

# Intelligence artificielle et nouvelles approches méthodologiques pour la **maîtrise** des **risques** industriels

*Guillaume Fayet, docteur de l'Université Paris VI, Responsable d'études et de recherche à l'Ineris.*

*Guillaume Fayet a soutenu à l'Université Paris VI en 2010, une thèse sur « Le développement de méthodes prédictives des propriétés d'explosivité de substances chimiques » financée par l'Ineris. Il intègre alors l'Institut en tant que Responsable d'études et de recherche sur la sécurité des substances et procédés. Il pilote actuellement un Axe de Recherche sur la « Sécurité des substances, des matériaux énergétiques et des réactions chimiques » à la Direction Incendie Dispersion Explosion de l'Ineris.*

*Ses recherches concernent en particulier le développement et l'utilisation de méthodes prédictives, par chimie computationnelle, pour étudier et gérer les risques associés aux substances et procédés industriels. Ces travaux portent notamment sur la prédiction des dangers physiques (inflammabilité, explosivité) des substances pures et des mélanges, ainsi que sur l'étude théorique des réactions chimiques dangereuses (telles que les incompatibilités chimiques).*

## Introduction

Les activités et produits industriels sont sujets à des risques qu'il convient de maîtriser afin d'éliminer ou au moins limiter leur impact sur l'Homme et l'environnement. L'évaluation des risques industriels repose sur des approches expérimentales et de modélisation qui visent entre autres à caractériser les dangers (éco)-toxicologiques et physiques des substances chimiques, évaluer les risques associés aux réactions chimiques dangereuses ou encore estimer les conséquences des phénomènes dangereux associés.

## L'expert public pour la maîtrise des risques industriels et environnementaux

L'Ineris est un institut public sous la tutelle du ministère en charge de l'environnement. Sa mission est de contribuer à la prévention des risques

que peuvent induire les activités économiques pour l'Homme, l'environnement et les biens. L'Institut intervient sur un champ d'activité large (**Figure 1**) incluant les risques d'incendie et d'explosion à l'origine des accidents majeurs, la toxicologie et l'écotoxicologie des substances, l'impact des rejets industriels sur les milieux, la qualité de l'air ou encore la sécurité des sols et des cavités souterraines.

Naturellement, les thèmes des recherches de l'Ineris s'intéressent aux enjeux sociétaux actuels et à la sécurité des nouvelles technologies, que ce soit dans un contexte de transition vers de nouvelles énergies, de l'économie circulaire ou en ce qui concerne l'exploitation post-industrielle des mines et des énergies fossiles.

Les activités de l'Ineris sont de trois ordres. En premier lieu, l'Ineris apporte un appui technique aux pouvoirs publics en développant et validant des outils et des méthodes, et en leur apportant son expertise technique sur les risques industriels dans la mise en place des politiques publiques ou via des expertises réglementaires. L'Ineris accompagne également les industries dans l'évaluation des risques auxquels elles sont confrontées et vers des solutions de maîtrise des risques, ce qui permet ainsi à l'Institut d'être en contact direct avec le monde industriel. Enfin, cette expertise s'appuie sur

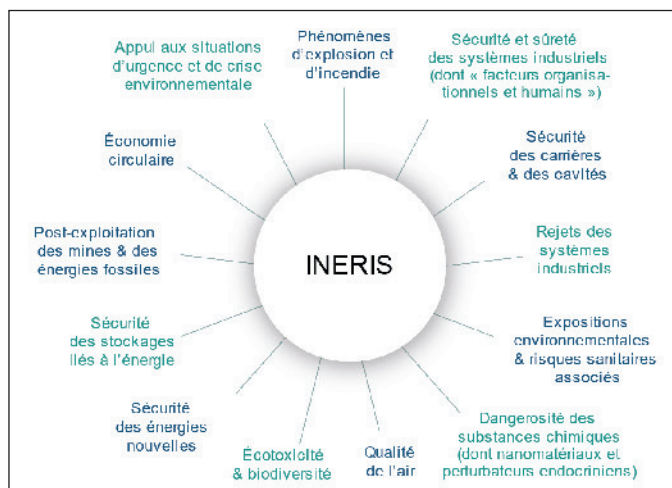


Figure 1

Champs d'activité de l'Ineris<sup>1</sup>.

1. <https://www.ineris.fr/fr/ineris/institut-bref/ineris-expert-public-institut-bref/ineris-expert-public-maitrise-risques-technologiques>

des activités de recherche de pointe, sur des programmes de recherche propres, via des projets de recherche collaboratifs nationaux et européens ou dans le cadre de projets de recherche partenariale.

### Une expertise basée sur l'approche expérimentale, la modélisation et la connaissance du monde industriel

L'Ineris met en œuvre des moyens expérimentaux et de modélisation de pointe. L'Institut dispose en effet d'un grand nombre d'installations expérimentales (*Figure 2*) depuis l'échelle laboratoire, pour étudier par exemple les dangers des nanomatériaux ou pour caractériser les paramètres clés des réactions chimiques dangereuses, jusqu'à l'échelle réelle. L'Ineris dispose notamment

pour cela d'une plateforme pyrotechnique pour réaliser des essais sur des matières explosives ou des réactions chimiques dangereuses à plus grande échelle, d'une plateforme incendie pour tester le comportement de produits pris dans des incendies ou encore de mésocosmes, des rivières artificielles représentant les écosystèmes naturels, dans lesquels sont étudiés les impacts que peuvent avoir des polluants sur les milieux naturels.

Les laboratoires de l'Ineris s'adaptent en permanence pour répondre aux enjeux réclamés par le développement des nouvelles technologies. Ainsi, une plateforme, dénommée STEEVE, a été mise en place pour l'évaluation de la sécurité des batteries.

L'Ineris dispose également de nombreux moyens de modélisation, là encore à différentes



Figure 2

Exemples d'installations expérimentales à l'Ineris.

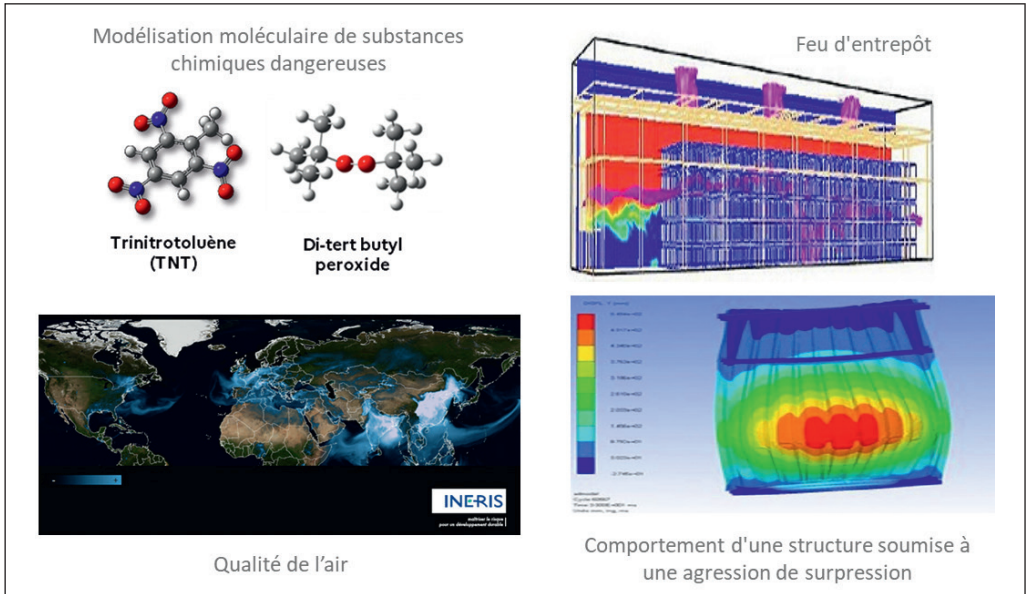


Figure 3

Exemples de modélisations réalisées à l'Ineris.



