La surveillance de risques émergents à partir d'algorithmes

Nécessité d'une participation active de la société

Florence CARRÉ(1), Jean-Yves CHATELIER(2), David GRANDGIRARD(3)

(1) Chargée de mission « multi-risques », Institut national de l'environnement et des risques (Ineris),
(2) Chargé de mission « intelligence artificielle », Institut national de l'environnement et des risques (Ineris),
(3) Enseignant-Chercheur en agroécologie, UniLasalle

Résumé : Le développement de notre société et les changements climatiques font apparaître des risques nouveaux qu'il faut pouvoir anticiper pour éviter des dommages sanitaires, environnementaux et financiers. Ce chapitre traite des différentes composantes du risque à prendre en compte dans l'anticipation. Les algorithmes d'identification, de traitement et de surveillance automatique de ces composantes sont, dans un second point, décrits en détail. Les limites des différents algorithmes sont évoqués dans une troisième partie, faisant apparaître le réel besoin d'une complémentarité entre algorithmes et implication sociétale dans l'exercice d'identification et surveillance des risques émergents.

Mots-clés : risque émergent, fouilles de données, identification, surveillance, société

Le développement de notre société (accroissement démographique, déploiement d'innovations technologiques…) et les changements climatiques font apparaître des risques nouveaux. La prise de conscience de ces risques par les gestionnaires des territoires et d'infrastructures, par les décideurs ou la société, apparaît parfois très tard, induisant des effets dommageables pour l'environnement et la santé ainsi que des pertes financières importantes. L'anticipation de ces risques et leur surveillance permet, pour les décideurs et gestionnaires du territoire, de les maîtriser le plus en amont possible, et pour les assureurs, d'éviter des pertes financières faisant suite aux catastrophes.

Aujourd'hui les outils de l'intelligence artificielle, notamment les algorithmes, permettent d'assister gestionnaires, décideurs et assureurs dans l'identification et la sur-

veillance des risques émergents. Ce chapitre présente dans un premier temps le concept de risque émergent, dans un deuxième temps, les algorithmes utilisés pour identifier et surveiller les sources potentielles de risques, et dans un troisième temps, la complémentarité entre algorithmes et société au regard des limites des algorithmes.

1 Définition d'un risque émergent

1.1 Définition du risque et des notions associées

Le risque est généralement défini comme la probabilité d'un événement non espéré ou comme sa probabilité multipliée par ses conséquences.[1]

L'événement non espéré, souvent appelé « aléa » ou « phénomène dangereux » peut être d'ordre naturel comme un tsunami, une sécheresse, une inondation, ou d'ordre technologique comme un court-circuit électrique dans un procédé industriel ou la présence d'une substance chimique nocive pour l'environnement et la santé. Dans le cas des effets cascade ou domino, un événement non espéré peut engendrer des conséquences qui seront elles-mêmes considérées comme un événement non espéré,[2] comme cela a été le cas à Fukushima en 2011 : un séisme d'une magnitude record de 9,0 (événement naturel 1) a généré un violent tsunami (événement naturel 2) sur les côtes nord-est du Japon entraînant des vagues d'une dizaine de mètres de haut (événement naturel 3) noyant 18500 personnes et endommageant la centrale nucléaire de Fukushima Daiichi (événement technologique 4), provoquant la fusion de ses réacteurs (événement technologique 5), des explosions en chaîne (événement technologique 6) et libérant ainsi une large quantité de césium 137 radioactif (événement technologique 7) dont 80% sont encore présents aujourd'hui[3] et impactent l'environnement et la santé.

Les conséquences sont évaluées au regard des enjeux ciblés sur un ou plusieurs territoires. Ces enjeux peuvent être ou plusieurs procédés technologiques, une ou plusieurs organisations, un ou plusieurs écosystèmes (incluant les hommes et l'environnement). Les conséquences peuvent se mesurer en pertes financières ou pertes de fonctionnalités du processus technologique, organisationnel et/ou des écosystèmes (pourcentage de fonctions techniques, organisationnelles ou (éco)physiologiques non assurées, nombre de morts, nombre et type de pathologies ou dommages corporels, perte de biodiversité…). Les conséquences dépendent de la capacité de l'enjeu à faire face à l'intensité de l'événement non espéré. Identifier un événement non espéré permet de s'y préparer et donc de mieux y faire face, d'où la nécessité de développer une culture du risque au sein des sociétés souvent exposées à des événements non espérés.

