



Méthodes robustes

pour l'aide à la décision

On doit souvent prendre une décision sans avoir toute l'information nécessaire. C'est vrai pour des tâches d'organisation (livraisons, plannings), de prospective (évolution des consommations, des marchés). Dans d'autres circonstances, les données existantes sont très pauvres et souvent approximatives : c'est typiquement le cas pour les préoccupations liées à l'environnement et l'épidémiologie ; les lois physiques sont en outre mal connues.

La modélisation n'est pas une science exacte, c'est un art somme toute assez récent : il a pris sa place industrielle avec les développements de l'informatique, qui ont fait croire que l'on pouvait tout modéliser, tout simuler, tout optimiser. C'est un art qui a plutôt mal commencé : en pratique, on se contente de quelques données éparses, que l'on considère comme fiables, de quelques lois empiriques, que l'on considère comme acquises, et d'un objectif unique, que l'on considère comme absolu. Après quoi, on fabrique une formule entièrement déterministe, reliant entre elles des données factices, au moyen de lois factices, pour satisfaire un objectif factice. On calcule pendant quelques heures, et tout le monde est content. Voici un exemple, issu de la sismologie :

$$\log_{10} PGA = -3.93 + 0.78 \times M_L - 1.5 \times \log_{10} R$$

où PGA (Peak Ground Acceleration) désigne la valeur maximale de la composante horizontale de l'accélération du séisme, M_L est la magnitude locale du séisme et R la distance entre le point d'observation et l'hypocentre du séisme. Cette loi (Smitt, 1998) a été établie à partir de seulement 14 séismes localisés dans les Alpes, les Pyrénées et le Massif Central. Peut-on s'en satisfaire ?

Dans la réalité, il y a des incertitudes sur les données, sur les lois et sur les objectifs. Notre métier est de concevoir des modèles, essentiellement probabilistes, tenant compte de toutes ces incertitudes. Nous les appelons "robustes", parce qu'ils incorporent les incertitudes dès la conception, et leur sortie n'est pas déterministe, mais probabiliste.

Les méthodes probabilistes permettent une première analyse grossière, qui conduit à une hiérarchie des ordres de grandeur : tel risque mérite d'être pris en compte ; tel autre est négligeable. On peut ensuite, si on le souhaite, affiner la connaissance de ceux qui émergent. Nous avons ainsi réalisé en 2005 pour le CEA Saclay une étude comparative des risques dus aux différents survols aériens et aux différents transports routiers de matières dangereuses.

Lorsqu'on met en œuvre des méthodes probabilistes, il n'est pas nécessaire que le phénomène étudié dépende du hasard : rien ne dépend jamais du hasard. Simplement, on décide qu'on "fait comme si" ; on ne recherche pas les causes intimes du phénomène. La raison essentielle est que, la plupart du temps, ces causes intimes dépendent de lois physiques mal connues, sur lesquelles les données sont rares.

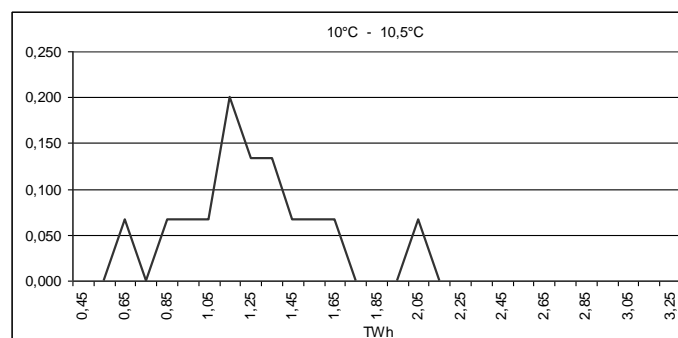
En 2005-2006, dans le cadre d'un contrat avec Veolia Environnement, Région Ouest, nous avons reconstitué les débits journaliers de 19 fleuves en Vendée : ils avaient été mesurés sur 37 ans, mais avec près de 50 % de "trous".

Nous avons utilisé des méthodes probabilistes (voir le livre de Bernard Beauzamy et Olga Zeydina : "Méthodes probabilistes pour la reconstruction de données manquantes", ref. plus bas). Cette reconstitution des débits a été grossière, mais suffisante pour répondre à la question posée : quelle est l'ampleur des pénuries d'eau l'été et comment y remédier ? Effectuer la reconstruction en se servant de modèles pluie-débit (les débits des fleuves à partir de la pluviométrie) aurait été impossible.

