Scénarios stochastiques, données textuelles et IA pour une meilleure quantification du risque cyber

Hugo RAPIOR

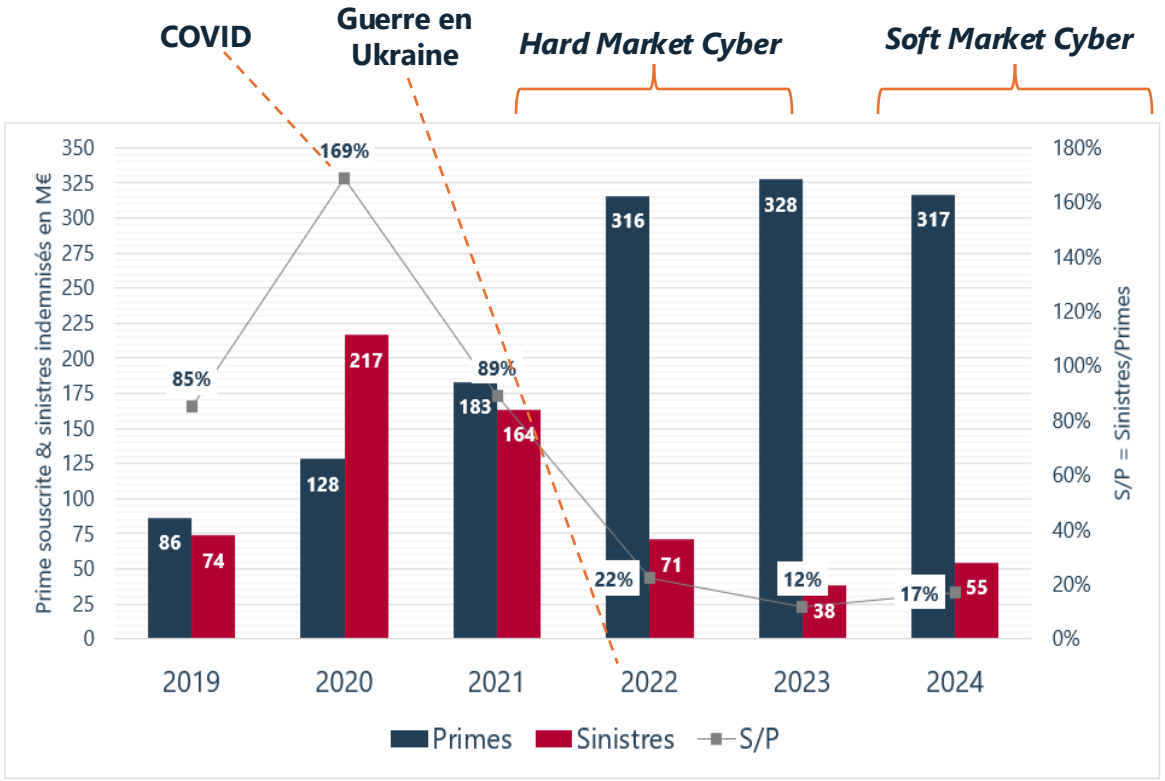
Caroline HILLAIRET

Chadi HANNA

Franck DOUNTIO

Etat du Marché de la Cyberassurance

État du marché



Historique

2020/2021 : Hard Market - Les assureurs supportent de lourdes pertes sur un risque très volatile et méconnu

2021/2022 : Réaction des assureurs/réassureurs :

- Hausse du coût de la réassurance
- Hausse des primes
- Baisse des capacités
- Revue des textes
- Revue de la politique de souscription (notamment mise en place de prérequis)

2022/2023 : Transition vers un Soft Market - Produit très rentable

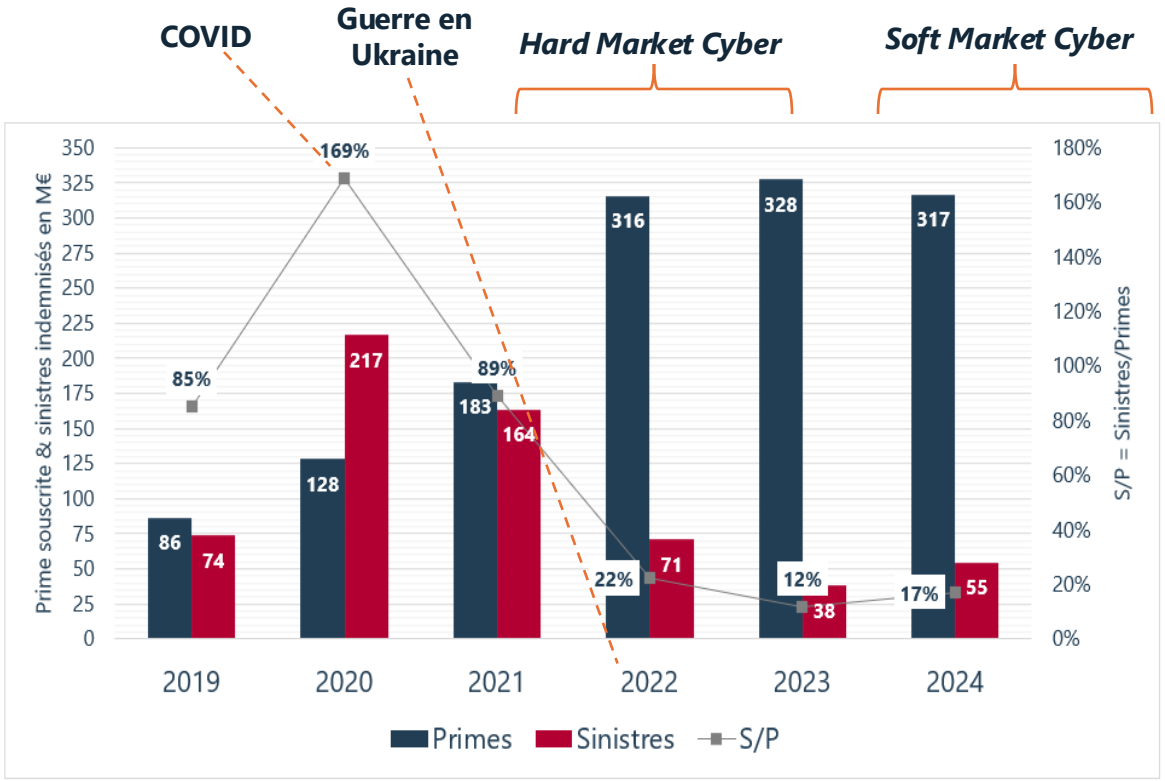
2023/2024 : Soft Market – Les assureurs regagnent de l'appétit pour ce risque :

- Baisse des primes
- Hausse des capacités
- Amélioration des textes
- Politiques de souscription beaucoup plus souples

2024/2025 : Soft Market – Marché très compétitif

- Retour à des niveaux de primes plus faibles que 2020/2021

État du marché



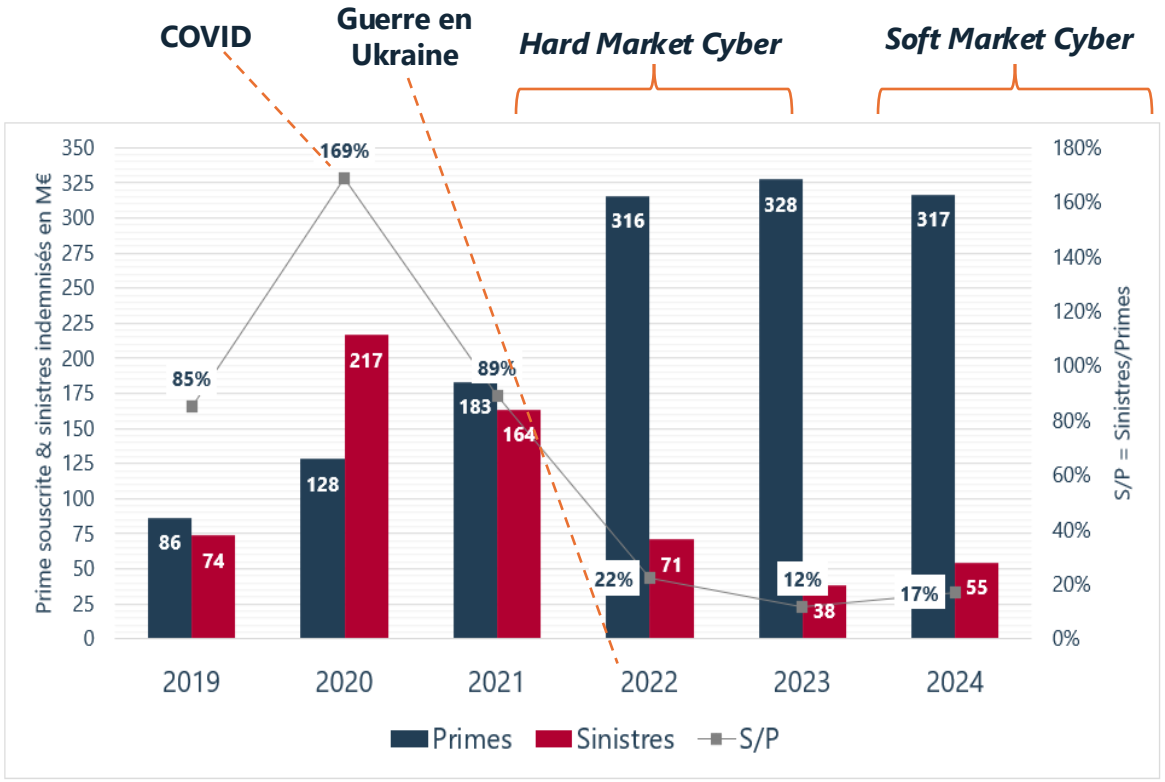
Question

Contexte :

- Les primes moyennes de 2024 sont plus faibles que les primes moyennes de 2020/2021 ;
- Les textes sont plus couvrants ;
- Les plafonds sont plus importants.

