

La Lettre de la S.C.M.



Mars 2022

Numéro 97

ISSN : 2112-4698

*Il n'y a point de plus cruelle tyrannie que celle que l'on exerce à l'ombre des lois
et avec les couleurs de la justice (Montesquieu)*

Éditorial par Bernard Beauzamy :

Méthodes probabilistes et élévation des températures

Nous ne pensons pas que les méthodes probabilistes soient appropriées pour juger de l'évolution des températures, en un point donné, sur de longues périodes (notion de scénario), vers le passé ou le futur. Il en va exactement de même de la propagation d'une épidémie.

L'évolution de la température repose sur trois déterminants fondamentaux :

- La Terre est sortie, voici environ 10 000 ans, d'une période glaciaire qui a duré environ 100 000 ans. Le réchauffement actuel pourrait n'être qu'un retour à la normale ;
- L'influence de l'activité solaire n'est pas du tout connue. On sait qu'elle n'est pas constante, mais non comment elle varie et comment cela influence la température du sol ;
- On sait que la géothermie existe (les volcans), mais personne ne sait estimer quelle est l'influence au niveau du sol.

Celui qui dispose de quelques milliers de mesures (des relevés de température) aimerait utiliser les méthodes probabilistes, s'emparer de leur vocabulaire, et en acquérir le pouvoir. Il se met à parler d'intervalle de confiance, d'indépendance, de durée de retour, etc., en oubliant que les probabilités sont une science exacte, basée sur la répétition des phénomènes. En particulier, nous n'avons aucun moyen de savoir si nos relevés sont suffisamment longs pour permettre une évaluation des phénomènes extrêmes.

Un autre point majeur de désaccord entre les deux communautés est que celui qui dispose de quelques relevés aura tendance à croire qu'ils sont issus de mesures indépendantes : cela simplifie les calculs. Mais le probabiliste sait bien que la température un jour et la température le lendemain ne sont pas indépendantes, non plus que la température une année donnée et l'année suivante.

Les notions de variance et d'intervalle de confiance vont caractériser la répartition des résultats en cas de répétition de l'expérience, mais quel sens leur donner si on dispose, en tout et pour tout, d'une unique série temporelle ?

La première chose à vérifier est que l'étude envisagée ne

contredit pas les critères de répétabilité. Pour tester de la durée de vie d'un appareil, on imaginera des dizaines d'appareils semblables, répartis sur des dizaines d'usines.

Les probabilités et le traitement de données numériques n'ont pas le même usage et la difficulté provient souvent de la confusion qui est faite. Les probabilités sont destinées à comprendre les lois de la Nature : c'est une science exacte, qui produit des théorèmes. A l'inverse, les données numériques apportent des outils d'investigation, par exemple le prolongement linéaire, qui ne constituent pas des théorèmes : n'importe qui a le droit de prolonger une série temporelle.

Prenons un exemple concret. Un système physique (où l'on mesure la température) sera considéré comme ayant une loi stationnaire s'il est constitué d'une grosse sphère, à l'intérieur de laquelle se trouvent de petits bulletins. Sur chaque bulletin on a inscrit une température. Chaque jour, un robot ouvre la sphère, prend un bulletin au hasard, règle la température et remet le bulletin dans la sphère. Cette loi est stationnaire, parce que la sphère est constamment la même, de même que le protocole de lecture. Naturellement, la température n'est pas constante. On note p la probabilité d'une température donnée, un jour donné.

Notons K le nombre de températures différentes. Supposons qu'un jour donné, la température 34°C soit apparue. Quelle est la probabilité de la revoir au bout de k jours, mais non avant ? $(1-p)^{k-1} p$. Quelle est l'espérance de ce

temps d'attente ? $E = p \sum_{k=1}^{+\infty} k (1-p)^{k-1} = 1/p$. On a démontré

un théorème : la durée de retour (durée moyenne du temps d'attente avant de revoir une étiquette) est égale à l'inverse de la probabilité.

On peut vouloir regarder ce que signifient ces mêmes concepts pour une série numérique quelconque. La probabilité p n'est pas connue et n'est pas la même pour toutes les étiquettes ; la durée de retour n'existe pas, parce que le phénomène correspondant ne s'est jamais produit. Nous n'avons plus un théorème, mais un abus de vocabulaire.