Figure 4

Modélisation de la dispersion des particules de plomb du panache de l'incendie de Notre-Dame (2019)<sup>2</sup>.

échelles (Figure 3). Des approches à l'échelle moléculaire permettent de prédire les dangers des substances et de comprendre leur réactivité. À plus grande échelle, l'Ineris modélise par exemple des feux

d'entrepôt, l'impact de phénomènes dangereux sur des structures et met en œuvre des modèles jusqu'à une échelle « géographique » pour l'évaluation de la qualité de l'air. L'Ineris intervient notamment dans un contexte post-accidentel pour modéliser des dispersions de fumées ou de particules, comme la dispersion des particules de plomb dans le panache de l'incendie de Notre-Dame en 2019 (Figure 4).

### Apport des nouvelles approches méthodologiques et de l'intelligence artificielle

Si l'approche expérimentale représente une pièce fondamentale de l'expertise de l'Ineris, depuis la caractérisation

2. <https://www.ineris.fr/fr/ineris/actualites/incendie-dame-rapport-ineris-est-paru>

en laboratoire jusqu'aux essais à grande échelle, les progrès scientifiques et techniques ont donc élargi au cours du temps le panel d'outils disponibles par des outils de modélisations de plus en plus poussés. Aujourd'hui, de nouvelles approches méthodologiques (notamment celles basées sur l'intelligence artificielle) permettent d'accéder à des informations plus nombreuses, plus rapidement et parfois plus complètes.

Il s'agit soit de compléter les moyens existants (pour exploiter de manière plus précise, plus étendue et efficace les données produites par les essais et modélisations), soit de disposer de méthodes alternatives pour gagner du temps ou optimiser des plans d'expérience. Ces outils s'intègrent dans une stratégie de digitalisation et de capitalisation des données disponibles à la fois à l'Ineris et à l'extérieur afin que ce retour d'expérience contribue à une meilleure maîtrise des risques.

Au niveau des activités expérimentales, au-delà de la gestion et de l'exploitation des données, des métrologies innovantes permettent le traitement d'informations plus nombreuses et en temps réel pendant les essais. Ces outils ouvrent la voie à des analyses et des traitements plus automatisés en faisant gagner du temps aux opérateurs et aux experts, leur permettant de se consacrer pleinement aux enjeux de recherche qui y sont associés.

Ces données permettent également de développer de nouveaux types d'outils basés en

construisant des modèles prédictifs et des méta-modèles de substitution ou compléments des modèles existants qui nécessitent dans certains cas des données et des temps de calculs importants.

Ces approches numériques nouvelles permettent d'accompagner *in fine* le développement de procédés et de substances plus sûres. Elles peuvent par exemple aider à la formulation et à la substitution en prenant en compte les risques industriels au plus tôt dans les démarches de développement des substances chimiques.

Quelques exemples d'application à l'Ineris de ces nouvelles approches méthodologiques sont ici présentés.

## 1 Prédire les dangers des substances

### 1.1. Le règlement REACH et les méthodes prédictives

Des recherches ont été initiées à l'Ineris il y a environ 15 ans pour le développement de méthodes prédictives des dangers physiques des substances, dans le contexte de la mise au point du règlement européen REACH<sup>3</sup>.

3. Le règlement européen REACH (Enregistrement, Évaluation et Autorisation de Produits Chimiques en français), adopté en 2007, vise à mieux protéger la santé humaine et l'environnement contre les risques liés aux substances chimiques. Toute substance mise sur le marché ou transitant dans l'Union européenne à plus d'une tonne par an doit y être enregistrée. Le règlement prévoit des mesures d'interdiction ou de restriction pour les substances les plus préoccupantes.



Cette nouvelle réglementation concerne l'autorisation de mise sur le marché des substances sur le territoire européen et a impliqué une quantité de travaux très importante, pour l'enregistrement d'un grand nombre de substances dans un calendrier défini et serré.

Or, les essais de caractérisation des dangers des substances peuvent dans certains cas être très coûteux et la disponibilité des laboratoires d'essai peut poser problème. Par ailleurs, la dangerosité des propriétés et substances testées peut imposer des contraintes importantes (essais sur des explosifs ou des substances toxiques). La caractérisation de la toxicologie et de l'écotoxicologie pose également des problèmes d'éthique avec une volonté de réduire le recours aux essais sur animaux, une recommandation toujours très actuelle pouvant entraîner le besoin d'utiliser des méthodes alternatives en substitution des essais actuels, qu'ils soient par des calculs numériques comme illustré par la suite, ou par des essais *in vitro*<sup>4</sup>.

Ainsi le règlement REACH encourage le recours au partage et à la capitalisation des données ainsi qu'à l'utilisation des méthodes alternatives, telles que les modèles QSAR<sup>5</sup>

ou QSPR<sup>6</sup>, qui reposent sur des relations structures chimiques-propriétés en utilisant des approches de *machine learning*<sup>7</sup>. Très connues pour la toxicologie et en général pour des produits purs, l'Ineris les développe et les utilise également sur les dangers physiques et dans le cas de mélanges.

L'approche QSAR/QSPR repose sur un principe de similarité selon lequel des molécules avec des structures chimiques similaires présenteront probablement des propriétés et des activités biologiques similaires. Tout l'enjeu de cette approche est donc de trouver des descripteurs pertinents pour caractériser la structure moléculaire qui soient corrélés avec la propriété ciblée (Figure 5).

Les données disponibles pour chercher ces corrélations sont la base du développement des modèles QSAR/QSPR. Ils sont entraînés sur un jeu de données expérimentales qui doivent être suffisamment nombreuses, fiables et obtenues dans des conditions homogènes.

Pour représenter les structures moléculaires, différents types de descripteurs sont utilisés. Certains sont simples à calculer et à utiliser comme des nombres

4. L'approche *In vitro* consiste à réaliser des tests sur des modèles cellulaires en laboratoire. Ces tests sont utilisés comme alternatives aux essais sur animaux.

5. QSAR pour *Quantitative Structure-Activity Relationships* : Relations Quantitatives Structure-Activité.

6. QSPR pour *Quantitative Structure-Property Relationships* : Relations Quantitatives Structure-Propriété.

7. Méthode d'apprentissage automatisée, forme d'intelligence artificielle, consistant à utiliser des approches statistiques pour entraîner un modèle à partir d'une base de données d'apprentissage.

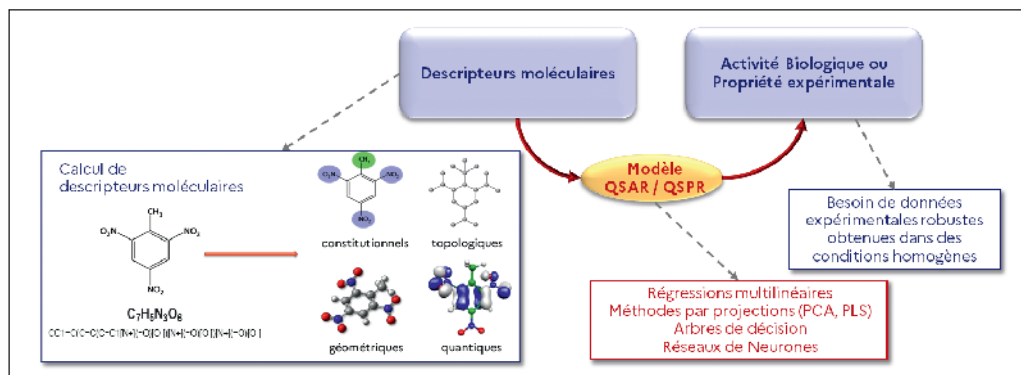


Figure 5

Approche QSAR/QSPR.

d'atomes. D'autres sont plus complexes pour caractériser des propriétés électroniques ou de réactivité (à partir de calculs de chimie quantique). Ces derniers seront plus compliqués à utiliser mais peuvent apporter une compréhension physico-chimique plus complète.