→ **Comment expliquez-vous que le loss ratio 2024 (17%) soit si inférieur à ceux de 2020/2021 (167%/89%) ?**

État du marché



Question

Contexte :

- Les primes moyennes de 2024 sont plus faibles que les primes moyennes de 2020/2021 ;
- Les textes sont plus couvrants ;
- Les plafonds sont plus importants.

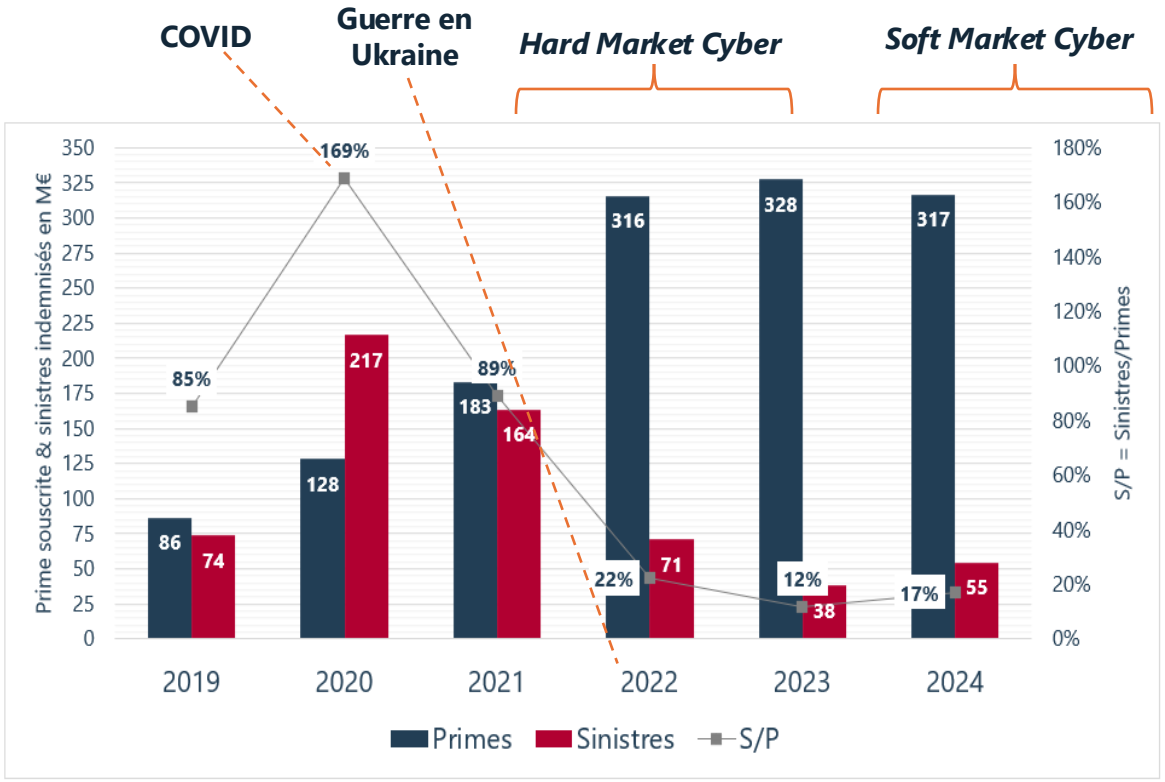
→ **Comment expliquez-vous que le loss ratio 2024 (17%) soit si inférieur à ceux de 2020/2021 (167%/89%) ?**

Réponse

Déformation du risque :

- **Fréquence** : Mise en place des prérequis (anti-virus, MFA, etc…) & Prévention
- **Coût moyen** : Mise en place des prérequis (sauvegardes déconnectées) & professionnalisation de la réponse à incident.

État du marché



Question au public

Contexte :

- Les primes moyennes de 2024 sont plus faibles que les primes moyennes de 2020/2021 ;
- Les textes sont plus couvrants ;
- Les plafonds sont plus importants.

→ **Comment expliquez-vous que le loss ratio 2024 (17%) soit si inférieur à ceux de 2020/2021 (167%/89%) ?**

Réponse

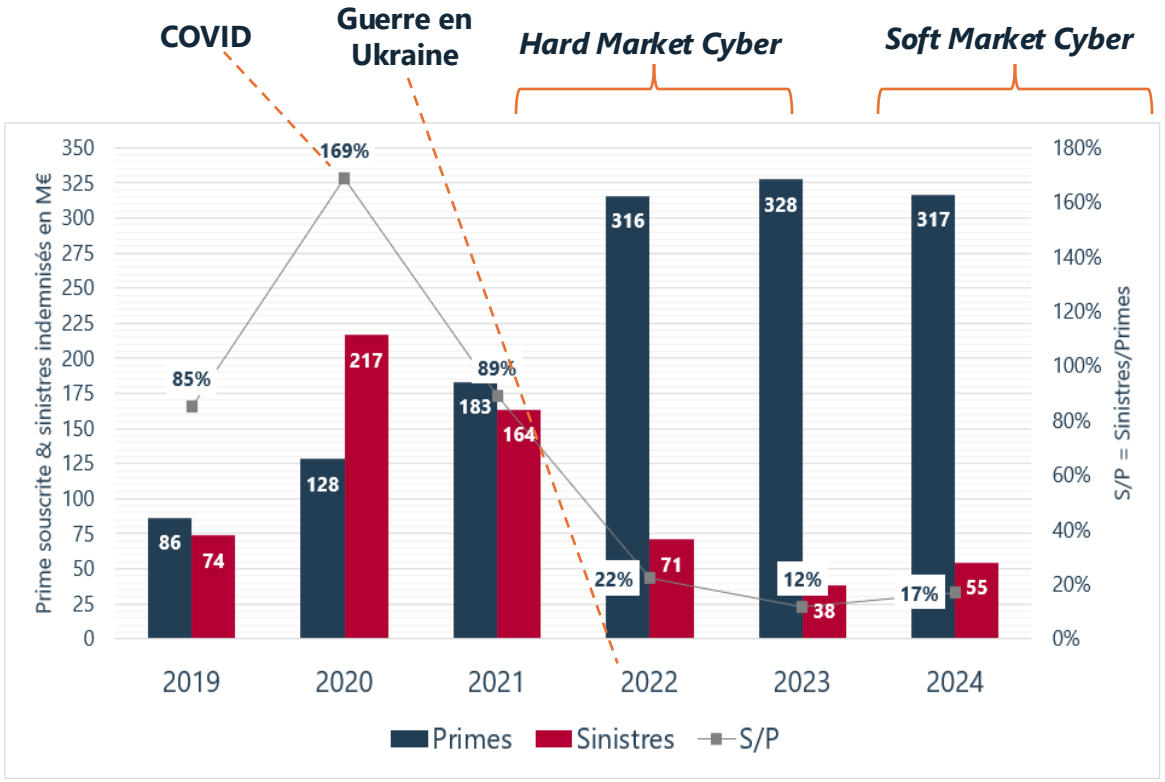
Déformation du risque :

- **Fréquence** : Mise en place des prérequis (anti-virus, MFA, etc…) & Prévention
- **Coût moyen** : Mise en place des prérequis (sauvegardes déconnectés) & professionnalisation de la réponse à incident.

Conclusions :

- **Le risque n'est plus le même d'un point de vue fréquence mais surtout d'un point de vue sévérité**

État du marché



Question au public

Contexte :

- Les primes moyennes de 2024 sont plus faibles que les primes moyennes de 2020/2021 ;
- Les textes sont plus couvrants ;
- Les plafonds sont plus importants.

→ **Comment expliquez-vous que le loss ratio 2024 (17%) soit si inférieur à ceux de 2020/2021 (167%/89%) ?**

Réponse

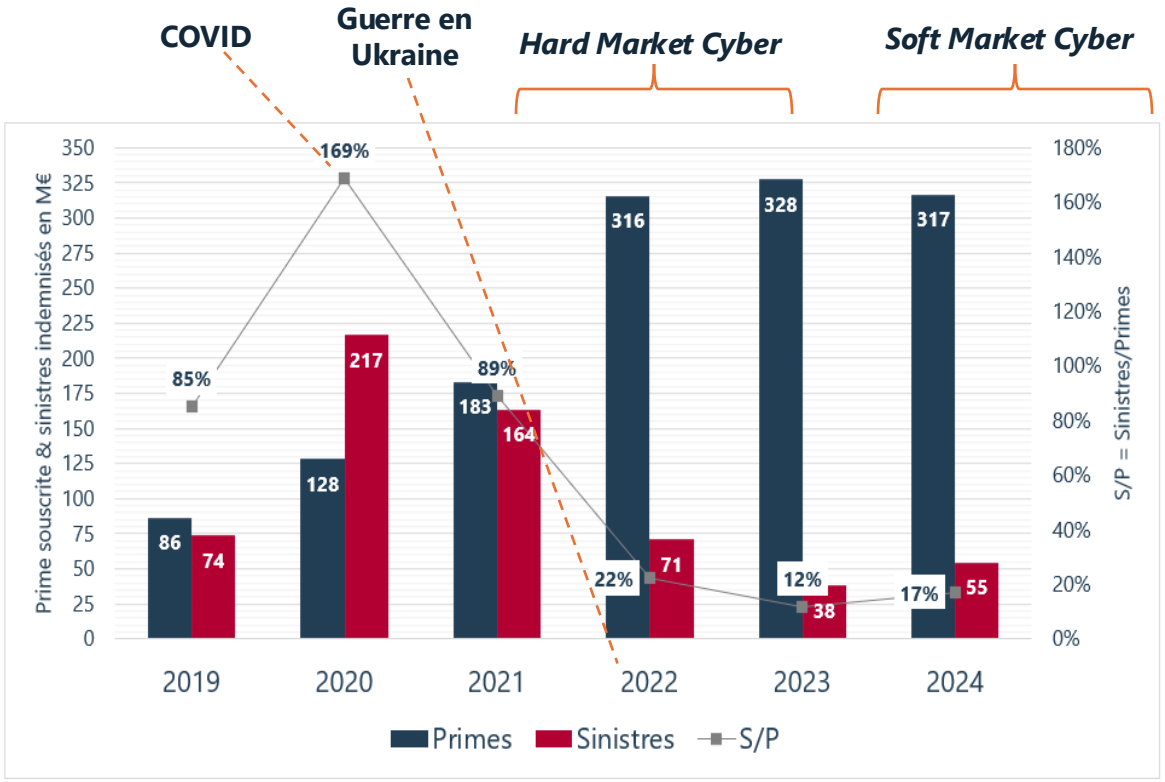
Déformation du risque :

- **Fréquence** : Mise en place des prérequis (anti-virus, MFA, etc…) & Prévention
- **Coût moyen** : Mise en place des prérequis (sauvegardes déconnectés) & professionnalisation de la réponse à incident.