Bernard Beauzamy

Sarp, site de Limay

Nous avons présenté à la mi-mars le rapport réalisé pour SARP (filiale de Veolia, site de Limay) ; Sarp exploite des fours, où les déchets dangereux sont incinérés. Le four no 3 est représenté par la valeur d'environ 180 paramètres, chacun étant mesuré toutes les minutes.

Un premier contrat, en 2019, concernait le réglage des fours, de manière à en optimiser le rendement. SARP Industries voulait que le fonctionnement des fours soit le plus stable possible, tout en respectant les seuils de rejet en sortie de cheminée. Le suivi que nous avons réalisé a montré que, dans l'ensemble, c'était bien le cas.

Un second contrat nous a été notifié en 2019 ; il avait pour objectif de réduire les émissions de CO₂. La SCM s'est vu remettre l'ensemble des mesures recueillies sur quelques années, par minute et pour chacun des paramètres, mais aucune indication sur la manière dont les indicateurs sont supposés être liés au fonctionnement du four ; à peine la définition de l'unité, dans le meilleur des cas. Certains fichiers, au format csv, pesaient plus d'un Gv et nécessitaient le recours à un éditeur spécialisé.

Notre premier travail a été la constitution d'une bdd "propre", faite de moyennes journalières ; nous avons insisté sur le fait, que, en rythme annuel, SARP devait pouvoir justifier de la qualité des mesures auprès des autorités de surveillance de l'environnement. Le grand public veut des données fiables et compréhensibles ; dans l'état actuel des choses, les données aberrantes sont beaucoup trop nombreuses.

Il est toujours possible de réduire les émissions de CO₂ en réduisant tous les paramètres, y compris ceux (que nous ne connaissons pas) qui caractérisent le bon fonctionnement du four, mais ne n'est évidemment pas ce que l'on souhaite. Il s'agit ici de réduire certains paramètres, tout en préservant d'autres.

De manière générale, comment peut-on procéder pour réduire les émissions de CO₂ sans sacrifier le rendement ? On suppose que le travail préliminaire relatif à la base de données a été fait. On note T_1 le paramètre relatif au rendement et T_2 celui relatif aux émissions de CO₂. On va chercher si, dans l'ensemble des données recueillies, certaines situations répondent, de près ou de loin, à la question posée.

Pour cela, on divise les données du tableau en trois plages : celles pour lesquelles le rendement est bon, celles pour lesquelles il est médiocre, celles pour lesquelles il est faible. On commence par regarder les lignes de bon rendement : quelles sont les situations où les émissions de CO₂ sont faibles, s'il y en a. S'il n'y en a pas, on passe au rendement inférieur, et ainsi de suite. Plus généralement, pour chaque valeur du rendement (ou pour chaque classe de valeurs), on détermine, par référence au tableau de données, la valeur minimale effective des émissions de CO₂.

L'avantage de cette approche est qu'elle peut être informatisée : c'est une suite de tris informatiques. L'inconvénient est qu'elle ne fera apparaître que les réglages déjà utilisés dans le passé, sans possibilité d'en suggérer de nouveaux et sans aucune explication quant à la nature des phénomènes.

Cette approche, simple à mettre en œuvre, pourra être utilisée de manière préliminaire. Une approche reposant sur une compréhension des lois de la physique a peu de chances d'aboutir, s'agissant de 180 paramètres ; toutefois, un expert du domaine saura immédiatement désigner les plus influents.

Monceau Assurances

Nous terminons prochainement le travail commencé en 2017 : fournir à Monceau Assurances des outils d'aide à la négociation des primes annuelles, pour les compagnies, appelées "cédantes", qui souhaitent partager avec Monceau tout ou partie des risques résultant des catastrophes naturelles, en particulier les tempêtes.