Le travail consiste ensuite à analyser les données et à établir des corrélations au sein d'un modèle mathématique reliant les descripteurs moléculaires les plus pertinents à la propriété que l'on cherche à prédire. Pour cela, différentes méthodes de *machine learning* sont utilisées, depuis les régressions linéaires simples jusqu'à des réseaux de neurones artificiels plus complexes et plus adaptés lorsque des quantités de données plus importantes sont disponibles. Les méthodes pourront être « supervisées » ou « non supervisées » pour obtenir des « données quantitatives » ou bien des « classifications » selon les besoins et les systèmes étudiés.

## 1.2. Cas des dangers physiques des matières auto-réactives

Un premier exemple de modèle QSPR développé à l'Ineris concerne les matières auto-réactives. Une matière auto-réactive est une substance thermiquement instable, susceptible de générer une décomposition exothermique<sup>8</sup>, potentiellement forte et rapide. Une des particularités de ces substances est qu'elles n'ont pas forcément besoin d'un apport de l'oxygène de l'air pour se décomposer. On retrouve dans cette classe de substances dangereuses différentes familles de composés chimiques (Figure 6).

La caractérisation des dangers physiques de ces substances

8. Exothermique : qui dégage de la chaleur.

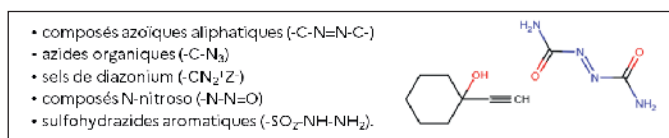


Figure 6

Groupes chimiques typiques rencontrés dans des matières auto-réactives.

est importante, car elles sont susceptibles de se décomposer de manière explosive. Les effets sont alors très importants. Elles brûlent rapidement. Prises dans un feu, elles peuvent amplifier l'incendie. Elles peuvent aussi présenter une sensibilité aux chocs ou à la friction, ce qui peut générer des accidents. Dans certains cas, elles sont susceptibles de réagir dangereusement au contact d'autres substances, on parle d'incompatibilité chimique.

Du fait de ces propriétés, elles ont un classement particulier dans le cadre des réglementations, par exemple pour le transport de marchandises dangereuses ou dans les réglementations liées à l'étiquetage et à l'emballage des produits. Des contraintes particulières et un étiquetage particulier leur sont alors appliqués.

9. G. Fayet *et al.*, "First QSPR models to predict the thermal stability of potential self-reactive substances", *Process Safety and Environmental Protection* 163 (2022) 191-199.

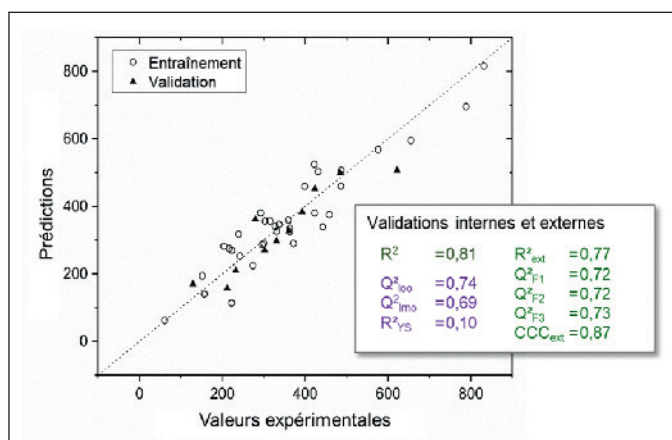


Figure 7

Modèle développé pour la chaleur de décomposition des matières auto-réactives<sup>9</sup>.

Jusqu'à présent, ces substances sont caractérisées exclusivement sur la base d'essais expérimentaux, car aucun modèle prédictif n'existait. L'Ineris a donc engagé des travaux sur ce sujet dans le cadre d'un projet européen (en collaboration notamment avec le BAM<sup>10</sup> en Allemagne) pour combler ce manque de modèles et pouvoir anticiper le caractère auto-réactif des substances chimiques.

Dans le cadre de ce travail, l'Ineris a développé un modèle pour évaluer la chaleur de décomposition de substances susceptibles de présenter un caractère auto-réactif<sup>11</sup> (Figure 7). Cette propriété est utilisée comme critère de présélection pour le classement en tant que matière auto-réactive. Il permet de décider s'il est nécessaire ou non d'engager toute une campagne d'essai sur le caractère auto-réactif de la substance.

Il s'agit du premier modèle développé pour cette famille de substances. Il a été développé à partir d'une base de données générée spécifiquement pour cette étude à partir d'essais sur une série d'échantillons fournis par la société Bayer<sup>12</sup>. Ils ont tous été réalisés par le BAM selon des protocoles identiques pour s'assurer de conditions expérimentales uniformes.

10. BAM pour *Bundesanstalt für Materialforschung und prüfung* : Institut fédéral pour la recherche et les essais des matériaux (Allemagne).

11. Chaleur de décomposition : la température à laquelle la substance se décompose chimiquement.

12. Société pharmaceutique et agrochimique allemande.



Dans ce type d'étude, l'enjeu clé du travail est de sélectionner les paramètres à retenir pour le modèle final. Ceci a été réalisé à l'aide d'un « algorithme génétique » qui a permis de choisir les descripteurs les plus pertinents. Une contrainte particulière pour l'Ineris réside dans la volonté que les modèles développés puissent être utilisés dans un cadre réglementaire. Ils doivent donc répondre à des principes de validation mis en place par l'OCDE<sup>13</sup>. Sont en particulier demandées : une définition claire de la propriété visée, une définition du domaine d'applicabilité du modèle, des validations statistiques approfondies et si possible une interprétabilité du modèle.

Néanmoins, le fait qu'un modèle satisfasse à ces principes ne suffit pas à valider toute prédiction issue de ce modèle. L'OCDE propose également des recommandations pour la validation des prédictions issues de ces modèles afin de vérifier qu'elles soient suffisamment fiables et permettent une prise de décision dans le contexte réglementaire visé.

13. OCDE : Organisation de coopération et de développement économiques. L'Ineris participe aux groupes de travail de l'OCDE définissant les principes de validation des modèles et prédictions QSAR pour un usage réglementaire.

### 1.3. Extension de la modélisation au cas des mélanges – Exemple du point d'éclair

En pratique, il est très courant qu'une substance chimique ne soit pas un composé pur mais un mélange dont la composition est déterminée pour obtenir des fonctions particulières. Cette situation nécessite de modifier la méthodologie en développant des approches permettant de prendre en compte les spécificités des mélanges. En effet, le principe d'un modèle QSPR est de corrélérer la variation de structure de la molécule à ses propriétés. Pour un mélange, plusieurs composés sont présents et différents facteurs influençant la propriété sont à prendre en compte comme leurs concentrations respectives et les interactions qu'ils peuvent avoir entre eux.

Une telle adaptation a été réalisée pour la prédiction du point d'éclair, la propriété qui sert de base au classement des liquides inflammables. Il s'agit de la température à partir de laquelle le liquide s'enflamme à l'approche d'une flamme. Plus la température du point éclair est basse, plus le liquide est inflammable (**Figure 8**).

L'adaptation de l'approche a consisté à définir des

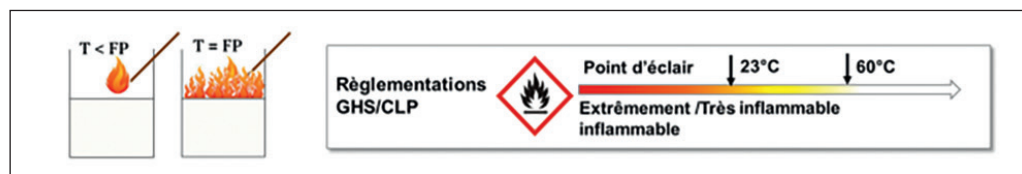


Figure 8

Point d'éclair et liquides inflammables.

descripteurs de mélanges, calculés à partir des descripteurs moléculaires et de la fraction molaire<sup>14</sup> des différents constituants du mélange sur lesquels la méthodologie QSPR est ensuite appliquée, de la même manière que pour les substances pures.