La probabilité de l'événement non espéré est calculée au regard des enjeux considérés. Il s'agit de la probabilité d'exposition de l'enjeu à l'événement non espéré.

[1] Boholm, M. (2019). *Risk and Quantification : A Linguistic Study. Risk Analysis*, *39*(6), p. 1243-1261. https://doi.org/10.1111/risa.13258

[2] Provitolo, D. (2005). Un exemple d'effets de dominos : La panique dans les catastrophes urbaines. *Cybergeo : European Journal of Geography*. https://doi.org/10.4000/cybergeo.2991

[3] https://www.notre-planete.info/actualites/4650-accident-nucleaire-Fukushima-pollution-2020

Published in : DELTORN, Jean-Marc ; PICHENOT, Evelyne (dir.) (dir.), *Algorithmes et Société*, Editions des archives contemporaines, ISBN : 9782813003928, doi : 10.17184/eac.9782813003928

1.2 Qu'est-ce qu'un risque émergent ?

Selon l'approche anthropocentrée de Flage et Aven[4] un risque est dit « émergent » lorsque les connaissances de base sont faibles mais contiennent des indications / croyances justifiées selon lesquelles un nouveau type d'événement non espéré (ou nouveau au regard de l'enjeu) pourrait se produire dans le futur et avoir potentiellement des conséquences graves sur l'homme. La faible connaissance de base, entre autres, entraîne des difficultés à spécifier les conséquences, l'intensité et la probabilité de l'événement non espéré.

Plus globalement, le International Risk Governance Council[5] définit les risques émergents comme étant de nouveaux risques ou des risques familiers qui apparaissent dans des conditions nouvelles ou non familières. Il est toutefois important de compléter cette définition par le fait qu'il y a un manque de connaissances pour bien anticiper, maîtriser et gérer un risque émergent, autrement dit, que de nombreuses incertitudes sont associées à l'événement non espéré et à la façon dont l'enjeu réagira face à celui-ci. Les prévisions relatives aux risques émergents portent généralement sur des cadres temporels s'échelonnant entre 5 et 20 ans.

Parmi les nouveaux risques, les assureurs évoquent l'antibiorésistance pouvant augmenter dans de proches horizons le nombre de décès liés à des infections bactériennes ne pouvant être traitées.[6] Par ailleurs, l'augmentation des températures dans les régions de pergélisols pourrait entraîner la fonte du permafrost et libérer des virus jusqu'ici inconnus ou connus mais apparaissant dans de nouvelles conditions comme la variole.[7]

Pour cette même catégorie de risques familiers apparaissant dans des conditions nouvelles, les nouvelles formes d'énergie comme les batteries électriques, les piles à combustibles ou la méthanisation peuvent être identifiées. En effet, les industries ont aujourd'hui de bonnes connaissances des événements non espérés associés à ces formes d'énergie tels que des phénomènes d'explosion. Les barrières de sécurité techniques et organisationnelles pour y faire face sont donc bien maîtrisées, garanties par des protocoles normés. Cependant, dans le cadre de la transition énergétique, ces technologies sont de plus en plus appliquées par les particuliers qui ont moins de connaissances sur les événements non espérés et les protocoles pour y faire face. La gestion des risques par les particuliers sera donc moins performante que par les industriels.[8] Ce genre d'exemples concerne également les imprimantes 3D de plus en plus présentes dans les milieux domestiques. Celles-ci peuvent relarguer bon nombre de nanoparticules

[4] Flage, R., & Aven, T. (2015). Emerging risk – Conceptual definition and a relation to black swan type of events. *Reliability Engineering & System Safety*, *144*, 61-67. https://doi.org/10.1016/j.ress.2015.07.008,

[5] International Risk Governance Council. IRGC. (2015). Guidelines for the Governance of Emerging Risks. *Guidelines for the Governance of Emerging Risks*. https://irgc.org/risk-governance/emerging-risk/a-protocol-for-dealing-with-emerging-risks/

[6] Hayward, T., & Hutchinson, E. (2019). Emerging risks in insurance : Antibiotic resistance. *Antibiotic Resistance*, 3.