Au fil des années, et sans que personne ne l'ait dit explicitement, nous nous sommes aperçus que plus le risque était grand (par exemple des tempêtes exceptionnelles), moins il était traité de manière rationnelle. Le gré à gré, les discussions entre collègues, remplaçant vite le recueil de l'information indispensable (force et trajectoire des tempêtes, position des habitations concernées, etc.). Le recours systématique à des lois de probabilité académiques (Gumbel, Weibull, Gauss, etc.) fait que la probabilité des phénomènes extrêmes est sous-évaluée de manière systématique. Enfin, le recours fréquent à un vocabulaire inapproprié (durée de retour, probabilité d'occurrence, voir plus haut) ne fait qu'ajouter à la confusion. Personne ne sait plus de quoi il parle, et ce ne sont plus que chuchotements pour se mettre d'accord.

Nous avons voulu proposer à Monceau Assurances un dernier travail, très simple à mettre en œuvre en principe, et fondé uniquement sur des données comptables, donc indiscutables. Prenons les dix dernières années (durant lesquelles il ne s'est quasiment rien passé, si bien que la mémoire collective est à son plus bas). Prenons un certain nombre de cédantes, présentes depuis longtemps, et faisons l'hypothèse qu'elles n'ont pas évolué : même portefeuille pour l'essentiel. Chaque année, pour chaque cédante, l'accès aux données comptables permet de savoir quelles primes elle a versées et quels remboursements elle a obtenus, et ce indépendamment du nombre de tempêtes de cette année-là.

Faisons la somme, une année donnée, sur l'ensemble des cédantes, nous avons le "bénéfice comptable" de la réassurance pour Monceau (pour bien faire, il faudrait tenir compte des frais de gestion, mais n'entrons pas dans les détails). Faisons maintenant la somme de ces bénéfices comptables, pour les dix ans en considération : nous avons le bénéfice comptable de l'opération "réassurance tempêtes" sur 10 ans, pour l'ensemble des cédantes. D'un côté, l'ensemble des sommes perçues, d'un autre côté l'ensemble des sommes versées. Actuellement, l'opération est légèrement positive pour Monceau, qui gagne un peu d'argent.

Maintenant, posons-nous la question : admettons que, en 2023, survienne une tempête exceptionnelle, du niveau de Lothar et Martin (1999). Elle n'a évidemment pas été anticipée, si bien que les primes n'ont pas augmenté. Les remboursements à octroyer, en revanche, seront beaucoup plus élevés, ce qui est susceptible de détruire l'équilibre du système "cat nat". Mais ce n'est pas le cas, les sommes en jeu à chaque fois étant insignifiantes. La réassurance en question ne s'adresse qu'à des particuliers et les dommages couverts sont très restreints : volets arrachés, perron endommagé, etc. Dans ces conditions, le retour de Lothar et Martin l'an prochain n'a rien qui puisse inquiéter Monceau et il n'est pas utile de proposer une refonte du système de primes, d'autant que l'on ne saurait pas sur quelles bases la calculer. La principale caractéristique factuelle du système est sa totale opacité : aucune donnée factuelle n'est présente ; toutes les données proviennent d'autres données, tout aussi invérifiables, au moyen de formules de calcul entièrement empiriques. Vouloir tout remettre à plat pour repartir sur des bases saines est pure naïveté : les données indispensables n'existent pas et, de toute façon, personne ne voudra du résultat. Il vaut donc mieux laisser les choses en l'état.

Bouygues Energies et Services

La SCM SA a récemment signé un contrat avec Bouygues Energies & Services ; il porte sur l'optimisation de la maintenance du Tribunal Judiciaire de Paris. Ce bâtiment est le résultat d'un Partenariat Public-Privé, et a été financé, conçu, construit et est exploité par Bouygues. Le contrat d'exploitation doit durer 27 ans à partir de son début d'utilisation. La livraison du bâtiment a eu lieu mi-2017 et les premiers usagers ont emménagé en avril 2018.

L'objectif principal de Bouygues E& S est de maintenir les niveaux de performance tout au long de la période d'exploitation, tout en s'efforçant de réduire les coûts. Deux modes d'organisation sont actuellement privilégiés. Le premier est la maintenance préventive : les équipements sont remplacés à des fréquences fixes, déterminées grâce à des retours d'expériences et aux données communiquées par les fabricants. La maintenance préventive est pratiquée par de nombreux industriels (y compris bien sûr dans le domaine de l'automobile), mais ne donne pas toujours satisfaction : tantôt elle est trop stricte et on remplace des pièces qui n'avaient pas vocation à l'être, tantôt elle est trop laxiste et les pannes restent nombreuses. De plus, la maintenance préventive doit dépendre des conditions d'exploitation, qui ne sont pas toujours connues et qui vont évoluer avec le temps : certains équipements s'usent plus vite dans certaines conditions d'utilisation.