Conclusions :

- **Le risque n'est plus le même d'un point de vue fréquence mais surtout d'un point de vue sévérité**
- **Le risque est rentable et "maîtrisé**

État du marché



Question au public

Contexte :

- Les primes moyennes de 2024 sont plus faibles que les primes moyennes de 2020/2021 ;
- Les textes sont plus couvrants ;
- Les plafonds sont plus importants.

→ **Comment expliquez-vous que le loss ratio 2024 (17%) soit si inférieur à ceux de 2020/2021 (167%/89%) ?**

Réponse

Déformation du risque :

- **Fréquence** : Mise en place des prérequis (anti-virus, MFA, etc…) & Prévention
- **Coût moyen** : Mise en place des prérequis (sauvegardes déconnectés) & professionnalisation de la réponse à incident.

Conclusions :

- **Le risque n'est plus le même d'un point de vue fréquence mais surtout d'un point de vue sévérité**
- **Le risque est rentable et "maîtrisé**
- **Le marché est cyclique donc il est nécessaire de rester vigilant**
- **Le risque est très hétérogène avec une forte composante systémique & extrême**

Caractéristiques du risque cyber

Risque émergent et non stationnaire

- **Risque relativement récent et en constante évolution** : très difficile à anticiper, nécessité de réagir très rapidement.
  - Développement rapide de nouveaux outils numériques (Gen IA)
  - Adaptation rapide des attaquants (en cas de cybermalveillant)
  - Évolution dans le temps des mesures de protection et des comportements de reporting, en raison de la réglementation et de l'évolution de la perception du risque.
- **Risque difficile à modéliser** car il nécessite une **bonne compréhension du comportement des différents acteurs** (utilisateurs et acteurs malveillants).
- **Très grande variance** due à :
  - Risque très volatile en lui-même
  - Risque très hétérogène (types d'événements et leurs conséquences, victimes…)
  - Imprécision des estimations statistiques car **peu de données disponibles** : biais, hétérogénéité dans les signalements, manque d'informations sur l'exposition dans les bases de données publiques.
- **Risque difficile à mesurer/quantifier et à anticiper** (même à relativement court terme).

Caractéristiques du risque cyber

Risque émergent et non stationnaire

- **Événements extrêmes** (pouvant entraîner des pertes considérables) : composante catastrophique.
  - Distributions à queue lourde (Théorie des Valeurs Extrêmes)

- **Risque d'accumulation** : composante systémique.
  - Dépendance entre les assurés
  - Effets de contagion
  - Concentration potentielle des incidents entraînant une perte de mutualisation

Ces caractéristiques peuvent mettre en danger la mutualisation des risques.

Approches Stochastiques pour une Quantification du Risque Cyber

Contagion et Clustering

Contagion et monitoring haute fréquence de la composante fréquence du risque

- **Modèles dynamiques** pour prendre en compte :
  - **Changements stochastiques** dans l'intensité d'arrivée des événements
  - **Autocorrélation** du nombre de cyber-événements

- **Modèle de Hawkes (auto-excitant)** pour capter les phénomènes de clusters (cf. [BBH21], [BCH25])
  - Modèle auto-excitant avec intensité stochastique, entièrement spécifiée par le processus ponctuel lui-même
  - Chaque événement augmente la probabilité qu'un nouvel événement se produise
  - Intégration de chocs exogènes (divulgation de vulnérabilités)

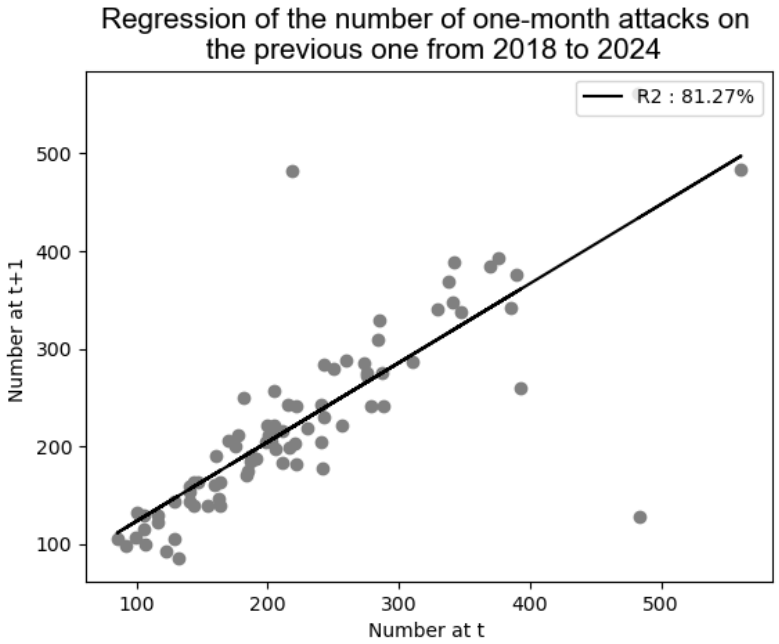
- **Modèle en temps continu** bien adapté pour un **monitoring haute fréquence du risque**.

Contagion et Clustering

Autocorrélation du nombre d'événements cyber

- Base de données Hackmageddon, gérée par Paulo Passeri (plus de 19 000 événements entre 2018 et 2024)

- Régression du nombre d'événements sur un mois $(t + 1)$ par rapport au mois précédent t (indépendance si modèle de processus de Poisson valide)



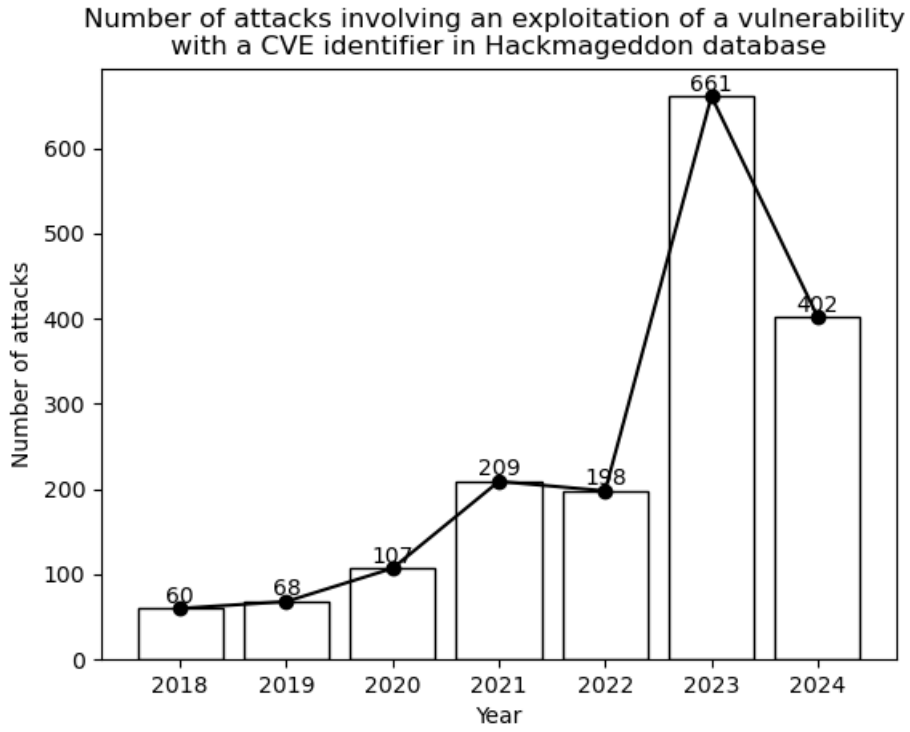
- Même observation sur les autres bases cyber (PRC database)

Contagion et Clustering

Excitation exogène

- Les clusters d'événements ne sont pas seulement dus à la contagion, mais aussi à l'arrivée de **chocs externes** tels que la **divulgation de vulnérabilités**.

- Dans la base de données Hackmageddon : nombre de cyberattaques impliquant une vulnérabilité avec un identifiant **CVE**.