[7] Cavicchioli, L. (2016, novembre 23). *Permafrost : Et s'il libérait des armées de virus… – Science & Vie*. https://www.science-et-vie.com/archives/permafrost-et-s-il-liberait-des-armees-de-virus-21237

[8] Carre, F., Adam, K., Cointe, R., Gentilhomme, O., Guillaume, M. B. et O., Gabillet, P., Gendron, C., Jakubowicz, M., Leber, J.-Y., Mocanu, S., Moretti, I., Vidalenc, E., & Villeneuve, J. (2020). *Les leviers d'action pour un mix énergétique propre et sûr au service de la transition énergétique dans les territoires*. http://www.annales.org/re/2020/resumes/avril/20-re-resum-FR-AN-avril-2020.html

Published in : DELTORN, Jean-Marc ; PICHENOT, Evelyne (dir.) (dir.), *Algorithmes et Société*, Editions des archives contemporaines, ISBN : 9782813003928, doi : 10.17184/eac.9782813003928

pouvant engendrer des fibroses pulmonaires.[9] Ces émissions sont plus maîtrisées par les services d'Hygiène, Santé, Environnement des entreprises que par les particuliers.

Parmi les risques pour lesquels peu de connaissances étaient disponibles jusque-là et dont les conséquences sur l'environnement et la santé sont importantes, les microplastiques sont un bel exemple. En effet, l'imprégnation environnementale des microplastiques est importante.[10] Les effets sur la santé et l'environnement sont, pour certains, de la perturbation endocrinienne intergénérationnelle pouvant conduire à l'extinction d'espèces.[11] Les microplastiques peuvent donc être catégorisés comme risques émergents.

1.3 Les besoins de surveillance

Les exemples cités dans la section précédente démontrent qu'il est important de surveiller : (a) l'évolution temporelle des sources d'événements non espérés comme les innovations technologiques sources de dangers (explosion, toxicité…) ou les événements climatiques extrêmes (b) le contexte spatio-temporel de leur déploiement reflétant (b1) le degré d'exposition des différents enjeux aux sources d'événements non espérés et (b2) la capacité de résilience des enjeux aux événements. Le contexte spatio-temporel de déploiement dépend en majeure partie des comportements humains, de leur propension à accepter et diffuser une innovation et à se protéger des dangers. Cette propension dépend elle-même des valeurs sociales accordées aux innovations technologiques[12] et de la perception sociétale des risques.[13] Seul le point (a) est traité dans la suite du chapitre.

2 Les algorithmes pour surveiller l'évolution temporelle des sources émergentes d'événements non espérés

Analyser l'évolution temporelle des sources émergentes d'événements non espérés requiert d'identifier tout type d'émergences, d'analyser si elles sont sources de dangers puis de les surveiller.

2.1 Algorithmes pour identifier des émergences

L'objectif est d'identifier des thèmes ou concepts relatifs à des événements sociotechniques (innovations technologiques, accidents technologiques) ou

[9] Zhang, Q., Pardo, M., Rudich, Y., Kaplan-Ashiri, I., Wong, J. P. S., Davis, A. Y., Black, M. S., & Weber, R. J. (2019). Chemical Composition and Toxicity of Particles Emitted from a Consumer-Level 3D Printer Using Various Materials. *Environmental Science & Technology*, *53*(20), 12054-12061. https://doi.org/10.1021/acs.est.9b04168

[10] Zhang, S., Wang, J., Liu, X., Qu, F., Wang, X., Wang, X., Li, Y., & Sun, Y. (2019). Microplastics in the environment : A review of analytical methods, distribution, and biological effects. *TrAC Trends in Analytical Chemistry*, *111*, 62-72. https://doi.org/10.1016/j.trac.2018.12.002