Le second mode d'organisation est le Gros Entretien Renouvellement (GER), beaucoup plus grossier comme son nom l'indique, mais aussi beaucoup moins coûteux puisqu'un certain nombre d'interventions sont faites en même temps. Il s'agit d'un plan pluriannuel de maintenance qui liste tous les équipements avec leur durée de vie approximative. On s'attend à ce qu'il soit mis en œuvre sur un ensemble de bâtiments de même nature, tout un quartier par exemple.

L'objectif de Bouygues E&S est celui de tous les industriels : préserver l'équilibre économique du projet, tout en assurant la qualité du service. Les industriels y parviennent rarement, parce qu'ils ont à traiter des réseaux vétustes (par exemple les réseaux d'adduction d'eau) dont l'historique n'est pas connu. Mais ici, le bâtiment est entièrement neuf et nous allons pouvoir enregistrer toutes les données pertinentes : pannes ou dysfonctionnements, réparations éventuelles, etc. Il s'agit d'une situation modèle, que Bouygues E&S contrôle entièrement.

Il est satisfaisant de voir que des outils mathématiques peuvent y aider ; ceux-ci auront nécessairement une forte composante probabiliste, parce qu'il ne s'agit pas d'optimiser des ressources connues dans le cadre d'un besoin connu, mais d'estimer des concepts imprécis, comme la durée de vie d'un système, le "mean time between failure", etc. Les estimations qui sont faites habituellement utilisent des lois académiques souvent inappropriées (lois exponentielles, etc.), entrées dans les traditions il y a cinquante ans et toujours exploitées depuis. Le fait que le bâtiment soit neuf va nous permettre de construire et d'exploiter notre propre retour d'expérience.

Un reporting annuel permettra très vite de savoir si certains secteurs sont plus fragiles que d'autres ; s'il faut y multiplier les capteurs (par exemple des caméras), si les composantes vieillissent bien ou mal, etc. En bref, dès la première année, on aura des indications solides quant à l'évolution prévisionnelle des budgets et des éléments comparatifs, d'un secteur à l'autre. Il sera également possible de faire des simulations : qu'aurait-on gagné ou perdu si, dans tel secteur, la maintenance avait été organisée de manière différente, plus fine ou plus grossière ?

ANDRA

Nous avons remis à l'Andra nos conclusions quant à l'élévation des températures à court terme (+3°C sur 50 ans), au voisinage de Bure. Pour cela, nous utilisons des séries de mesures recueillies dans des stations voisines de Bure ; il y a en tout 9 stations, situées au maximum à 200 km de Bure. Le problème est que les enregistrements ne commencent pas tous en même temps, ne finissent pas tous en même temps, et comportent beaucoup de "trous". La méthode que nous utilisons, purement probabiliste, est présentée dans le livre "Reconstruction de données manquantes" [RDM], de Bernard Beauzamy et Olga Zeydina ; éditions de la SCM SA, avril 2007.

L'idée de base est que si deux séries X, Y ont été enregistrées en même temps, on peut déterminer la loi de probabilité conjointe du couple, puis la loi "marginale" de X sachant Y . On calcule alors l'espérance conditionnelle de X sachant Y . Si maintenant, dans le tableau de données, on trouve une ligne où Y a été mesurée, mais non X , on affectera à X la valeur de l'espérance conditionnelle sachant Y .

Pollution atmosphérique

Par le passé, nous avons été associés à de nombreux projets visant à quantifier la pollution atmosphérique : il s'agissait de réaliser une carte de la France métropolitaine, montrant la concentration en différents polluants et comment elle pouvait évoluer d'une heure à l'autre. Ces cartes servaient généralement de base à une décision politique : réduction de la circulation ou de la vitesse des véhicules. La démarche scientifique était très discutable : personne ne connaît réellement l'origine d'une pollution, ni comment elle se déplace. Disons qu'il s'agissait de donner un alibi scientifique à une décision politique déjà prise.