Ce travail a été un succès puisque, comme le montre la *Figure 9*, le modèle obtenu fournit de bonnes prédictions, assez proche des performances attendues pour des modèles de produits purs. Même si d'autres approches (thermodynamiques) peuvent être utilisées pour estimer le point d'éclair de liquides inflammables, cette modélisation fait maintenant partie

de la boîte à outils à disposition de l'Ineris.

#### 1.4. Utilisation des modèles QSPR

##### 1.4.1. Sécurité des substances et aide à la formulation/substitution

Les modèles QSPR sont intéressants d'un point de vue réglementaire pour combler des données manquantes. Au-delà des dangers physiques, l'Ineris s'intéresse de la même manière aux modèles QSAR pour la toxicologie et l'écotoxicologie. De telles prédictions permettent également d'accéder à des données utiles dans des études de sécurité des procédés. Elles permettent en particulier de faire varier des paramètres comme la concentration de tel ou tel composé et d'évaluer ainsi l'influence de la composition d'un mélange sur la sécurité d'une installation industrielle. La modélisation QSPR est très précieuse dans le cadre d'études R&D<sup>16</sup> pour le développement de substances nouvelles ou alternatives à des substances existantes. Elle permet de prendre en compte leurs dangers le plus en amont possible dans les phases de recherche. Cela peut éviter d'importants efforts de recherche, synthèses et caractérisations sur des molécules a priori intéressantes pour leurs propriétés fonctionnelles, mais qui présentent finalement

14. Grandeur quantifiant la proportion de quantité de matière d'un constituant dans un mélange. La fraction molaire est comprise entre 0 et 1 (1 pour un composé pur, 0 pour un composé absent).

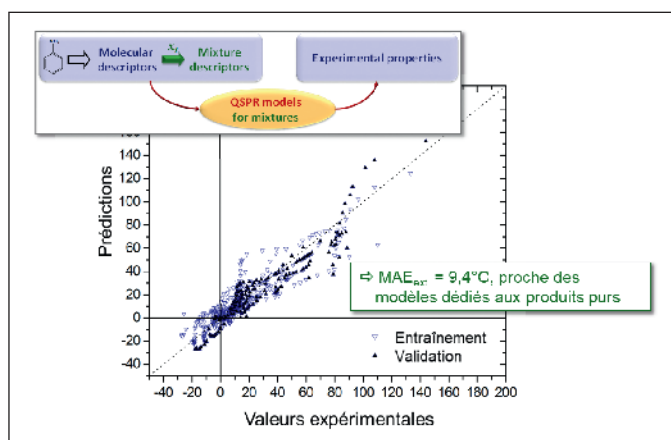


Figure 9

Modèle développé pour le point d'éclair de mélanges liquides inflammables<sup>15</sup>.

15. G. Fayet & P. Rotureau, "New QSPR Models to Predict the Flammability of Binary Liquid Mixtures", *Molecular Informatics* 38 (2019) 1800122.

16. R&D : Recherche et Développement.

des dangers induisant des contraintes de sécurité trop importantes.

Les modèles QSPR sont donc utilisés non seulement pour prédire les propriétés de substances particulières, mais aussi à des fins de criblage au sein de listes de molécules. À partir d'une base de données importante de substances, des modèles QSPR peuvent être utilisés dans une analyse multicritère incluant les propriétés dangereuses et la toxicité (Figure 10). À l'aide de ces prédictions, il est ainsi possible de sélectionner les candidats les plus pertinents pour les étapes ultérieures de synthèse et de caractérisations expérimentales. Il est également possible d'aller plus loin en proposant de nouvelles molécules par *Design in Silico*<sup>17</sup>. Le principe est alors de se baser sur les modèles et les paramètres qu'ils génèrent pour construire des bases de données de

nouvelles molécules (virtuelles) sur lesquelles est réalisé le criblage (Figure 11).

Une telle approche permet la prise en compte précoce de la sécurité et sera particulièrement intéressante dans la recherche de solutions durables et plus sûres dans des démarches de type « *safe and sustainable by design*<sup>18</sup> ».

#### 1.4.2. Application au cas des tensioactifs dérivés de sucre

Des travaux visant ce type d'application ont été menés dans le cadre de projets financés par l'ITE<sup>19</sup> Pivert, à savoir Amphipred et Amphifoam dont l'objectif était de développer des tensioactifs biosourcés, afin de les substituer aux produits pétro-sourcés actuellement sur le marché.

Ce travail a été réalisé en collaboration avec différents

17. *Design in Silico* : Conception par ordinateur.

18. *Safe and Sustainable by Design* : sûr et durable par conception.

19. ITE : Institut de Transition Énergétique.

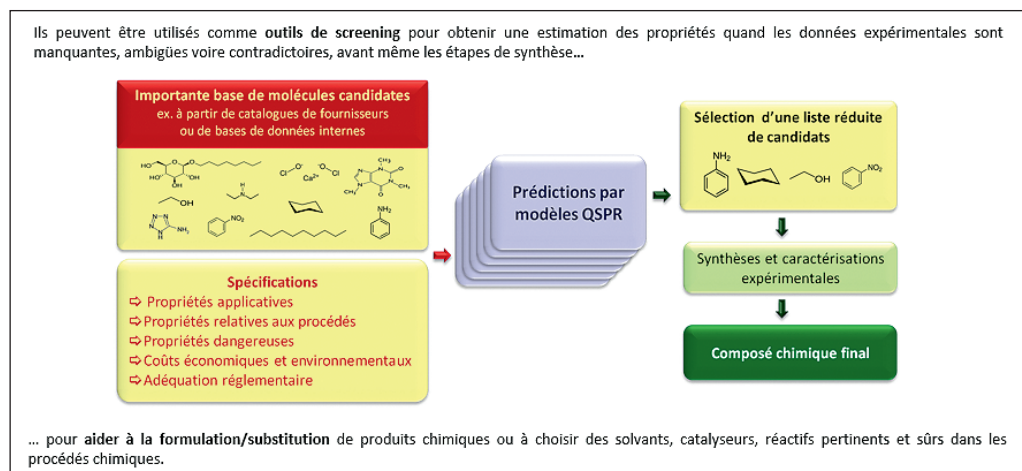


Figure 10

Utilisation des modèles QSPR comme outils de screening.

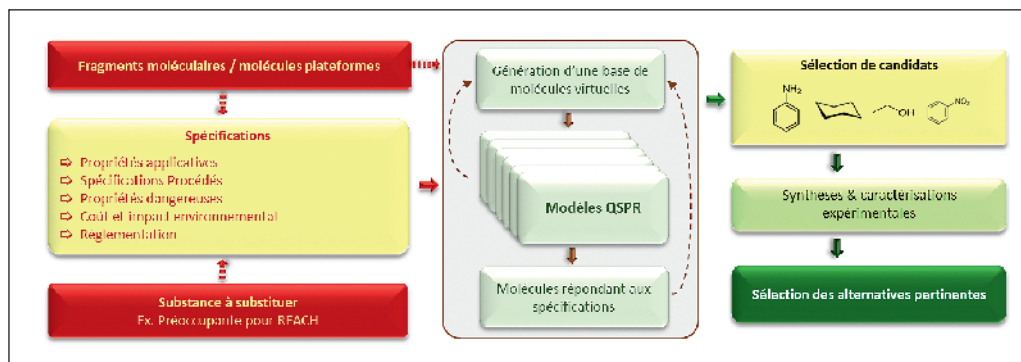


Figure 11

Modèle du Design in silico.

partenaires dont l'Université de Technologie de Compiègne et l'Université Picardie Jules Vernes à Amiens. Une des caractéristiques de ces substances est qu'elles sont constituées d'une tête polaire<sup>20</sup> (hydrophile) et d'une chaîne alkyle<sup>21</sup> (hydrophobe) qui leur donnent des propriétés spécifiques en solution (formation de mousses). Cette structure moléculaire particulière a encouragé l'utilisation de descripteurs de fragments pour le développement des modèles QSPR permettant de prédire les propriétés de ces substances mais aussi pour créer de nouveaux tensioactifs. Les propriétés visées étaient la concentration micellaire critique<sup>22</sup> (la CMC), la tension de

surface<sup>23</sup> à la CMC, l'efficacité (pC20) et le point de Krafft (TK)<sup>24</sup>.