Contagion et Clustering

Processus de Hawkes avec intensité de base stochastique

- **Processus de Hawkes** caractérisé par l'intensité stochastique (hazard rate) $s\lambda$:
  - **Intensité de base stochastique** :
    - une composante μ_0 déterministe
    - une composante exogène avec **noyau d'excitation externe $\bar{\phi}$ (vulnérabilité)**
  - **Noyau d'excitation interne ϕ (contagion)**

$$\lambda(t) = \underbrace{\mu_0(t)}_{partie\ déterministe} + \underbrace{\sum_{\bar{\tau}_k < t} \bar{\phi}\left(t - \bar{\tau}_k,\ \bar{Y}_{\tau_k}\right)}_{Choc\ externe} + \underbrace{\sum_{\tau_n < t} \Phi\left(t - \tau_n,\ \boldsymbol{Y}_{\boldsymbol{\tau_n}}\right)}_{Auto\text{-}excitation}$$

- **Marques** $\{\bar{Y}_k\}_{k\geq 1}$ **et** $\{Y_k\}_{k\geq 1}$(vulnérabilités et contagion) : indicateurs de sévérité modulant les noyaux d'excitation
  - **Pour les vulnérabilités** : CVSS score
  - **Pour les attaques** : caractère contagieux

Construction de scénarios stochastiques d'accumulation

Scénarios les plus préoccupants selon la Geneva association

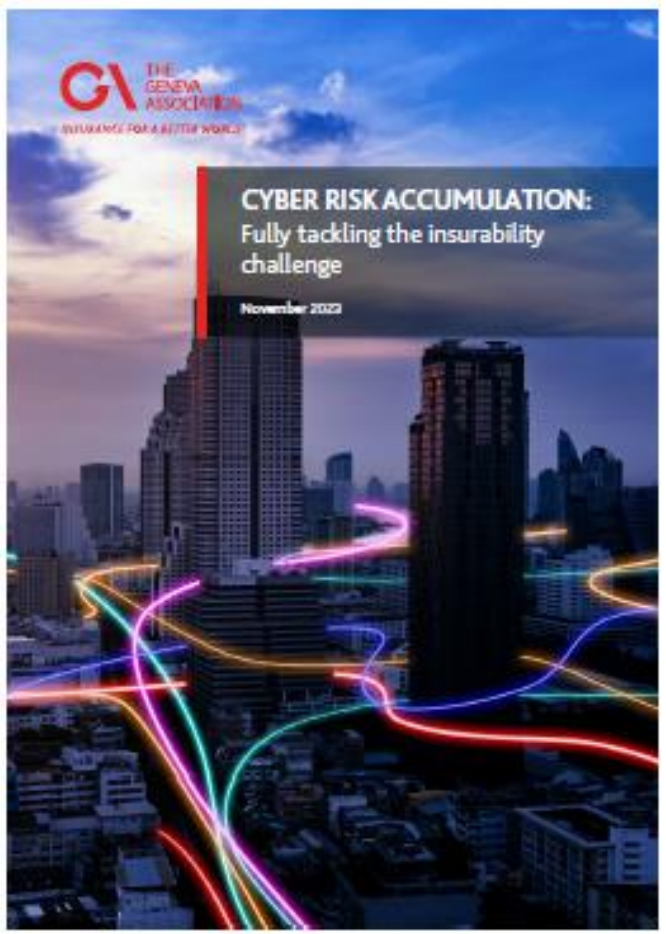


TABLE 3: RE/INSURERS' RANKING OF EXTREME CYBER SCENARIOS

Extreme cyber scenarios	Average ranking of scenario
Denial of service/interruption of operations	
Worm-like malware epidemic	1
Widespread ransomware attack	2
Mass data breach	
Exfiltration of sensitive information (PII, encrypted passwords, etc.) at key organisation/institution which has widespread effects on customers/suppliers	4
Disruption to critical infrastructure	
An extortion of supervisory control and data acquisition (SCADA) networks of industrial control systems	4
A cyberattack on a crucial participant in an industry/sector (e.g. hospital, food manufacturer/distributor, etc.)	5
A cyberattack on a key utility provider (power, water etc.)	2
A compromise of state/municipal services	5
Cross-sector IT failure	2

Refers to median ranking score assigned by survey respondents (1 being the highest-ranked scenario). Based on the results from a poll of 11 GA member cyber re/insurers

Source: The Geneva Association

Construction de scénarios stochastiques d'accumulation

Accumulation : élaboration de scénarios

- **Deux types de scénarios d'accumulation :**
  - **déterministe** (nécessite une définition précise de l'événement)
  - **stochastique** (plus approximatif mais permettant d'explorer de nombreuses situations)

- **Élaboration de scénarios d'accumulation stochastiques avec réseau** (cf. [HL21], [HL22]).
  - cadre général de conception de scénarios stochastiques d'accumulation, en utilisant l'adaptation de modèles épidémiologiques au cyber-risque : « cyber-pandémie »
  - Spécificité par rapport à une pandémie biologique : modélisation de la stratégie d'attaque, de la stratégie de prévention/réaction
  - **Quelques applications typiques** : tester robustesse du portefeuille, évaluer son degré de diversification, quantifier l'impact de la prévention, évaluer le risque de saturation.

- **Difficulté du calibrage.**

Construction de scénarios stochastiques d'accumulation

Modèles de contagion avec effets de réseau

- **Modèle SIR** (Susceptible–Infecté–Retiré) multi-groupes avec différentes sous-populations.

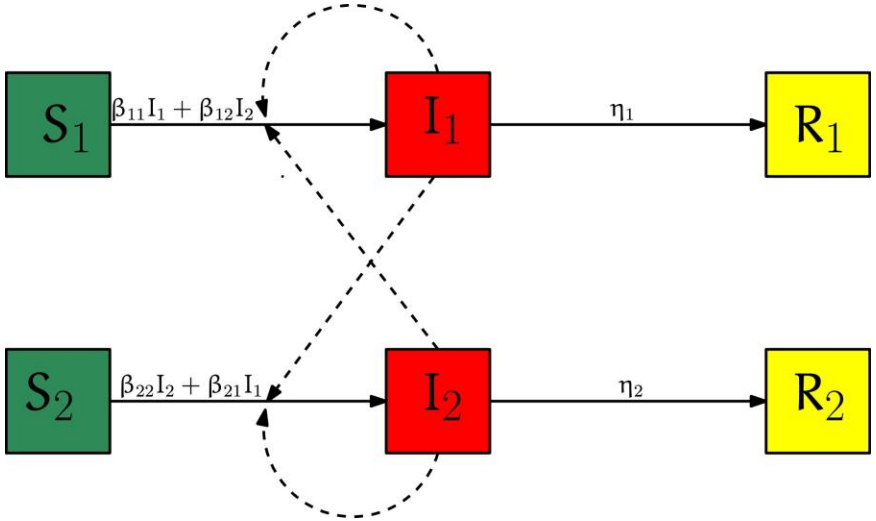


Figure tirée de Magal et al. (2018)

- $\mathcal{B} = (\boldsymbol{\beta}_{i,j})_{1 \leq i,j \leq K}$ matrice des taux d'infection : $\beta_{i,j}$ matérialise la manière dont j contamine i.

Construction de scénarios stochastiques d'accumulation

Modèle SIR Multi-groupes

- SIR Multi-groupes avec K **sous-groupes :** $1 \leq i \leq K$

$$\frac{dS_i(t)}{dt} = -\eta_i(t)\left(\alpha_i(t) + \sum_{j=1}^{K} \beta_{i,j}\, I_j(t)\right) S_i(t)$$

$$\frac{dI_i(t)}{dt} = \eta_i(t)\left(\alpha_i(t) + \sum_{j=1}^{K} \beta_{i,j}\, I_j(t)\right) S_i(t) - \gamma_i I_i(t)$$

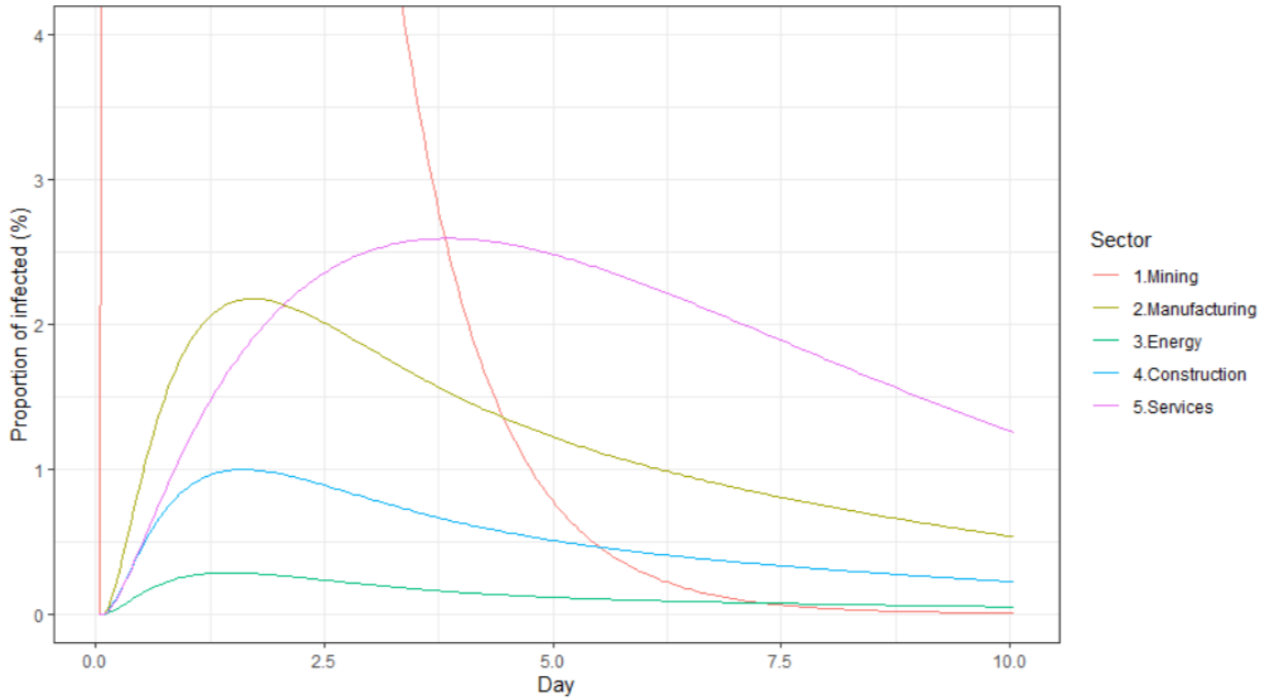
$$\frac{dR_i(t)}{dt} = \gamma_i I_i(t)$$

- $\mathcal{B} = (\beta_{i,j})_{1 \leq i,j \leq K}$ matrice des taux d'infection.
- $\alpha_i(t)$ représente l'intensité des attaques dans la classe i.
- $\eta_i(t)$ représente la manière dont la classe i tend à se protéger et prend ses valeurs dans [0, 1].