Un principe simple et robuste

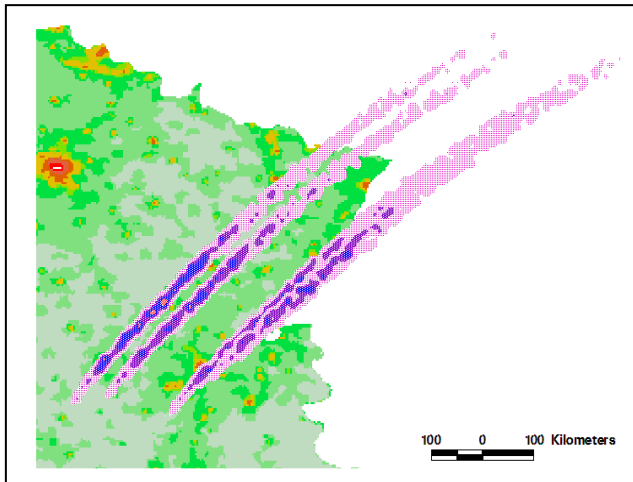
Le principe des méthodes probabilistes est simple : on considère que tout ce que l'on ne connaît pas de manière précise est régi par une loi de probabilité. Cela peut-être :

- Une donnée imprécise : on dira par exemple que la consommation de gaz, en France, est liée à la température, mais non pas de manière déterministe.



Le graphe ci-dessus est issu d'une étude que nous avons menée en 2006 pour l'Observatoire de l'Energie, Direction Générale de l'Energie et des Matières Premières : pour une température donnée (plage 10°C - 10,5°C), on trouve en abscisse une consommation (en TWh) et en ordonnée la probabilité de cette consommation : diverses consommations sont possibles, pas toutes avec la même probabilité.

- Une loi mal connue. En 2004-2005, le CNES nous avait demandé d'étudier les risques associés à la chute des débris provenant de la désintégration de satellites. Il y a très



peu de données et les lois sont mal connues. En particulier, dans la formule donnant la résistance de l'air, nous avons considéré que la densité de l'air, aux différentes altitudes, était probabiliste ; de même pour la forme du débris, sa masse, le coefficient de traînée, et même l'exposant de la vitesse : personne n'est sûr que la résistance soit proportionnelle au carré de la vitesse, pour des mobiles à 7 km/s en atmosphère raréfiée.

Le résultat n'a pas été un point de chute précis, mais une "carte probabiliste" : voici où les débris peuvent tomber, avec quelle probabilité.

On voit une telle carte sur l'image ci-dessus ; elle correspond à la chute de plusieurs objets issus du même satellite ; les zones foncées ont une probabilité plus élevée que les zones claires.

Amélioration des mesures

Les méthodes probabilistes permettent l'amélioration des mesures, grâce à des "tables de calibration" appropriées ; ce sont des tables de probabilités conditionnelles. Sachant qu'un capteur indique tant, voici la probabilité que l'erreur soit de tant. Ces lois sont différentes d'un bout à l'autre de la gamme de mesure : un capteur n'a pas partout la même précision (cela s'appelle le facteur d'échelle). Nous avons employé ces méthodes d'abord dans le cadre de contrats avec la Délégation Générale pour l'Armement (Service des Programmes de Missiles Tactiques), puis avec l'IRSN, depuis 2003 : amélioration des mesures de matières nucléaires, et en 2012-2016 avec l'IFSTTAR : amélioration du positionnement GPS en environnement urbain.

Différences avec les statistiques

Les statistiques permettent de multiples traitements des données : ajustements, régressions, tests, mais toujours en faisant des hypothèses sur les lois des échantillons recueillis. Or, dans les situations où nous travaillons, ces lois ne sont pas connues et introduire des lois factices est tout aussi inacceptable qu'introduire des données factices.

Les méthodes probabilistes n'ont pas ce handicap, puisque notre travail est tout d'abord de constituer la loi de probabilité : voir l'exemple de la consommation de gaz ci-dessus. Nous partons d'une loi uniforme sur un intervalle (si on ne sait rien), et nous l'enrichissons petit à petit en fonction du retour d'expérience.

On peut dire grossièrement que les statistiques sont une forme raffinée des probabilités (lorsque la loi est connue) et que les probabilités sont une forme préliminaire des statistiques (avant que la loi soit connue). Bien entendu, si un historique abondant existe (exemple : ventes d'un produit), il n'y a aucune raison de ne pas utiliser les outils statistiques usuels.

Modélisation robuste

Les méthodes probabilistes sont un élément essentiel de notre programme de recherche "modélisation robuste", développé en partenariat avec environ 70 institutions (entreprises et universités), françaises et étrangères (voir <http://scmsa.eu/robust.htm> pour une description complète du programme).