[11] Chen, Q., Allgeier, A., Yin, D., & Hollert, H. (2019). Leaching of endocrine disrupting chemicals from marine microplastics and mesoplastics under common life stress conditions. *Environment International*, *130*, 104938. https://doi.org/10.1016/j.envint.2019.104938

[12] Deffuant, G., Huet, S., & Amblard, F. (2005). An Individual-Based Model of Innovation Diffusion Mixing Social Value and Individual Benefit. *American Journal of Sociology*, *110*(4), 1041-1069. https://doi.org/10.1086/430220

[13] Mohtar, S., & Abbas, M. (2020). Consumer resistance to innovation due to perceived risk : relationship between perceived risk and consumer resistances to Innovation. *Journal of Technology and Operations Management*, *10*(1), 1-13. https://doi.org/10.32890/jtom2015.10.1.10116

Published in : DELTORN, Jean-Marc ; PICHENOT, Evelyne (dir.) (dir.), *Algorithmes et Société*, Editions des archives contemporaines, ISBN : 9782813003928, doi : 10.17184/eac.9782813003928

socio-environnementaux (émeutes, incidences de maladies, phénomènes climatiques extrêmes) qui vont croître de façon significative au fil du temps et qui doivent donc alerter.

Pour ce faire, six étapes, présentées en Figure 1, sont nécessaires.

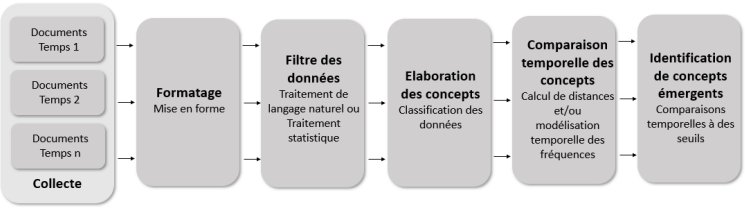


Fig. 1 : Les différentes étapes algorithmiques associées à l'identification des sources de dangers

La première étape concerne la collecte des données à chaque pas de temps d'analyse. Si l'on recherche de nouvelles innovations, la fréquence d'analyse sera plutôt d'ordre annuelle alors que pour une recherche de préoccupations sociétales, la fréquence sera plutôt d'ordre mensuelle voire hebdomadaire. Dépendant de l'objectif visé, les documents peuvent concerner des rapports techniques ou articles de recherche d'un ou plusieurs journaux spécifiques au sujet global recherché, ou des tweets ou posts Facebook pour analyser les sujets de préoccupation sociétale. Dans tous les cas, cette étape requiert une sélection des sources de documents, tweets ou posts à partir de mots-clés.

Une fois les documents textuels collectés, il est nécessaire de les mettre au format d'analyse requis par les outils de traitements statistiques utilisés lors des étapes ultérieures (étape 2). Lorsque effectué pour la première fois, ce travail requiert en général un tiers du temps de l'ensemble des traitements, d'où l'utilité d'automatiser cette procédure.

Les données sont ensuite filtrées (étape 3), i.e. nettoyées de termes non pertinents, souvent les plus représentés au sein des différents documents (e.g. auxiliaires, vocabulaire fonctionnel). Deux types de traitements sont en général utilisés : (a) des traitements de langage naturel qui utilisent des dictionnaires existants, remplacent un mot par son synonyme et/ou comparent la fréquence relative des mots du corpus de documents à celle de dictionnaires comme Wikipedia afin de ne sélectionner que les mots de fréquence plus élevée dans le dictionnaire utilisé,[14] (b) des traitements statistiques comme la procédure fréquence de termes – fréquence inverse de documents (TF-IDF) qui permet de pondérer la fréquence d'un terme d'un document en fonction de sa représentation au sein de l'ensemble des documents analysés puis de ne sélectionner que les termes de fréquence pondérée importante.[15]

[14]Raveh, E., Ofek, Y., Bekkerman, R., & Cohen, H. (2020). Applying Big Data visualization to detect trends in 30 years of performance reports : *Evaluation*. https://doi.org/10.1177/1356389020905322