De nouveaux modèles spécifiques aux tensioactifs dérivés de sucre ont tout d'abord été développés (Figure 12), car les modèles existants pour les propriétés d'intérêt étaient dédiés aux molécules pétrosourcées et ne montraient pas les mêmes niveaux de performance pour les tensioactifs dérivés de sucre.

Une démarche de *Design in Silico* a ensuite été développée pour obtenir de bons candidats pour de futures synthèses et caractérisations (Figure 13).

La première étape consiste à définir un cahier des charges des propriétés désirées pour ces substances. Étant donné qu'il s'agissait d'une démarche de substitution, il a

20. Polarité : non homogénéité de la répartition des charges au sein de tout ou partie d'une molécule.

21. Succession d'atomes de carbone, chacun lié à deux hydrogènes.

22. Concentration en surfactants au-delà de laquelle il leur est favorable de se regrouper sous forme de structures supramoléculaires nommées micelles pour minimiser leur surface de contact avec l'eau.

23. Grandeur caractérisant la résistance d'un fluide à augmenter sa surface de contact avec un autre fluide.

24. Température minimale pour qu'un tensioactif forme des micelles. Cette propriété permet de vérifier si un tensioactif est utilisable à température ambiante.

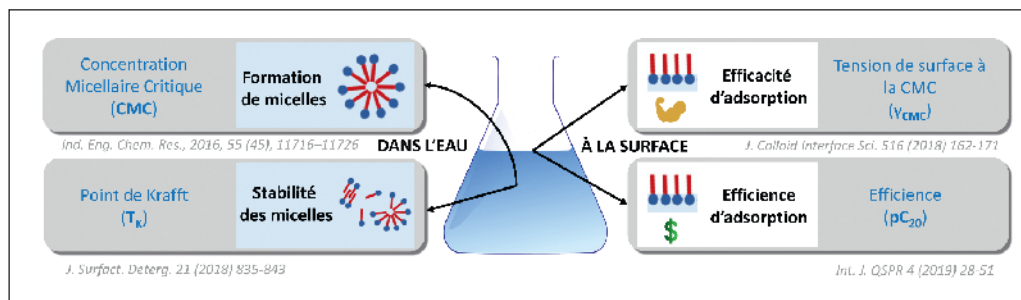


Figure 12

Modèles développés pour les tensioactifs dérivés de sucre.

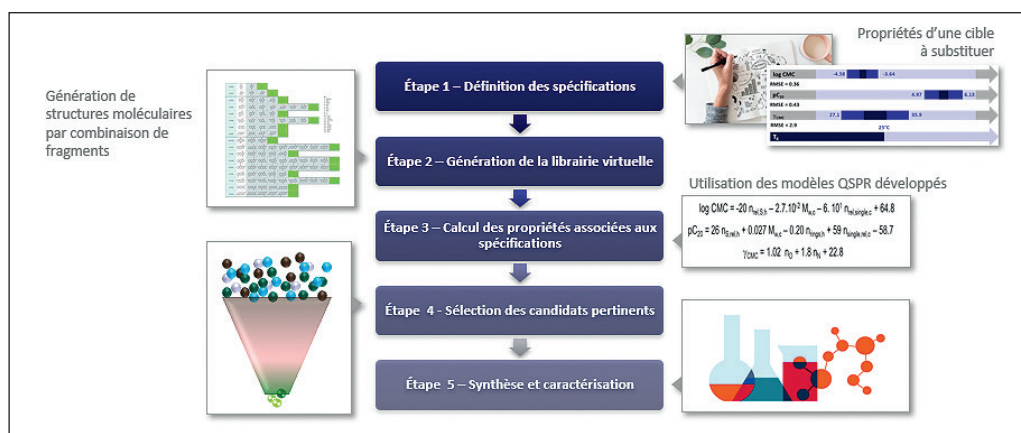


Figure 13

Méthode de Design in Silico.

été basé sur une sélection de cibles à substituer et sur leurs propriétés, l'objectif étant de rechercher des molécules qui pouvaient avoir des propriétés proches.

Les fragments moléculaires modélisés pour le développement des modèles ont été exploités pour générer une base de données de molécules virtuelles par combinaison des fragments hydrophiles et hydrophobes. De cette manière, plus de 2 500 molécules ont déjà pu être générées à partir de fragments

représentés dans les molécules étudiées pour le développement des modèles QSPR et cette base pourrait être facilement étendue en ajoutant plus de fragments.

Un criblage au sein de cette base de tensioactifs virtuels a ensuite été réalisé pour identifier les candidats les plus pertinents, qui pourraient être considérés dans une étape suivante de synthèse et de caractérisation.

Le nombre de candidats proposés était bien supérieur à

ce qui a pu être trouvé dans la base de données AmphInnov qui correspondait au recensement de surfactants dérivés de sucre le plus exhaustif connu (Figure 14). Les candidats recensés répondaient à chaque critère du cahier des charges. Par ailleurs, les prédictions issues des modèles étaient en bon accord avec les valeurs expérimentales disponibles, validant la pertinence de l'approche proposée.

## 2 Faciliter et améliorer l'identification de contaminants dans l'environnement

### 2.1. Identifier les substances contaminant l'environnement

La présence de contaminants dans l'environnement, notamment dans les milieux aquatiques, peut avoir un impact néfaste et interroger l'Ineris. Pour identifier les substances potentiellement polluantes, des échantillons sont prélevés puis analysés au laboratoire à l'aide d'instruments de pointe. Malgré les technologies disponibles très avancées,

l'identification de contaminants peut s'avérer longue et difficile.

Une étude a donc été lancée avec pour objectif d'aider à l'identification des contaminants dans le cadre d'analyses LC-HRMS (chromatographie liquide<sup>25</sup> couplée à la spectrométrie de masse à haute résolution<sup>26</sup>), en utilisant les nouvelles techniques issues de l'intelligence artificielle. La Figure 15 présente la procédure expérimentale de base, qui conduit à l'identification d'un très grand nombre de molécules, plusieurs milliers de molécules, via environ 20 000 spectres générés pour un échantillon.

L'immense quantité de données générées par chaque analyse est traitée par des logiciels constructeurs ou libres afin d'aider l'opérateur à identifier des substances présentes dans les échantillons. Concrètement, les données

25. Méthode de séparation de composés basée sur leur temps d'écoulement dans une colonne.

26. Méthode d'analyse d'un mélange de composés basée sur leurs différences de masse.

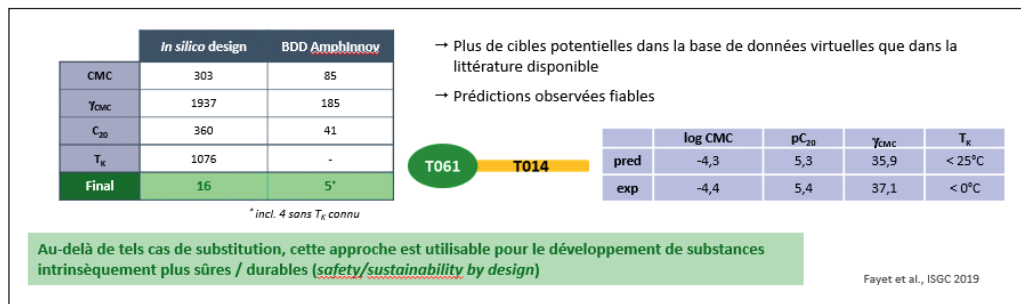


Figure 14

Comparaison du nombre de tensioactifs proposés par l'approche de Design in Silico avec ceux trouvés dans la littérature scientifique disponible (Base de données AmphInnov).