Construction de scénarios stochastiques d'accumulation

Exemple de dynamique cyber de type Wannacry

Évolution de la proportion d'infections – Attaque sur secteur mining.



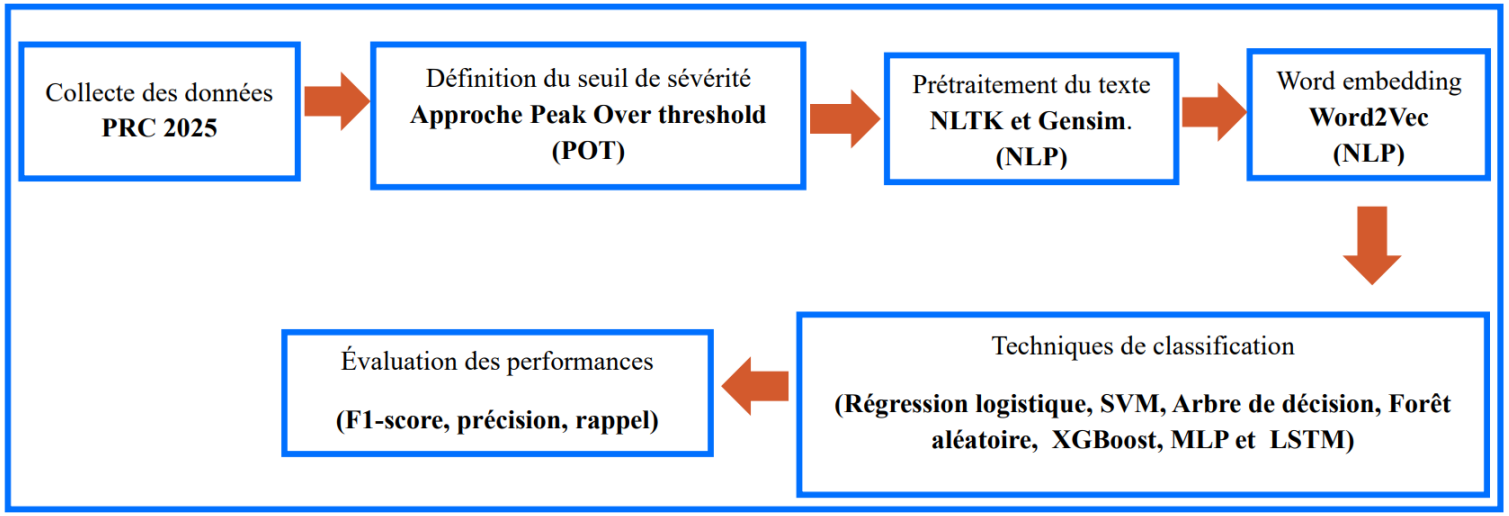
La valeur maximale pour le secteur mining est de 70 % (après 10 heures).

L’IA et le NLP pour la Modélisation de la Sévérité des Incidents Cyber

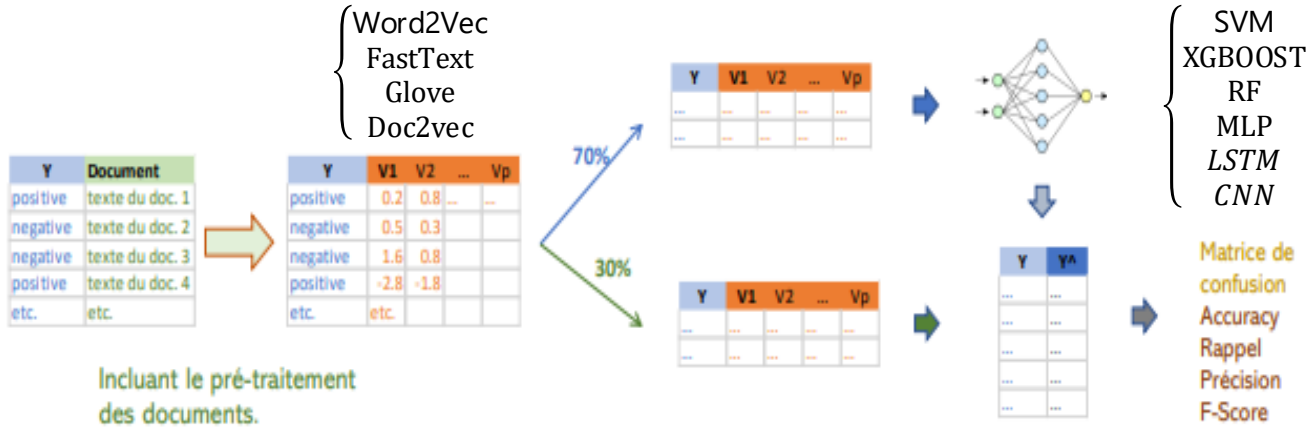
Évaluation de la sévérité des incidents cyber de type *data breach* à partir des descriptions textuelles

Méthodologie générale

On crée des *embeddings* à partir des rapports textuels d'incidents de la base PRC grâce aux techniques de NLP, puis on peut appliquer sur ces vecteurs des **algorithmes de Machine Learning (ML)** et de **Deep Learning (DL)** (classification d'incidents, détection de menaces, clustering de rapports, etc.).



EXPRESSIONS RÉGULIÈRES (*RegEx*)
(Reclassifier les incidents à la sortie des modèles de ML et de DL pour réduire le nombre de faux négatifs et de faux positifs)



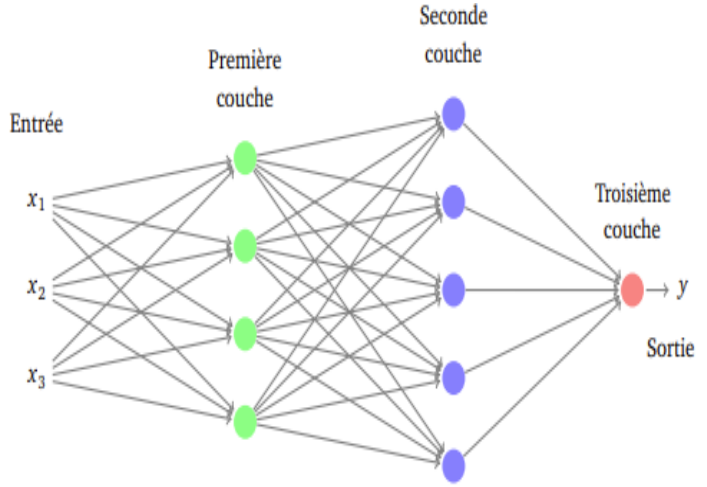
Le Deep Learning

MLP (Perceptron Multicouche) et LSTM (Long Short-Term Memory)

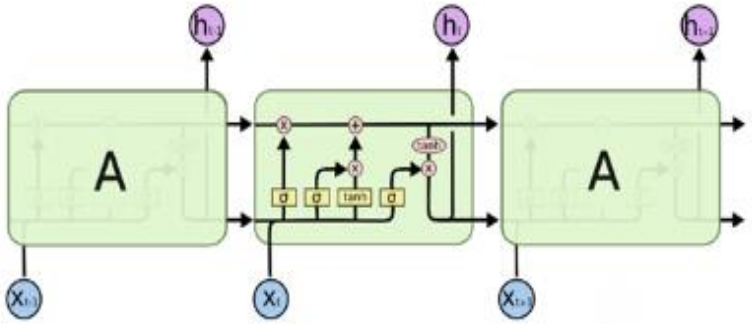
1. MLP – Multi-Layer Perceptron

Idée clé : imite un neurone biologique pour apprendre des relations entre les données d'entrée et la sortie.

- **Entrées** : $x_1, x_2, \ldots, x_n$,chacune associée à un **poids** a_i plus un biais.
- **Activation** : fonctions non linéaires (sigmoïde, ReLU, tanh…).
- **Architecture** :
  - **Couche d'entrée** : reçoit les données (ex. texte, vecteurs).
  - **Couches cachées** : apprennent des représentations complexes et non linéaires.
  - **Couche de sortie** : produit la prédiction finale y (grave vs Attritionnel).



Schématisation d'un MLP (Tavenard, 2023)



Structure simplifiée d'une cellule LSTM (Olah 2015)

2. LSTM – Long Short-Term Memory : une solution au problème de mémoire

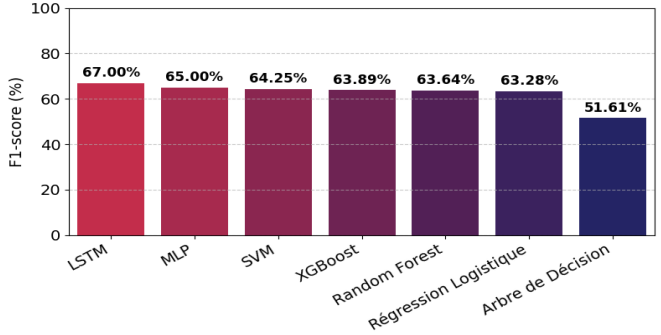
Idée clé : résout le problème des dépendances longues (vanishing gradient) dans les RNN.