[15]Aiello, L. M., Petkos, G., Martin, C., Corney, D., Papadopoulos, S., Skraba, R., Göker, A., Kompatsiaris, I., & Jaimes, A. (2013). Sensing Trending Topics in Twitter. *IEEE Transactions on Multimedia*, *15*(6), 1268-1282. https://doi.org/10.1109/TMM.2013.2265080

Published in : DELTORN, Jean-Marc ; PICHENOT, Evelyne (dir.) (dir.), *Algorithmes et Société*, Editions des archives contemporaines, ISBN : 9782813003928, doi : 10.17184/eac.9782813003928

L'élaboration des concepts pour chaque pas de temps (étape 4) consiste le plus souvent à analyser, au sein d'un gramme i.e. d'un groupe de mots de taille fixe ou variable (phrase nominale), la fréquence de termes sélectionnés à l'étape 3 puis à classer les grammes en fonction de la fréquence des termes qui les composent afin d'en déduire des concepts thématiques pouvant représenter une innovation ou une situation.[16]

Lors des étapes 5 et 6, les concepts sont comparés entre pas de temps à partir de calculs de distance de similarité basés sur les fréquences des termes les composant. Les distances au-dessus d'un certain seuil, déterminé arbitrairement ou par des algorithmes de réseaux neuronaux ou de *machine learning*, indiquent des concepts marginaux, innovations ou situations potentiellement émergentes.[17] Les autres concepts, non-marginaux, peuvent être sujets à un accroissement temporel important et donc représenter potentiellement un concept émergent déjà émergé. Dans ce cas, les fréquences sont modélisées temporellement et la significativité du taux d'accroissement temporel est évaluée par des tests statistiques.[18] Les concepts dont le taux d'accroissement temporel est positif et statistiquement significatif sont considérés comme émergents.

2.2 Algorithmes pour sélectionner les sources de dangers parmi les concepts émergents

Tous les concepts émergents ne sont pas forcément porteurs de dangers. La sélection des concepts dangereux était jusque-là établie le plus souvent par des experts à partir du type de procédés de fabrication ou des substances et produits sous-jacents à une innovation technologique ou d'un type d'événements ou situations. Néanmoins, elle est de plus en plus assistée par des algorithmes se basant sur l'apprentissage à partir de données existantes.

Les types de procédés, substances et produits considérés comme dangereux sont ceux pouvant provoquer des phénomènes accidentels tels qu'incendies, explosions, dispersions atmosphériques et transferts dans les différents milieux dont les effets lors d'ingestion, inhalation ou réaction de contact sont toxiques et écotoxiques.

Dans le cas des incendies, l'inflammabilité d'une substance ou d'un mélange de substances, peut être prédite à partir des modèles QSPR (*Quantitative Structure Property Relationship*) permettant de calculer le point d'éclair i.e. la température la plus basse à laquelle ce liquide libère assez de vapeur pour s'enflammer en surface. Ces modèles combinent algorithme génétique et régression linéaire multiple.[19] Un produit/substance sera d'autant plus évalué comme porteur de danger que son point d'éclair sera faible.

[16] Pottenger, W. M., Kim, Y.-B., & Meling, D. D. (2001). HDDI™ : Hierarchical Distributed Dynamic Indexing. In R. L. Grossman, C. Kamath, P. Kegelmeyer, V. Kumar, & R. R. Namburu (Éds.), *Data Mining for Scientific and Engineering Applications* (p. 319-333). Springer US. https://doi.org/10.1007/978-1-4615-1733-7_18

[17] Duvvuru, A., Kamarthi, S., & Sultornsanee, S. (2012). Undercovering research trends : Network analysis of keywords in scholarly articles. *2012 Ninth International Conference on Computer Science and Software Engineering (JCSSE)*, 265-270. https://doi.org/10.1109/JCSSE.2012.6261963

[18] Färber, M., & Jatowt, A. (2019). Finding Temporal Trends of Scientific Concepts. *BIR@ ECIR*, 132–139.