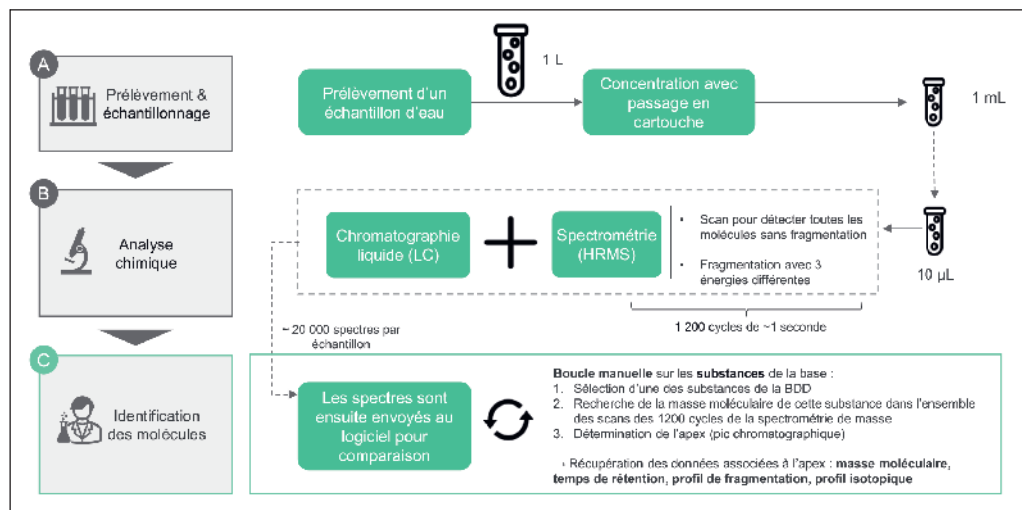


Figure 15

Schéma du processus analytique pour l'identification de contaminants par LC-HRMS.

générées sont comparées à des bases de données, et le « matching » (la correspondance) entre les données générées et celles des bases de données permet d'identifier des substances. L'opérateur doit tout de même analyser les données générées et traitées par le logiciel pour valider ou non les identifications. Cette procédure est longue, fastidieuse et peut s'avérer peu reproductible d'un opérateur à un autre.

D'autre part, les bases de données utilisées pour identifier des correspondances, fournies par les constructeurs ou créées par les laboratoires utilisateurs, sont peu fournies ; ceci limite les possibilités d'identification de substances dans l'échantillon. Les méthodes d'automatisation permises par l'IA couplées à de plus larges bases de données disponibles en libre accès éviteraient ces inconvénients et limitations.

Un autre intérêt serait le gain de temps qu'elles offriraient : l'opérateur doit actuellement vérifier manuellement chaque concordance, substance par substance pour chaque échantillon. L'objectif de ce travail était également de chercher à automatiser et accélérer ce travail de vérification.

## 2.2. Intégrer l'intelligence artificielle dans le processus métier

L'automatisation des identifications grâce à l'IA a été réalisée dans le cadre d'un « projet empreinte environnementale » financé par un Appel à Manifestation d'Intérêt (AMI IA) en 2020, et une aide dans le cadre des projets de France Relance<sup>27</sup> 2021. L'enjeu était de créer et d'utiliser des

27. Aide publique au financement de projets à visée écologique ou sociale.

*features*<sup>28</sup> particuliers à partir des spectres de masse pour construire un modèle capable de reproduire le travail de vérification de l'opérateur, a minima pour une partie des identifications, afin de lui permettre de se concentrer sur les cas les plus litigieux. L'utilisation de l'IA de type *machine learning* a permis d'élaborer un modèle capable d'identifier de façon automatique en moyenne la moitié des substances dans un échantillon.

De plus, l'inclusion des connaissances propres à l'Ineris ainsi que des données librement disponibles dans des bases de données européennes (MassBank) a permis de multiplier par 4 la taille de la base de données originale (la base constructeur contenait 1 200 molécules).

### 2.3. Plus-values observées et perspectives

Les gains obtenus avec l'approche IA sont importants. L'outil constructeur (Figure 16)

est basé sur un matching automatique focalisé sur la masse moléculaire et la masse de certains fragments, l'opérateur doit ensuite vérifier la correspondance des intensités de fragment, du temps de rétention et du profil isotopique pour valider ou non l'identification. L'outil développé avec l'intelligence artificielle peut vérifier, plus rapidement, les correspondances entre échantillon et base de données au regard de ces cinq paramètres. Le taux d'analyse est accru et la systématisation de l'analyse faite par l'outil informatique permet également d'augmenter la reproductibilité et la fiabilité des résultats obtenus. Par ailleurs, l'augmentation de la taille de la base de données a permis de multiplier le nombre de substances identifiées par 2.

En pratique, pour l'analyse d'environ 50 échantillons (Figure 17), le temps de travail qui était réalisé jusqu'à présent avec l'outil constructeur était de 39 jours pour 500 identifications. Avec l'outil IA, le temps de travail de l'opérateur est divisé par plus de deux et

28. *Features* : caractéristiques.



Figure 16

Données traitées par le logiciel constructeur (en vert) et par l'opérateur (en rouge) pour conduire à une identification.

un nombre légèrement plus élevé d'identifications est obtenu, en utilisant la même base de données. L'utilisation de la base de données étendue permet d'identifier deux fois plus de substances en 50 jours environ.

L'outil IA est plus fiable parce que plus reproductible, mais des perfectionnements supplémentaires pour l'automatisation seront encore possibles dans l'avenir. Cependant, il faut une nouvelle fois rappeler que la présence d'un opérateur est nécessaire pour les cas les plus litigieux.

Sur la thématique, d'autres travaux sont encore à conduire avec l'aide de l'IA, comme la détermination des sources de polluants (par exemple, relier les rejets aux sources de formation de particules).

### 3 D'autres applications de l'intelligence artificielle

Deux autres applications moins en lien avec les substances chimiques et leurs dangers sont citées pour illustrer d'autres utilisations de l'IA à l'Ineris.

#### 3.1. Surveillance microsismique

La première concerne la surveillance microsismique, classiquement utilisée pour surveiller l'exploitation industrielle du sous-sol : mines, réservoirs, stockages géologiques, géothermie profonde entre autres. Dans certains cas, les anciennes exploitations souterraines font l'objet d'une activité microsismique

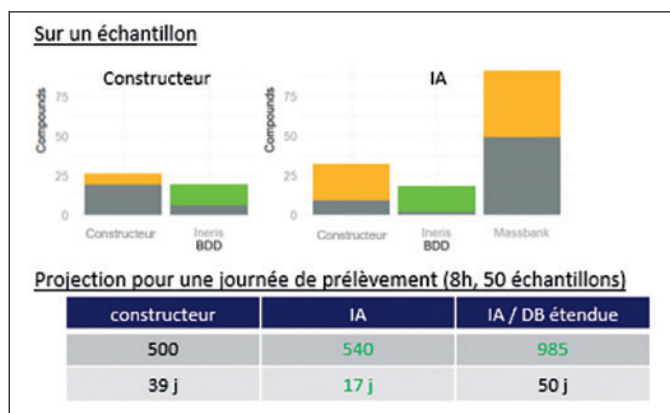


Figure 17

Comparaisons des performances entre un outil constructeur et l'outil IA par l'intelligence artificielle.

résiduelle de long terme durant leur phase de post-exploitation, le temps nécessaire au sous-sol de retrouver un nouvel équilibre hydromécanique.