Aujourd'hui, conscients des aspects arbitraires des décisions politiques, certains industriels ont décidé de prendre leur destin en main, et on ne peut que les en féliciter. C'est ainsi que, récemment, nous avons été en contact avec les Industriels de la Luzerne, représentant approximativement 15 coopératives, exploitant 40 usines au total.

Tout récemment, la Luzerne de France a pris possession d'un logiciel, au standard Excel, destiné à enregistrer les éléments d'inventaire du CITEPA (pollutions atmosphériques) et, avant de l'exploiter publiquement, veut s'assurer du bon fonctionnement de cet outil. On estime que 100 données par usine seront disponibles, sous la forme d'un portail en ligne.

Nous recommandons aux industriels d'être beaucoup plus vigilants quant à la perception que le grand public et les politiques peuvent avoir de leurs activités. Après avoir fait disparaître des pans entiers de l'activité industrielle, les politiques parlent de ré-industrialisation, mais ce n'est qu'un vœu pieux parfaitement hypocrite : attendons de voir la première usine construite et de mesurer l'hostilité de l'opinion publique.

Ce sont les rejets dans l'environnement et les démonstrations de sûreté qui vont inquiéter l'opinion publique et qui doivent faire l'objet d'une attention particulière. Les industriels ne devraient pas hésiter à prendre les devants, publier tous les éléments de mesure dont ils disposent, montrer en quoi les choses ont progressé sur 20 ans, etc. Mais ne soyons pas naïfs : même si l'industriel a créé des millions d'emplois, toute gêne, si minime qu'elle soit, apportée aux déplacements ou aux séjours du Grand Hamster d'Alsace verra la fin du projet. La France du 21^{ème} siècle a ses priorités.

Chercheurs en entreprise

Lorsque nous écrivons à une entreprise pour proposer nos services, c'est souvent pour améliorer la qualité des données recueillies par les différents capteurs et systèmes d'information. Mais l'entreprise, qui n'a pas compris ce que nous faisons, nous répond : nous avons en interne des chercheurs qui répondent bien à notre besoin et nous n'avons pas besoin de compétences externes.

Cela montre que l'entreprise en question ne sait pas utiliser les chercheurs dont elle dispose. Ceux-ci ont vocation, non au repli sur soi, mais à participer au développement de la recherche internationale, à y apporter une contribution et à en extraire ce qui intéresse l'entreprise. La Direction Générale de l'entreprise devrait soutenir la production de données anonymisées, susceptibles d'être présentées dans des congrès, l'écriture d'articles conjoints, etc. En général, elle fait tout l'inverse, en interdisant à "ses" chercheurs toute relation avec l'extérieur.

Très vite, les chercheurs-maison souffrent d'une réduction de leur périmètre d'activité, d'autant qu'ils n'ont plus guère de stimulus. Les ingénieurs en charge des process sont réticents à leur parler des éventuels défauts de ces process et si d'aventure un chercheur a une idée, les ingénieurs refuseront de l'écouter : on ne va tout de même pas arrêter un four toutes les 20 minutes parce ce que quelqu'un, quelque part, croit avoir une idée.

Les chercheurs-maison devraient être incités, par la direction générale, à donner un avis sur l'organisation de l'entreprise en général, et pas seulement sur un petit nombre de sujets techniques qui relèvent de leur qualification. Un haut responsable nous a dit un jour que l'entreprise se plaignait de ne pas connaître ses clients. Ceci inclut le profil de ceux qui entrent et de ceux qui sortent, les raisons de leur choix, la durée de leur présence, etc. Bien souvent, on conquiert ou l'on perd un marché pour des raisons qui n'ont rien de scientifique, et l'amélioration de la qualité du produit n'y fait pas grand-chose : la question fondamentale n'est pas là.