[19] Fayet, G., & Rotureau, P. (2019). New QSPR Models to Predict the Flammability of Binary Liquid Mixtures. *Molecular Informatics*, *38* (8-9), 1800122. https://doi.org/10.1002/minf.201800122

Published in : DELTORN, Jean-Marc ; PICHENOT, Evelyne (dir.) (dir.), *Algorithmes et Société*, Editions des archives contemporaines, ISBN : 9782813003928, doi : 10.17184/eac.9782813003928

L'(éco)toxicité des substances peut être évaluée à partir de modèles QSAR (*Quantitative Structure Activity Relationship*) se basant sur la structure moléculaire de composés chimiques de même famille.[20] Aujourd'hui, les QSAR commencent à intégrer les mélanges de substances. Une substance évaluée comme (éco)toxique sera considérée comme porteuse de dangers.

La dispersion atmosphérique de substances potentiellement (éco)toxiques s'effectue à partir de modèles de dynamique computationnelle des fluides (CFD) alliant analyse numérique et algorithmes informatiques pour simuler, analyser et résoudre des problèmes d'écoulement de fluide.[21] Les substances (éco)toxiques de forte dispersion atmosphérique seront donc considérées comme porteuses de danger.

Les situations ne faisant pas directement référence à des procédés ou substances issues d'innovation technologique sont par exemple l'apparition d'événements climatiques extrêmes, de nouveaux pathogènes ou des pollutions émergentes dans les différents milieux. L'évaluation de leur dangerosité s'effectue en simulant leur impact sur les milieux (homme inclus) et les systèmes techniques à partir, pour l'impact des substances sur l'homme et l'environnement, de modèles (éco)toxicologiques existants comme les modèles pharmaco-cinétiques[22] ou, pour l'impact d'aléas naturels sur les infrastructures technologiques, de modèles de courbes de fragilité.[23]

Tous ces modèles requièrent des données d'entrée exhaustives. Si ce n'est pas le cas, les résultats seront incertains ; il sera alors nécessaire d'effectuer des expérimentations pour tester la réelle dangerosité des produits, substances, aléas et alimenter les bases de données existantes.

2.3 Surveillance de la diffusion des dangers émergents

Une fois les dangers émergents identifiés, il faut pouvoir suivre leur diffusion dans le temps pour bien s'y préparer. Pour ce faire, des algorithmes de veille de l'actualité scientifique, normative, des réseaux sociaux, de surveillance des marchés ou des algorithmes d'analyse de données des milieux (homme inclus) en temps réel sont utilisés.

Les algorithmes de veille sont programmés à partir d'équations de recherche intégrant le danger émergent à surveiller afin d'interroger bases de données et moteurs de recherche. Les données sont ensuite récupérées des pages Web concernées.[24] Elles sont alors filtrées afin de ne conserver que l'information pertinente, puis analysées tem-

[20] Mombelli, E. (2008). An evaluation of the predictive ability of the QSAR software packages, DEREK, HAZARDEXPERT and TOPKAT, to describe chemically-induced skin irritation. *Alternatives to Laboratory Animals : ATLA*, *36*(1), 15-24. https://doi.org/10.1177/026119290803600104

[21] Batt, R., Gant, S., Lacome, J. M., Truchot, B., & Tucker, H. (2018). CFD modelling of dispersion in neutrally and stably-stratified atmospheric boundary layers : Results for Prairie Grass and Thorney Island. *International Journal of Environment and Pollution*, *63*(1/2), 1. https://doi.org/10.1504/IJEP.2018.093026

[22] Brochot, C., Willemin, M.-É., & Zeman, F. (2014). *Chapitre 13. La modélisation toxico-pharmaco cinétique à fondement physiologique : Son rôle en évaluation du risque et en pharmacologie*. Éditions Matériologiques. https://www.cairn.info/modeliser-et-simuler--9782919694730-page-453.htm

[23] Mebarki, A., Willot, A., Jerez, S., Reimeringer, M., & Prod'Hommetan, G. (2014). Vulnerability and resilience under effects of tsunamis : Case of industrial plants. *Procedia Engineering*, *84*, 116-121. https://doi.org/10.1016/j.proeng.2014.10.520