La surveillance microsismique est habituellement réalisée via l'enregistrement et l'analyse de signaux provenant de stations de surface ou en forages, réparties autour de la zone d'intérêt à surveiller. Dans le cas de la surveillance d'anciennes mines, les signaux enregistrés peuvent bien évidemment être générés par l'endommagement et/ou la rupture d'une partie des ouvrages souterrains, mais également avoir pour origine des activités anthropiques de surface (travaux, bruits parasites divers, tirs) ou une origine naturelle (séismes naturels, orages et surtensions) (Figure 21). Les signaux, issus de ces différents phénomènes, présentent des similarités et sont usuellement analysés dans un délai court par des opérateurs spécialistes, pour

en déterminer l'origine (phase de qualification), les traiter (phase de traitement du signal et de calculs à la source) et ainsi aider l'expert à définir s'ils correspondent au déclenchement d'un évènement redouté. L'automatisation de la phase de qualification de données en quasi-temps réel et de manière fiable est donc un enjeu important.

Les méthodes d'apprentissage (*machine learning*) associées à l'intelligence artificielle sont bien adaptées pour réaliser de la classification automatique d'informations. Plusieurs types de réseaux neuronaux (CNN, Inception et LSTM) ont pu être testés pour évaluer leur capacité à prédire, sans intervention humaine, 24 h/24, l'origine des signaux microsismiques issus de travaux souterrains. Le cas d'application de l'ancien bassin minier de Gardanne, objet d'une microsismicité résiduelle de long terme, a été utilisé car tout à fait représentatif de

la surveillance d'opérations industrielles sismogéniques en milieu souterrain.

La **Figure 18** montre que jusqu'à plus de 98 % des signaux sont correctement classés pour ce site, ouvrant la perspective vers d'autres applications. Mais l'outil ne remplacera pas l'expert qui intervient pour confirmer les prédictions de l'IA et évaluer les conséquences possibles des signaux précurseurs d'endommagement potentiel en surface.

### 3.2. Méta-modèles – Explosions en milieu confiné

L'intelligence artificielle peut également être utilisée pour développer des méta-modèles, c'est-à-dire des modèles prédictifs, basés sur des approches d'apprentissage automatiques (*machine learning*), en substitution des modèles phénoménologiques existants (parfois trop lourds ou nécessitant la connaissance

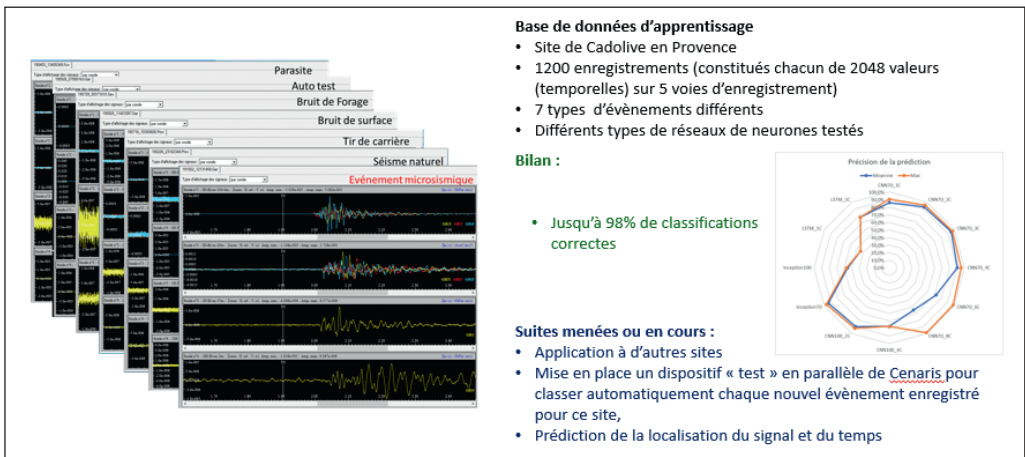


Figure 18

Résultats obtenus pour la surveillance microsismique.

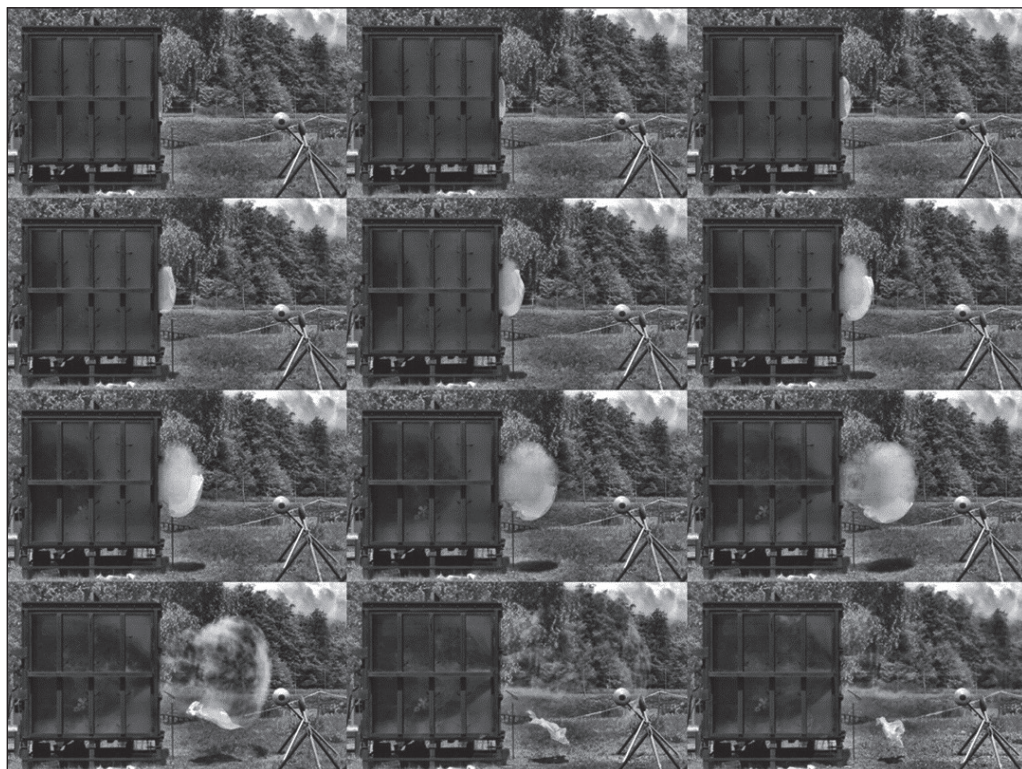


Figure 19

Exemple d'essais d'explosion de mélange hydrogène-air.

de nombreux paramètres). Ce principe est appliqué ici à l'étude des explosions en milieu confiné.

La Figure 19 présente des images acquises avec une caméra rapide lors d'une explosion de gaz dans une enceinte équipée d'un événement d'explosion<sup>29</sup>. On y observe l'explosion d'un mélange air-hydrogène, ainsi que l'éjection de gaz frais puis brûlés, à l'extérieur, par le biais de l'ouverture.

29. Événement d'explosion : il s'agit d'une sorte de porte calibrée pour s'ouvrir à une pression choisie, qui vise à s'ouvrir en cas d'explosion pour évacuer les gaz et limiter la surpression dans l'enceinte.