Une méthode "robuste" prend en compte, dès le début du programme, les incertitudes sur les lois et sur les données, mais aussi sur les objectifs. L'expérience nous montre que les programmes industriels n'ont jamais un seul objectif : il y a le court terme et le long terme, les problèmes de production, de maintenance, d'organisation, etc. L'ensemble est toujours complexe, et ramener la question posée à un simple optimum (souvent sur le coût) est réducteur et conduit généralement à des solutions inadaptées.

Nous considérons donc qu'il faut abandonner la recherche d'un optimum et traiter l'ensemble des objectifs comme des contraintes ; par exemple : "dépenser 10 % de moins que l'année dernière", "réduire les retards de 5 %", etc. Ceci doit être obtenu rapidement : c'est ce que nous appelons une "*Quick Acceptable Solution*". Une fois celle-ci trouvée, rien n'empêche de recommencer, en affinant les contraintes. Le décideur préférera avoir une solution grossière, mais rapide et robuste, permettant de savoir que telle configuration est satisfaisante, et que telle autre l'est moins.

Hiérarchisation de paramètres

Etant donné un process industriel, qui dépend évidemment d'un très grand nombre de paramètres (des températures, des pressions, des compositions chimiques, etc.), une question revient souvent : "hiérarchiser" les paramètres, en fonction de leur influence sur une variable de sortie (la qualité du process, quelle qu'en soit la définition). Nos méthodes permettent cette hiérarchisation, quel que soit le nombre des données : voir la fiche spécifique http://scmsa.eu/fiches/SCM_Hierarchisation.pdf

Un exemple très simple de mise en œuvre a été réalisé par la SCM pour l'Agence de l'Eau Artois-Picardie en 2008 et est disponible : http://scmsa.eu/archives/SCM_AEAP_2008_12_01.pdf.

Recherche de zones à risque

Ceci vient en complément du paragraphe précédent. Etant donné un process dépendant d'un très grand nombre de paramètres, il est possible de déterminer les configurations de l'ensemble des paramètres qui conduiront à une "situation à risque" pour la variable de sortie du process, par exemple une qualité inférieure à un seuil fixé. Ceci se fait uniquement à partir des informations recueillies, et sans introduire d'hypothèse factice. La méthode est décrite au chapitre XIII du livre ci-dessus.

La "Méthode" d'Archimède

Considérée par Archimède comme son chef d'œuvre, elle a été perdue pendant plus de 2000 ans. Elle consiste à comparer l'information inconnue à une information connue, générée pour la circonstance, et cette comparaison est toujours robuste. Les perspectives sont étonnantes ; voir le livre "Archimedes Modern Works", ref. ci-dessous.

Nos réalisations

Livres :

1. Bernard Beauzamy : Méthodes probabilistes pour l'étude des phénomènes réels, ISBN : 2-9521458-0-6, Editions de la SCM, mars 2004. Seconde édition, juin 2016.
2. Bernard Beauzamy et Olga Zeydina : Méthodes probabilistes pour la reconstruction de données manquantes, ISBN : 2-9521458-2-2, Editions de la SCM, avril 2007.
3. Bernard Beauzamy : Nouvelles méthodes probabilistes pour l'évaluation des risques. ISBN : 978-2-9521458-4-8, ISSN : 1767-1175, Editions de la SCM, avril 2010.
4. Bernard Beauzamy : Archimedes Modern Works. ISBN 978-2-9521458-7-9, ISSN 1767-1175, Editions de la SCM, août 2012. .
5. Olga Zeydina et Bernard Beauzamy : Probabilistic Information Transfer. ISBN: 978-2-9521458-6-2, ISSN: 1767-1175. Editions de la SCM, mai 2013.

Références récentes :

- Agence Européenne pour l'Environnement, 2006-2015 : contrat cadre "méthodes probabilistes pour la qualité de l'eau en Europe"
- IRSN, 2006-2009 : Méthodes probabilistes pour la Sécurité Nucléaire : mise en place de la méthode de l'Hypersurface Probabiliste (méthode créée par la SCM)
- Observatoire de l'Energie, DGEMP, 2006 : Etude probabiliste concernant la sécurité des approvisionnements en gaz pour la France
- Agence Nationale pour la Gestion des Déchets Radioactifs (ANDRA), 2007-2012 : Analyse probabiliste des modèles de transferts de radionucléides
- Délégation à la Sécurité Nucléaire et à la radioprotection pour les activités et installations intéressant la Défense (DSND), 2007-2008 : calculs de sûreté pour les systèmes d'armes nucléaires
- Agence de l'Eau Artois-Picardie, 2008 : Etude probabiliste concernant la qualité des eaux de rivière et caractérisation des situations de bonne qualité