[24] Thomas, A., Bonny, P., Deschamps, C., Cherbonnier, M., Cottave, M., Beauvieux, A., & François, P. (2008). Les outils de la veille. *Documentaliste-Sciences de l'Information*, *vol. 45*(4), 46-57. https://www.cairn.info/revue-documentaliste-sciences-de-l-information-2008-4-page-46.htm

Published in : DELTORN, Jean-Marc ; PICHENOT, Evelyne (dir.) (dir.), *Algorithmes et Société*, Editions des archives contemporaines, ISBN : 9782813003928, doi : 10.17184/eac.9782813003928

porellement au regard du nombre d'occurrences et du contexte d'apparition selon les méthodes d'identification des émergences décrites en 2.1. Des occurrences qui tendent à croître et l'apparition de nouveaux contextes laissent supposer un accroissement des dangers et la nécessité d'une acculturation aux risques qui en découlent.

Les polluants émergents qui font l'objet d'analyses en temps réel peuvent être de nature chimique (substances pharmaceutiques, pesticides, plastifiants…) ou de nature biologique (virus, prion, ADN de bactéries résistantes aux antibiotiques, toxines…). La mesure des polluants peut être directe (concentration des polluants concernés) ou indirecte (mesure de l'effet global de la substance ou d'un mélange sur un modèle biologique). Quelle que soit la méthode d'acquisition de la mesure, l'identification du polluant est effectuée à partir d'algorithmes de comparaison à des données existantes.[25] Plus les données existantes sont exhaustives, plus la détermination analytique du polluant sera de bonne qualité. L'analyse temporelle est ensuite effectuée à partir des algorithmes évoqués en 2.1. Il est par ailleurs important de noter que ces substances ne sont pas stables dans le temps et les différents milieux. Le suivi temporel de ces substances doit donc intégrer les molécules issues de leurs transformations.

3 Discussion : limites des algorithmes et nécessité d'une implication active de la société pour la surveillance des risques émergents

3.1 Les limites des algorithmes

Les algorithmes ont comme limites de ne pouvoir être appliqués qu'avec une grande quantité de données afin d'obtenir des modèles statistiquement robustes. De même, les règles de décisions associées à ces modèles sont également basées sur des seuils établis à partir d'algorithmes de *machine learning* ou de réseaux neuronaux qui ne sont significatifs que lorsque les données d'apprentissage sont exhaustives.

Toutes ces données proviennent de sources différentes. Il faut donc pouvoir harmoniser non seulement les protocoles d'acquisition mais également les formats de dépôts qui doivent être accessibles à tout utilisateur ou producteur de données.

De nombreuses données sont aujourd'hui utilisées sans avoir d'information sur leur qualité. Celle-ci est évaluée à partir de 6 critères : degré de représentation temporelle, exhaustivité, unicité, validité (conformité au format de dépôt), cohérence vis-à-vis d'une définition, précision (degré de description correcte vis-à-vis de la réalité)[26] Ainsi, les algorithmes de traitement des données produisent des résultats le plus souvent erronés. Il faut donc développer un sens critique vis-à-vis des algorithmes sur lesquels se basent les décisions. Ce sens critique est d'autant plus nécessaire que les algorithmes sont souvent présentés sous forme de « boîtes noires » dans des logiciels qui n'offrent pas la possibilité de contrôler la façon dont ils ont été développés.

[25] Agüera, A., Martínez Bueno, M. J., & Fernández-Alba, A. R. (2013). New trends in the analytical determination of emerging contaminants and their transformation products in environmental waters. *Environmental Science and Pollution Research*, *20*(6), 3496-3515. https://doi.org/10.1007/s11356-013-1586-0

[26] Satterfield, A. (CDC/ONDIEH/NCBDDD) (CTR). (s. d.). The Six Dimensions of EHDI Data Quality Assessment. 4. https://www.cdc.gov/ncbddd/hearingloss/documents/dataqualityworksheet.pdf.