- **RNN standard** : chaque sortie dépend de l'entrée courante et de l'état précédent, mais **oublie le long terme**.
- **LSTM amélioré (Hochreiter & Schmidhuber, 1997)** :
  - Introduit une **cellule mémoire** C_t qui transporte l'information.
  - Utilise **3 portes** :
    - **Forget gate** : décide ce qu'on oublie.
    - **Input gate** : décide ce qu'on ajoute.
    - **Output gate** : décide ce qu'on transmet.
- **Avantage** : apprend un **contexte long terme** sans perte d'information.

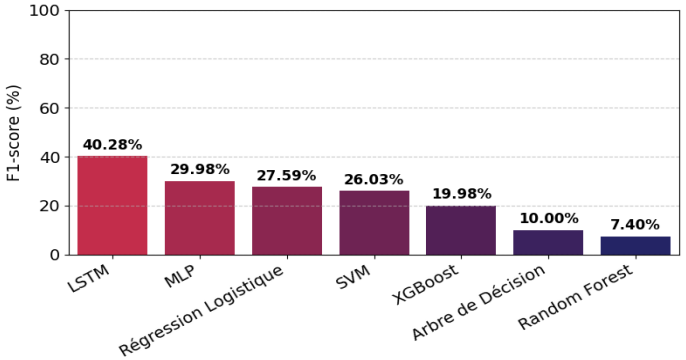
Résultats de la modélisation

Résultats MLP et LSTM

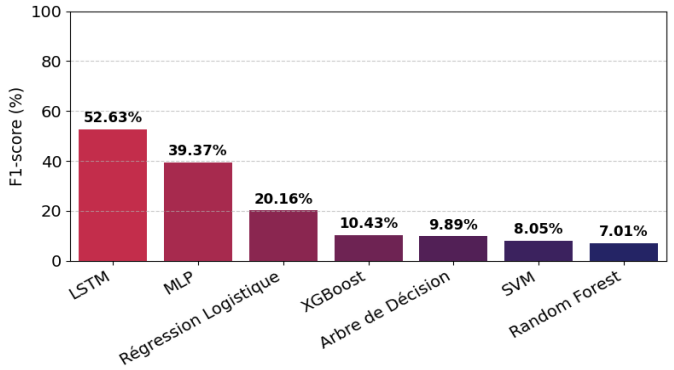
Analyse des résultats au seuil de sévérité du 60e percentile



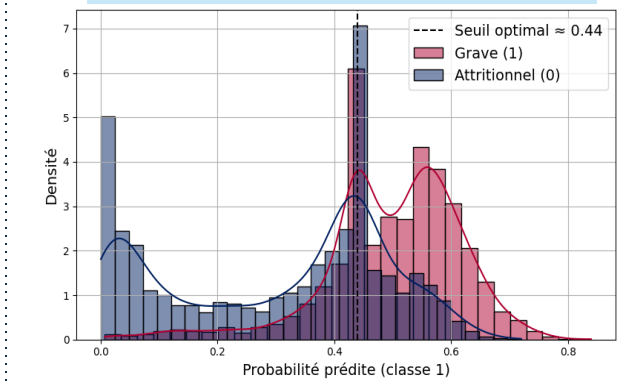
Analyse des résultats au seuil de sévérité du 95e percentile

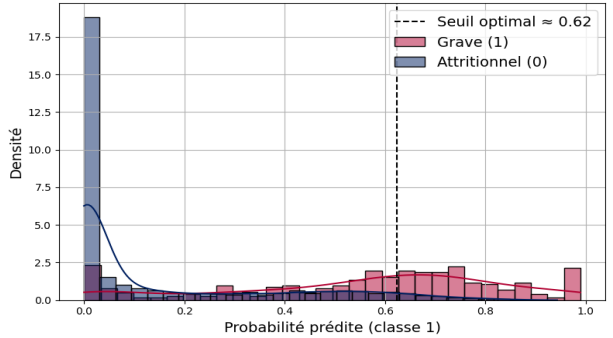


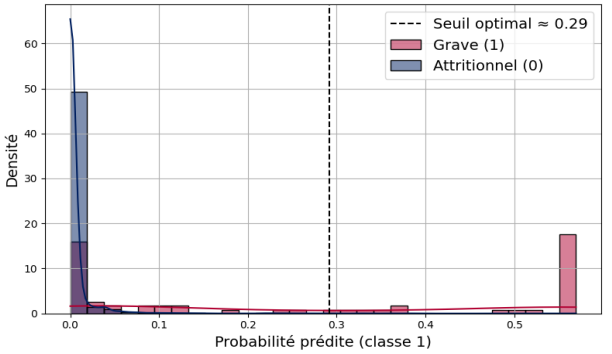
Analyse des résultats au seuil de sévérité du 99e percentile



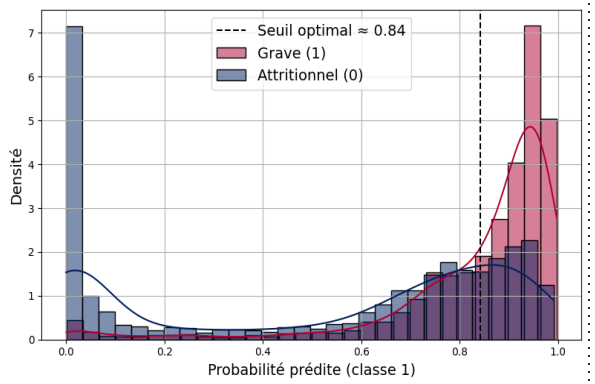
Résultats du LSTM

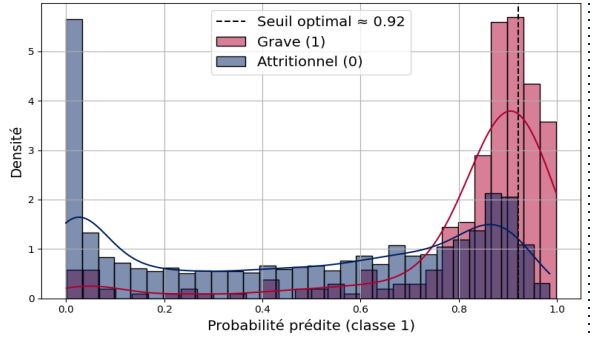


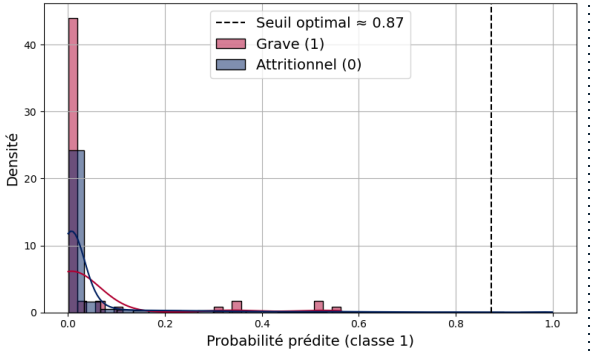




Résultats du MLP







L'IA générative au service de la prédiction

L'intuition de base : le bootstrap

Idée : repartir du concept du Bootstrap

Le Bootstrap : pour une fonctionnelle linéaire (comme la moyenne), l'estimateur bootstrappé converge asymptotiquement sans biais vers la vraie moyenne. Pour la moyenne, la variance de l'estimateur bootstrappé converge vers la variance empirique ,ce qui garantit la convergence asymptotique sans biais.

En clair : Pour un échantillon de données $[10; 40; 100; 200]$, la moyenne empirique est de 70

Si on Bootstrap cet échantillon suffisamment de fois, et qu'on en fait la moyenne, on convergera aussi vers 70

Est-ce toujours vrai pour du texte ?

Intuitivement, un LLM (ChatGPT, etc…) pourrait être assimilé à un très gros système de Bootstrap ?
Puisqu'il réutilise des mots provenant de sources sur lesquelles il s'est entrainé

« Est-ce que la propriété de convergence asymptotique sans biais du Bootstrap est conservée lorsqu'on utilise du texte généré par LLM ? »

L'IA générative au service de la prédiction

Méthode de génération de descriptifs de sinistres

Idée : Définir une approche (deux en réalité) permettant de correctement simuler des descriptifs d'incidents cyber
Deux possibilités : Laisser complètement faire le LLM, ou l'aider un peu (non-paramétrique v.s. paramétrique)

Approche Commune

On fournit des exemples de sinistres graves et attritionnels de la base PRC (25 de chaque)

Non-paramétrique

On demande au LLM de générer des descriptifs de sinistres graves, sans contrainte supplémentaire dans la génération

Paramétrique

1) On simule au préalable des réalisations de variables clés (type d'organisation, type de brèche,…)
2) On demande au LLM de générer des descriptifs de sinistres graves, en tenant compte pour chaque description des valeurs simulées des variables clés

Dans les deux approches, beaucoup d'itérations ont été nécessaires afin de converger vers un prompt qui produisait des résultats satisfaisants (nécessité de réaliser du *prompt engineering*)

L'IA générative au service de la prédiction

Exemples de sinistres générés

Incidents réels

{"organization_type": "BSR",

"breach_type": "CARD",

"incident_details": "The Wisconsin Department of Justice reported on December 19, 2013, that Target Corp. experienced a data breach affecting approximately 40 million customers. The breach occurred between November 27 and December 15, 2013, involving the theft of information from credit and debit cards, including customer names, card numbers, expiration dates, and CVV codes.",

"niveau_gravite_99": "grave"}

{"organization_type": "BSR",

"breach_type": "UNKN",

"incident_details": "The Indiana Office of the Attorney General reported a data breach involving The Neiman Marcus Group on September 30, 2021. The breach occurred on May 2, 2020, affecting a total of 4,354,346 individuals, including 31,992 residents.",

