches/fiches.htm>.

Nos principaux domaines d'intervention sont :

- **Qualité des données** : détection de données aberrantes, reconstruction de données manquantes, régularisation des données bruitées, amélioration des systèmes d'information. Nous avons écrit deux livres sur ces questions, détaillant les méthodes que nous employons.

Principale référence : Nuclear Energy Agency (OCDE) : détection de données aberrantes dans les bases de données nucléaires, 2010-2012, 2014, 2015, 2016, 2017.

- **Amélioration des process métier** : contrôle de la qualité, hiérarchisation des paramètres intervenant au cours d'un process (déterminer ceux qui sont critiques et doivent être surveillés en particulier), définition d'une maintenance préventive, amélioration des réglages, réduction de la variabilité du process en sortie.

Références : SNECMA Propulsion Solide : amélioration de la fiabilité des composants, ArcelorMittal : amélioration de la qualité d'un usinage d'aciers spéciaux, DCNS (Naval-Group) : amélioration du soudage par faisceau d'électrons.

- **Analyses de sûreté**, tout particulièrement dans le domaine du nucléaire ; analyse critique des rapports d'experts, prise en compte des incertitudes, analyse et quantification des risques, évaluation de la fiabilité, prise en compte des phénomènes extrêmes, rédaction des démonstrations de sûreté, analyse des dysfonctionnements des capteurs, mise en évidence de fausses alarmes, dimensionnement des réseaux de surveillance.

Références : Convention-cadre avec l'IRSN, 2010-2015 : méthodes probabilistes pour la sûreté nucléaire. EDF/SEPTEN, 2015 : Prise en compte des incertitudes dans les Etudes Probabilistes de Sûreté. ANDRA, 2016 et 2017-2018 : Optimisation du placement de capteurs dans un site de stockage de déchets radioactifs.

Amélioration de la politique commerciale : analyse statistique des comportements de la clientèle, analyse des situations "à risque", mise en évidence des situations les plus satisfaisantes, construction d'un tableau de bord, définition d'outils marketing et construction de panels de consommateurs.

Références : Rhodia, puis Solvay : Préviation des ventes mondiales d'automobiles, 2009-2010, 2012-2013, 2014 ; Monceau Assurances, 2014, 2017-2018 ; Carrefour/Bazar, 2016-2017.

Défense : Contre-mesures, vulnérabilités, algorithmes temps réel, algorithmes embarqués, simulations, jeux de guerre, traitement de l'image. Leurrage anti-missiles (guidage IR et visible) et anti-torpilles.

Références : US Army Research Office, Délégation Générale pour l'Armement, Etats-Majors (EMAA, EMM). Dernier contrat traité : le "Bouclier Naval" : protection d'un navire de surface contre une menace terroriste à courte portée, en cotraitance avec Thales Naval France et TDA Armements SAS, la SCM étant responsable de la faisabilité du système complet.

Transports : comparaison de divers modes de transports, organisation, rentabilité, maintenance, sécurité, amélioration de la logistique, planning de remplacement d'équipements ;

Références : RATP, 2016 : Assistance scientifique pour la définition du planning de remplacement pour des équipements critiques ; SNCF Réseau, 2016 : Appui scientifique pour l'analyse des scénarios relatifs à une ligne nouvelle ; RATP, 2016-2017-2018 : Modélisation du comportement des trains en situation de freinage d'urgence ; RATP, 2017 : Réalisation d'un outil de simulation des temps d'acheminement des trains de travaux ; SNCF/Transilien, 2017 : Analyse critique de modèles de représentation des déplacements ; réalisation d'un outil de simulation.

Environnement : politique de l'eau, maintenance des réseaux ; modèles de propagation des pollutions, dans l'eau, dans l'air, dans les sols. Caractérisation des pollutions. Reconstruction de pollutions globales à partir d'échantillons incomplets. Evolution dans le temps.

Références : contrat cadre "Méthodes probabilistes pour l'environnement", avec l'Agence Européenne de l'Environnement (2006-2010, 2011-2015) ; Direction Générale Energie-Climat, MEDD : définition d'une méthode statistique robuste relative à l'évaluation de la qualité de l'air 2014-2015.

es livres que nous écrivons et diffusons

La SCM édite et commercialise une série de livres "Mathématiques du réel", essentiellement à destination des ingénieurs. Il s'agit d'une introduction aux Lois de la Nature et les probabilités y sont utilisées comme outil principal. Voir :

http://www.scmsa.eu/livres/SCM_livres.pdf.

olloques

Chaque année, la SCM organise deux ou trois colloques, sur les sujets d'intérêt du moment. Ils sont toujours gratuits et ouverts à tous. La liste des colloques qui se sont tenus est disponible dans la rubrique "archives scientifiques" <http://scmsa.eu/archives.htm> ; on y trouvera également les liens pour télécharger les exposés.