Published in : DELTORN, Jean-Marc ; PICHENOT, Evelyne (dir.) (dir.), *Algorithmes et Société*, Editions des archives contemporaines, ISBN : 9782813003928, doi : 10.17184/eac.9782813003928

Par ailleurs, les données permettant l'exécution des algorithmes sont stockées dans des dépôts virtuels qui peuvent être sujets aux cyberattaques ; leur vulnérabilité fragilise également les algorithmes.

Enfin, les algorithmes permettant de détecter des dangers émergents se basent sur des données historiques qui rendent l'émergence continue au cadre historique. Or, certaines émergences (émergences disruptives) ne peuvent être prédites à partir de données historiques. Elles peuvent en revanche être imaginées par différentes personnes de culture professionnelle différente.

3.2 La nécessité d'une implication sociétale

Au vu de ce qui a été évoqué précédemment, la société doit nécessairement avoir un regard critique sur les algorithmes et les données sur lesquels se basent les prédictions d'émergence. Elle est également utile à l'anticipation d'émergences disruptives à partir de séances d'intelligence collective mêlant sciences humaines et sociales et sciences de l'ingénieur.

Les données traitées par les algorithmes proviennent de plus en plus d'observatoires citoyens répartis à travers le globe. Les mesures concernent par exemple, la sismicité[27] ou la qualité de l'air.[28] Ces observatoires contribuent au développement de *big data* et à des modèles robustes statistiquement.

Les limites des algorithmes et des machines peuvent être palliées grâce à l'humain pour travailler ensemble à la résolution de problèmes : les machines sont utiles au traitement de grand nombre de données, alors que l'humain, outre son potentiel d'acquisition de données, est capable d'interpréter de façon critique les résultats, de développer ainsi des savoirs et de prendre des décisions.[29]

4 Conclusions

Le développement de notre société et les changements climatiques font apparaître des risques nouveaux qu'il faut pouvoir anticiper pour éviter des dommages sanitaires, environnementaux et financiers. Les risques sont la combinaison d'événements non espérés (dangers) et de leurs conséquences possibles sur l'environnement, la société et les infrastructures. Anticiper un risque revient donc à caractériser le danger, évaluer son degré de diffusion auprès d'enjeux (environnement, société et infrastructures) et les capacités de résilience de ces différents enjeux. Ce chapitre n'a traité que de l'utilisation d'algorithmes aux différentes étapes de la surveillance des dangers : collecte, sélection, traitement des données, identification et surveillance de l'émergence pouvant représenter un danger. Les limites des différents algorithmes ont été évoqués, faisant

[27] Rochford, K., Strauss, J. A., Kong, Q., & Allen, R. M. (2018). MyShake : Using Human-Centered Design Methods to Promote Engagement in a Smartphone-Based Global Seismic Network. *Frontiers in Earth Science*, *6*. https://doi.org/10.3389/feart.2018.00237

[28] Schaefer, T., Kieslinger, B., & Fabian, C. M. (2020). Citizen-Based Air Quality Monitoring : The Impact on Individual Citizen Scientists and How to Leverage the Benefits to Affect Whole Regions. Citizen Science : Theory and Practice, 5(1), 6. https://doi.org/10.5334/cstp. 245

[29] Russakovsky, O., Li, L.-J., & Fei-Fei, L. (2015). Best of both worlds : Human-machine collaboration for object annotation. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2121-2131. https://doi.org/10.1109/CVPR.2015.7298824

Published in : DELTORN, Jean-Marc ; PICHENOT, Evelyne (dir.) (dir.), *Algorithmes et Société*, Editions des archives contemporaines, ISBN : 9782813003928, doi : 10.17184/eac.9782813003928

apparaître une réelle complémentarité entre algorithmes et implication sociétale dans l'exercice d'identification et surveillance des dangers.

Les risques ne pouvant être évalués sans connaître la propension de la société à diffuser ou contrer une innovation potentiellement porteuse de dangers, ou à faire face aux dangers, il convient de prendre en compte également les aspects psychosociaux dans la surveillance des risques émergents.

Published in : DELTORN, Jean-Marc ; PICHENOT, Evelyne (dir.) (dir.), *Algorithmes et Société*, Editions des archives contemporaines, ISBN : 9782813003928, doi : 10.17184/eac.9782813003928