Lors de ce type d'expériences, les mesures effectuées consistent en des signaux de pression, comme illustré dans la Figure 20. Pour modéliser ce

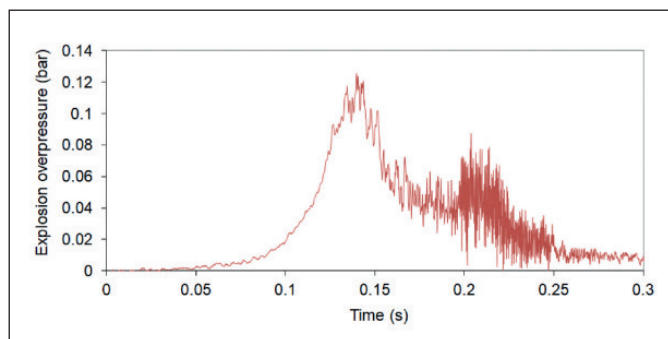


Figure 20

Exemple de signal de pression mesuré.



type de phénomène, il existe des modèles empiriques<sup>30</sup> ou phénoménologiques, qui reposent sur des modèles de physique connus et assez bien maîtrisés, mais dépendent de nombreux paramètres, dont certains ne sont pas aisément accessibles. L'idée sous-jacente est de développer, grâce à l'intelligence artificielle, des métamodèles substitutifs plus facilement utilisables, plus robustes lorsqu'il manque une partie des données ou dans le cadre de criblages préliminaires.

Des premiers travaux ont été engagés en ce sens en exploitant des données historiques recensées dans la littérature, englobant environ 270 essais réalisés au cours des 50 dernières années pour différentes configurations d'explosions et différents types de gaz. La **Figure 21** schématise la structure du réseau de neurones retenu qui utilise en entrée des variables aisément accessibles

30. Basés sur l'expérience.

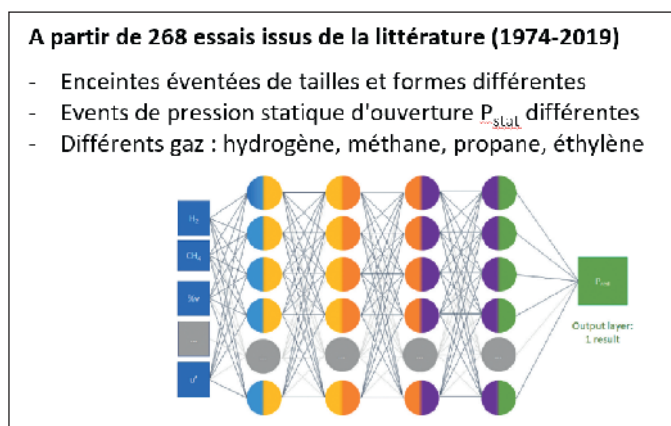


Figure 21

Structure du réseau de neurones du métamodèle développé.

décrivant à chaque fois une configuration d'un essai d'explosion, telles que la nature du gaz (hydrogène, méthane, propane, éthylène), sa concentration dans l'air, le volume de l'enceinte, la taille de l'évent. Il a été choisi de ne pas introduire dans ces données des paramètres plus difficilement accessibles tels que la vitesse de flamme.

Malgré l'absence d'une optimisation complète de la structure du réseau, des premiers résultats prometteurs ont été obtenus. La **Figure 22** compare un signal simulé par le métamodèle avec un signal expérimental issu de mesures réelles, démontrant une bonne adéquation entre les prédictions du modèle et l'expérience. La comparaison des prédictions obtenues avec ce réseau de neurones avec les modèles empiriques et/ou phénoménologiques existants, montre une bonne cohérence et met en avant une grande performance de cet outil, globalement plus précis que tous les autres.

Néanmoins, les cas testés restent proches des données d'entraînement et il est encore difficile de se prononcer sur les capacités réelles de généralisation de ce modèle. Également, le non-respect de certaines tendances physiques, telles que l'écart entre pression atteinte dans l'enceinte et pression d'ouverture de l'évent qui devrait toujours être positif, a été mis en évidence.

Ces travaux sont donc poursuivis afin d'optimiser le réseau de neurones sur lequel est basé le modèle, mettre à jour



la base de données (qui avec l'attrait actuel pour l'hydrogène s'est considérablement étoffée) et l'intégration de contraintes physiques.

Au-delà de la mise en place de ce métamodèle, des recherches sont également en cours pour exploiter l'intelligence artificielle afin d'analyser les séquences d'images issues de vidéos d'essais de l'Ineris. En effet, ils mettent en évidence des phénomènes explosifs, caractérisés par des durées et des effets externes souvent extrêmes et difficiles à mesurer. L'imagerie rapide se révèle être une méthode particulièrement intéressante pour étudier ces phénomènes, car elle permet d'enregistrer localement ou globalement les explosions ou leurs conséquences de

manière non intrusive avec une cadence d'acquisition et une durée d'enregistrement correspondant, dans une certaine mesure, aux temps caractéristiques des explosions. Les travaux menés visent par exemple à mesurer les vitesses caractéristiques des éléments projetés ou des nuages à partir de ces enregistrements vidéo. Un autre enjeu concerne le traitement des images enregistrées dont la qualité est très variable (en fonction notamment de la nature de l'essai réalisé et des conditions extérieures).

31. Y. Grégoire, J. Daubech, C. Proust, E. Leprette, "Vented gas explosion overpressure calculation based on a multi-layered neural network", *Journal of Loss Prevention in the Process Industries* 74 (2022) 104641.

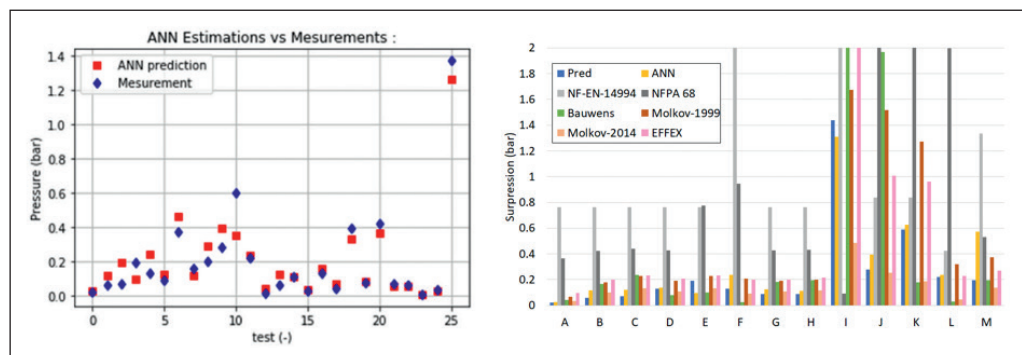


Figure 22

Performances du métamodèle en comparaison des données de pressions (pression) mesurées lors des essais (à gauche) et des modèles phénoménologiques existants<sup>31</sup>.

## Conclusion

L'intelligence artificielle et les nouvelles approches méthodologiques donnent accès à des outils et des méthodes aux potentiels indéniables dans les différents secteurs d'activité de l'Ineris, et trouvent peu à peu leur place en complément des outils expérimentaux et de modélisation plus classiques.

Ces développements sont conduits en parallèle d'une stratégie de digitalisation des process et des plateformes d'essais de l'Ineris, et ne se limitent pas au développement de modèles prédictifs. Il s'agit également de disposer d'outils et de méthodes de travail améliorés, plus rapides et plus efficaces, dans nos travaux expérimentaux et de modélisation permettant de se focaliser sur les aspects les plus importants et complexes de l'évaluation des dangers et des risques.

Les démarches actuelles vont permettre d'optimiser les campagnes expérimentales, pour ne réaliser que les essais les plus intéressants, notamment lorsqu'ils nécessitent une logistique importante et en prenant en compte la dangerosité des essais.

Ces démarches visent également à capitaliser au mieux toute l'expérience disponible de manière étendue, à partir des données d'essais, mais aussi en exploitant le retour d'expérience de l'accidentologie (pour identifier les signaux faibles et des risques émergents par exemple).

Au-delà de l'exploitation de ces approches pour les besoins de l'Ineris (tel que présenté ici), l'Ineris s'intéresse aux enjeux de sécurité posés par ces nouvelles technologies et notamment la question de la certification des systèmes et barrières de sécurité utilisant l'intelligence artificielle ou celle de cybersécurité qu'ils pourraient poser.

### Remerciements

Guillaume Fayet remercie Azziz Assoumani, Yann Gregoire et Jean-Bernard Kazmierczak pour leurs contributions respectives à cette présentation.