"niveau_gravite_99": "grave"}

Incidents générés

{"organization_type": "BSO",

"breach_type": "HACK",

"incident_details": "The California Office of the Attorney General reported that a natiaonal online services provider experienced a network intrusion attributed to ransomware between April 3 and April 8, 2024. Forensic review indicates unauthorized access to systems containing names, Social Security numbers, and authentication data for approximately 6.1 million individuals across multiple states. The actor exfiltrated files before encrypting servers.",

"niveau_gravite_99": "grave"}

{"organization_type": "MED",

"breach_type": "HACK",

"incident_details": "The U.S. Department of Health and Human Services reported that a multi-state clinic network was attacked through a supply-chain exploit on May 9, 2024. Protected health information for 2.9 million patients, including medical histories, was compromised.",

"niveau_gravite_99": "grave"}

L'IA générative au service de la prédiction

Métriques d'évaluation du « biais » et de la « variance »

1. Jensen–Shannon Divergence (JSD)

Objectif : comparer les distributions de probabilité entre données réelles et simulées

Évalue la similarité entre :

- La distribution des variables catégorielles (*ex.* organization_type, breach_type ou du vocabulaire présent)
- Le vocabulaire généré vs. le texte original.

$$JSD(P_{real}, P_{syn}) = \frac{1}{2} D_{KL}(P_{real} \,||\, M) + \frac{1}{2} D_{KL}(P_{syn} \,||\, M)$$

avec $M = \frac{1}{2}(P_{real} + P_{syn})$.

Interprétation : Si proche de 0 → distributions similaires, si proche de 1 → divergence forte

2. Biais de centre (1 – cos)

$$Biais_{centre} = 1 - cos(\mu_{real}, \mu_{syn})$$

Objectif : mesurer le décalage global entre le centre des embeddings réels et simulés.

μ= barycentre des vecteurs de mots.

Interprétation : 0.00–0.05 → très proches ; 0.05–0.10 → léger décalage 0.10 → biais notable (*contenu généré "à côté du réel"*)

3. MMD RBF (Maximum Mean Discrepancy – Radial Basis Function)

Objectif : comparer la forme complète des distributions d'embeddings (variabilité, multi-modalité), et Capter les écarts de forme qu'un simple centre ne détecte pas.

$$k(x, y) = e^{-\gamma \|x-y\|^2},$$

$$MMD^2 = E[k(x, x')] + E[k(y, y')] - 2E[k(x, y)]$$

Interprétation : $MMD^2 \approx 0$ → distributions très proches, 0.10–0.20 → écart modéré ; 0.20 → biais sémantique perceptible

.

L'IA générative au service de la prédiction

Evaluation du biais selon les différentes approches de prompt

Après avoir généré 400 sinistres graves selon les différentes approches présentées supra, les mesures de biais sont évaluées en comparant à la base Réel de 313 sinistres graves :

Métriques d'évaluation du biais / Techniques de prompt	JSD (sur la distribution organization_type)	JSD (sur la distribution breach_type)	Biais de centre	JSD lexical	MMD RBF
Génération Non paramétrique	0,021	0,016	0,0780	0,753	0,2364
Génération Non paramétrique avec avec la variable "information_affected_explanation"	0,013	0,017	0,0854	0,776	0,2522
Génération paramétrique	0,004	0,004	**0,0453**	**0,692**	**0,1635**
Génération paramétrique avec la variable "information_affected_explanation"	**0,000**	**0,001**	0,0686	0,787	0,2214

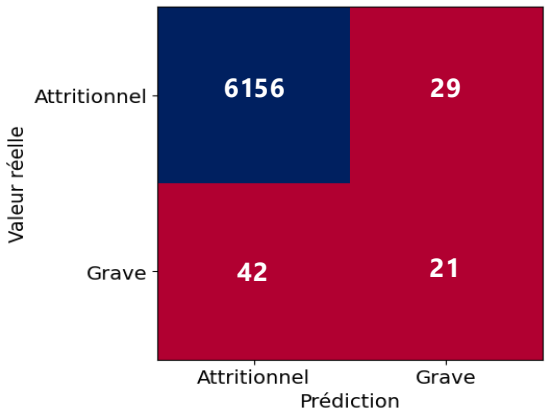
- Les résultats montrent que les **approches paramétriques** réduisent nettement l'ensemble des biais, traduisant une **meilleure similarité sémantique et structurelle** avec les sinistres réels.
- L'intégration des distributions statistiques dans le prompt améliore donc **la fidélité et la cohérence** des données synthétiques générées.

L'IA générative au service de la prédiction

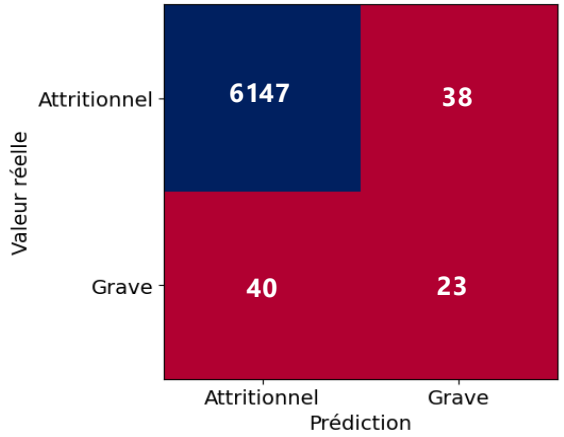
Résultats ajout des données synthétiques sur le modèle optimal de MLP obtenu
Au seuil de sévérité 99-ième percentile

Résultat après ajout de 500 sinistres attritionnels générés selon une approche paramétrique

$F1 - score$ = 0, 3937

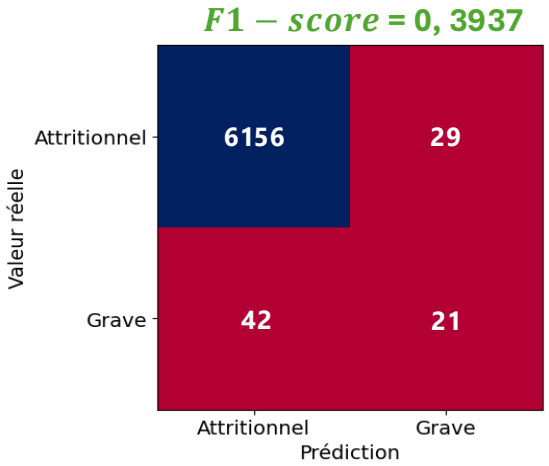


$F1 - score$ = 0, 3710

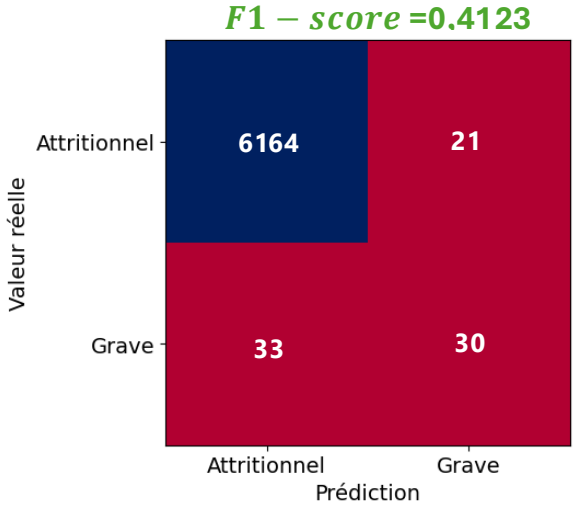


Résultat après ajout de 5000 sinistres graves générés selon une approche paramétrique

$F1 - score$ = 0, 3937



$F1 - score$ =0.4123



Le Deep Learning + RegEx

Amélioration des prédictions par les expressions régulières (RegEx)

Motivation : pallier les limites du Deep Learning

- Les réseaux neuronaux peinent à interpréter les expressions numériques dans le texte (ex. : *« 27 000 employés »*, *« 315 patients »*).
- Les embeddings (Word2Vec) séparent les nombres et les mots, perdant leur lien sémantique.
- L'apprentissage reste dépendant des exemples vus durant l'entraînement, limitant la généralisation sur de nouvelles valeurs.

Solution : combiner la puissance du LSTM avec une analyse complémentaire par expressions régulières (RegEx) pour capturer les motifs *nombre* × *mot*.

Principe de l'approche hybride

Décision finale = Prédiction neuronale (LSTM) OU Prédiction RegEx

- Le modèle RegEx extrait et agrège les combinaisons *nombre* × *mot* (ex. : *« 20 787 individuals »*).
- Un score RegEx est calculé pour chaque description :

$$\text{Score}_{\text{REGEX}}(D) = \sum_{u=1}^{n} N_u \cdot 1_{M_u \in V}$$

- L'incident est classé grave si

$$\text{Score}_{\text{REGEX}}(D) > \text{Seuil}_{\text{REGEX}}(k) \begin{cases} 2\,521 \; pour \; k = 60\% \\ 181\,820 \; pour \; k = 95\% \\ 2\,068\,450 \text{ pour k} = 99\% \end{cases}$$

où $k \in \{60,95,99\}$ correspond aux percentiles de sévérité.