Un élément essentiel est que l'entreprise doit s'efforcer de rendre aussi robustes que possible ses process métier, alors que, au contraire, elle s'efforce constamment de les normaliser. Si tel composé chimique entre pour 10% dans le produit final, il faudrait essayer de voir ce qui se produit si la proportion monte à 11% ou descend à 9% et faire ainsi des "plans d'expérience" qui reflètent l'essentiel des variations possibles qui, de toute façon, seront voulues par la Nature un jour ou l'autre. Comme nous l'avons déjà écrit, c'est tout le contraire d'une certification (qui procure un confort intellectuel sans véritable sécurité) et l'analyse de ces variations devrait être confiée aux chercheurs eux-mêmes, avec un mandat clair de la DG : s'il y a des changements dans les réglages, quel sera l'impact sur la sécurité, la rentabilité du process ? Faut-il rechercher ces changements, ou bien au contraire tenter de s'en préserver ? On désigne parfois sous le nom de "stress tests" l'ensemble des analyses qui sont faites pour déterminer la robustesse ; ils devraient être confiés à des chercheurs et, si possible, à des chercheurs confirmés.

Lorsqu'un industriel nous montre cinq années de recueil de données toutes les minutes, sur le fonctionnement d'un process, il s'attend à ce que nous lui disions "là, à ce moment, le réglage de tel paramètre n'était pas correct". Mais il ne réalise pas que la Nature lui a offert gracieusement un gigantesque plan d'expé-

rience, au sein duquel certains résultats peuvent être supérieurs à ce qu'il sait faire avec ses propres réglages. Il devrait les sélectionner et essayer de voir pourquoi. Là encore, des chercheurs confirmés devraient être associés à la démarche, mais c'est rarement le cas : lorsque nous menons ce travail à bien, parce qu'il nous est contractuellement demandé, nous trouvons rarement quelqu'un qui soit disposé à écouter le résultat.

Un point très important est que l'exploitation de ces résultats peut demander beaucoup de temps, largement plus que ce que prévoit notre propre contrat. Cela signifie que les chercheurs maison vont devoir travailler pendant des mois, parfois des années, en collaboration avec les ingénieurs, pour tenter d'améliorer le process industriel. Mais, comme on dit communément "le jeu en vaut la chandelle".

Revenons à nos propres activités : nous aidons nos clients à surmonter leurs difficultés, qui sont en général d'organisation ou de mauvaise compréhension des lois de la Nature. Dans la première catégorie, on trouve souvent ceci : le 1^{er} du mois, l'entreprise a commandé un disque d'acier de 20 cm de diamètre. Le 25 du mois, on lui a livré un carré de cuivre de 25 cm de côté. Il lui a fallu encore un mois pour prendre conscience des erreurs commises, et renvoyer la pièce. Bien entendu, ceci se traduit par des dérives de budget et de qualité.

Dans la seconde catégorie, on trouve une croyance systématique dans une absolue stabilité des lois de la Nature, qui empêche en particulier de prendre en compte l'apparition (pourtant fréquente) de phénomènes divers : inondations, tempêtes, grèves, accidents de toute nature, etc. La meilleure façon de s'y préparer est de constituer progressivement un système d'information, qui dira quels ennuis ont été rencontrés et quelles conséquences ils ont eues. Mais les Industriels, de manière générale, n'aiment pas cette manière de procéder, considérée comme "ringarde" et préfèrent l'achat de nouveaux équipements, présentés comme "innovants".

Prenons un exemple bien simple : en plus de trente ans d'exploitation d'une ligne porte-containers, un transporteur peut se faire une idée très nette des retards auxquels il est susceptible d'être confronté : fréquence, localisation, ampleur. Cela lui permettra de s'y préparer, au moins en corrigeant les horaires qu'il met à la disposition de ses clients. Vouloir acquérir de nouveaux équipements, permettant en particulier de détecter les variations dans la houle, les courants, les vents, etc. va coûter une fortune (il faut en équiper tous les bateaux) sans aucune certitude a priori qu'ils vont donner satisfaction.

Cette remarque a une portée très générale : vouloir comprendre un phénomène physique est la plupart du temps très difficile, parce que le phénomène dépend de quantité de paramètres qui sont mal connus. A l'inverse, vouloir faire une exploitation statistique des manifestations du phénomène est généralement assez simple, pourvu que l'historique soit suffisant.

Il y a une différence philosophique entre les deux approches : les tenants de la première sont complètement convaincus que la Terre est en perpétuelle mutation ; ils croient qu'elle est détraquée de partout et que l'homme en est la cause. Ceux de la seconde constatent que la Terre est très stable ; la moindre mutation peut prendre des milliers d'années, sans que l'homme y ait quelque responsabilité que ce soit.