- CEA, Direction de l'Energie Nucléaire, Département Modélisation des Systèmes et Structures : méthodes probabilistes en séismologie et en épidémiologie, 2007-2008
- Réseau de Transport d'Electricité, 2008 : Méthodologie probabiliste relative à une décision d'investissement
- Caisse Centrale de Réassurance, 2009 : Méthodes probabilistes pour l'étude du débit des rivières ; 2010-2011 : méthodes probabilistes pour l'évaluation des phénomènes extrêmes "hauteurs des marées"
- SNECMA Propulsion Solide, 2009-2010 : Méthodes probabilistes pour la fiabilité
- Nuclear Energy Agency, 2010, 2012, 2014, 2015, 2016, 2017 : Détection de données aberrantes dans les bases de données
- ArcelorMittal, 2011-2012 : Méthodes probabilistes pour la qualité d'un usinage
- Air Liquide, 2011 : Construction d'un "indice de proximité" entre pipelines
- IFSTTAR, 2012-2015 (dans le cadre d'un contrat avec le Ministère de l'Ecologie) : Méthodes probabilistes pour l'amélioration du positionnement GPS en environnement urbain
- Areva, 2013 : Hiérarchisation des paramètres intervenant dans un process industriel.
- DCNS, site d'Indret, 2013 : Soudage par Faisceau d'Electrons ; hiérarchisation de l'influence des paramètres par analyse probabiliste
- COSEA (Ligne à Grande Vitesse Sud Est Atlantique), 2013 : détermination des durées de retour pour les crues extrêmes sur la Vienne et la Creuse
- Coop de France déshydratation, 2013 : Réalisation d'un outil d'analyse des Composés Organiques Volatils Non Méthaniques
- IRSN, 2014-2015 : Analyse critique du dimensionnement du réseau de mesure "Tele-ray"
- Direction Générale Energie-Climat (MEDD), Bureau Qualité de l'Air, 2015 : Relations probabilistes entre trafic et émissions de polluants sur le boulevard périphérique autour de Paris
- Solétanche-Bachy, 2015 : Hiérarchisation des paramètres influant sur la déformation d'un ouvrage d'art
- Telcap, 2015 : Prévisions de trafic en téléphonie
- Carrefour, 2016 : Hiérarchisation des paramètres intervenant sur la vente de jouets
- COSEA, 2016 : Etude probabiliste relative à la turbidité de l'eau
- Taxis G7, 2016 : Analyse probabiliste de bases de données et hiérarchisation de paramètres
- L'Oréal, 2016 : Etude des données disponibles pour les accidents de la route entre le domicile et le lieu de travail
- SNCF Réseau, 2016 : Appui scientifique pour l'analyse des scénarios relatifs à une ligne nouvelle
- ANDRA, 2016, 2017, 2018, 2019 : Optimisation de la position des capteurs pour la surveillance d'un site de stockage
- COSEA, 2016 et 2017 : Etudes statistiques relatives à la turbidité de l'eau
- RATP, 2016-2017 et 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
- Monceau Assurances, 2017 : Amélioration de la politique commerciale
- Monceau Assurances, 2017 : Modélisation des catastrophes naturelles et de leur impact sur le portefeuille
- Syndicat des Eaux d'Ile de France, 2017 : appui méthodologique
- Taxis G7, 2017 : Etudes liées à la logistique
- Carrefour/Bazar, 2017 : Amélioration d'un outil de recommandation pour les Plans d'Achat
- SNCF Mobilités, 2018 : Etude des déplacements au voisinage du bipôle Nanterre-La Défense
- RTE, 2018 : Analyse des maintenances pour une famille d'équipements
- Atlandes (autoroute A63) : Etude du comportement des véhicules sur les bretelles de sortie de l'autoroute
- RATP, 2018 : Etude probabiliste des efforts dus aux tractions et freinages des matériels roulants
- Eramet, 2018-2019 : Amélioration d'un process industriel
- BRGM, 2018-2019 : Outils probabilistes relatifs à la pollution des sols
- RATP, 2018-2019 : Etude probabiliste des efforts dus aux tractions et freinages des matériels roulants sur la structure des viaducs
- SARP Industries, 2019 : Hiérarchisation des paramètres intervenant dans un process industriel
- Industriel, 2019 : Amélioration d'un process de fabrication
- Orano Mining, 2019 : Hiérarchisation de paramètres intervenant dans un process industriel
- CEA, 2019 : Hiérarchisation de paramètres
- Groupe Atlantic, 2019 : Analyse probabiliste des appels au Service Après-Vente