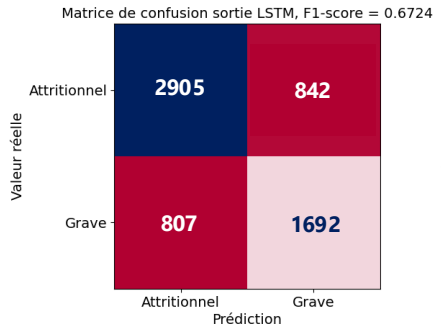
Méthodologie opérationnelle (pour chaque description D et niveau de sévérité k)

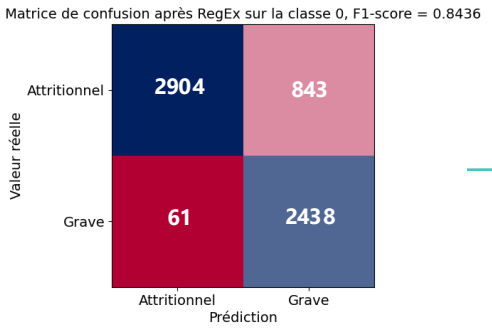
Étape	Description	Sortie
1. Prédiction neuronale	Prédiction_neuronale ← modèle.predict}(D)	Probabilité issue du LSTM
2. Analyse RegEx	Si D contient un motif « nombre × mot » : calculer $\text{Score}_{\text{REGEX}}$	Prédiction binaire RegEx
3. Décision finale	prédiction_finale = prédiction_neuronale OU prédiction_regex	Classification finale

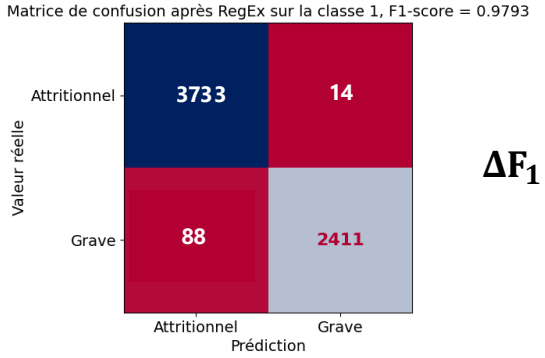
LES REGEX

Resultats (RegEx)

Résultats REGEX au seuil de sévérité du 60e percentile

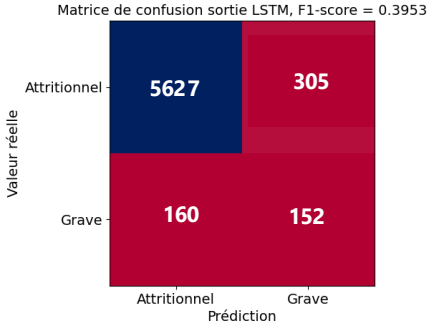


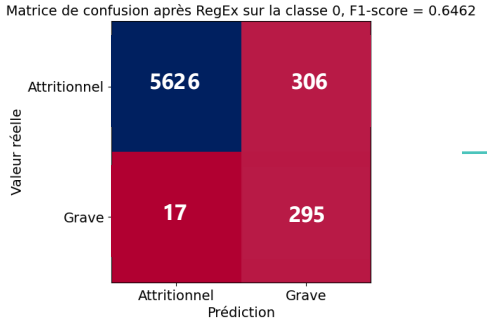


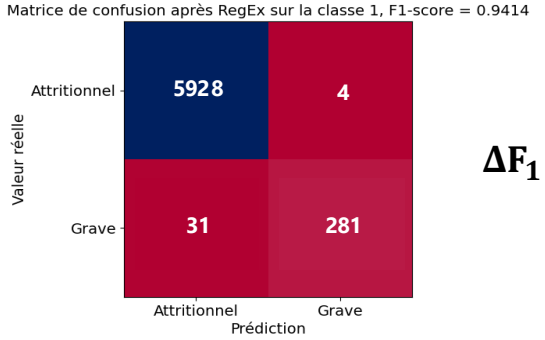


$$\Delta F_1 = F_1^{hybride} - F_1^{LSTM} = 0,3069$$

Résultats REGEX au seuil de sévérité du 95e percentile

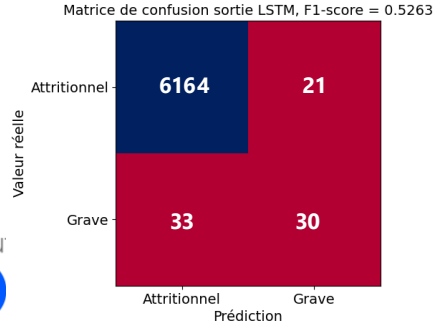


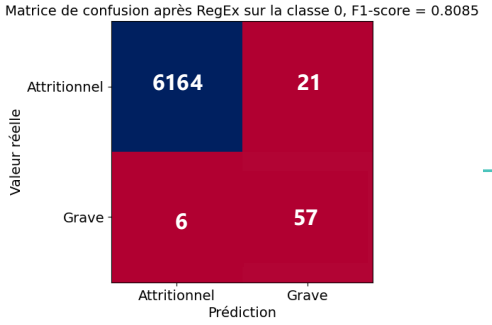


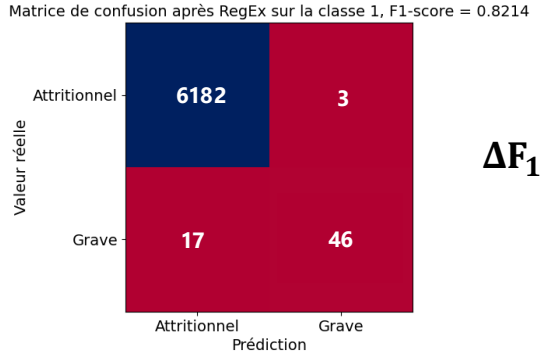


$$\Delta F_1 = F_1^{hybride} - F_1^{LSTM} = 0,5461$$

Résultats REGEX au seuil de sévérité du 99e percentile







$$\Delta F_1 = F_1^{hybride} - F_1^{LSTM} = 0,2951$$

Les Apports Clés de la Modélisation Avancée en Cyberassurance

Les apports clés de la modélisation avancée en Cyberassurance

Assurer la rentabilité technique du produit

Les modèles avancés permettent d'estimer plus finement la fréquence et la sévérité des sinistres, en tenant compte des événements extrêmes et des effets de contagion.
Cela garantit une tarification plus juste, évite les sous-provisionnements et préserve la performance technique du portefeuille.

Quantifier son exposition aux risques et assurer sa solvabilité

En simulant des scénarios catastrophes et des stress tests, l'assureur peut mesurer sa vulnérabilité à des chocs massifs. Ces analyses servent à dimensionner le capital réglementaire et à démontrer la solidité financière de l'entreprise face aux exigences de Solvabilité II.

Challenger le coût de la réassurance

Les réassureurs basent souvent leurs modèles sur des données américaines, surestimant les pertes attendues en Europe. Disposer de modèles internes robustes permet de négocier des traités plus adaptés et de **réduire le coût de la réassurance** sans dégrader la protection.

Structurer l'équipe de réponse à incident (CERT interne Dattak)

Les modèles prédictifs (comme les processus de Hawkes) aident à anticiper les vagues d'attaques et à prioriser les ressources internes. Ils permettent au CERT de Dattak d'adopter une approche proactive, en renforçant la veille, la prévention et la coordination lors des pics de sinistralité.

Modéliser pour assurer la rentabilité technique du produit

Assurer la rentabilité technique du produit

Les modèles avancés permettent d'estimer plus finement la fréquence et la sévérité des sinistres, en tenant compte des événements extrêmes et des effets de contagion.
Cela garantit une tarification plus juste, évite les sous-provisionnements et préserve la performance technique du portefeuille.

Contexte :

- Données insuffisantes et non représentatives du champ des possibles
- Risque de sinistres de masse (failles communes, attaques globales)
- Risque en constante évolution (géopolitique, technologique, réglementaire)
- Forte corrélation entre assurés (fournisseurs, logiciels, secteurs)

Pourquoi ces modèles sont essentiels pour assurer la rentabilité technique :

→ Pallier le manque de données pour la tarification et le provisionnement

- **Fréquence :** l'historique ne reflète pas le risque futur → Simuler des scénarios rares ou encore jamais observés
- **Sévérité :** la sinistralité passée sous-estime les pertes extrêmes → Mieux représenter la queue du risque (sinistres extrêmes)
- Intégrer la partie "invisible et évolutive" du risque

→ Mieux mesurer son exposition

- Identifier les zones d'accumulation (même fournisseur, même secteur)
- Ajuster la politique de souscription et les critères d'éligibilité

→ Renforcer la prévention pour réduire la sinistralité

- Exploiter les signaux faibles pour anticiper les vagues d'incidents
- Alerter les assurés les plus exposés en amont

Comprendre son exposition pour mieux se protéger

Quantifier son exposition aux risques et assurer sa solvabilité

En simulant des scénarios catastrophes et des stress tests, l'assureur peut mesurer sa vulnérabilité à des chocs massifs. Ces analyses servent à dimensionner le capital réglementaire et à démontrer la solidité financière de l'entreprise face aux exigences de Solvabilité II.

Contexte :

- Risque cyber très corrélé : un même événement peut toucher des centaines d'assurés
- Sinistres extrêmes difficiles à anticiper sans données suffisantes
- Forte incertitude sur les montants à régler (délai, évolution des coûts, litiges)
- Nécessité de démontrer la solidité financière (réglementation Solvabilité II)

Pourquoi ces modèles sont essentiels :

→ Mesurer précisément l'exposition au risque d'accumulation

- Identifier les zones de concentration (même fournisseur cloud, même secteur, même vulnérabilité)
- Quantifier la perte potentielle maximale pour le portefeuille global

→ Tester la résistance financière de l'assureur

- Simuler des scénarios catastrophes (faille logicielle mondiale, ransomware en chaîne)
- Évaluer la capacité de l'entreprise à absorber le choc sans défaillance

→ Dimensionner le capital et les couvertures de réassurance

- Ajuster le niveau de capital requis selon les stress tests
- Dimensionner et calibrer les traités de réassurance en fonction de l'exposition réelle
- Optimiser la structure de protection (seuils, priorités, plafonds) pour garantir la solvabilité à moindre coût
- Trouver le **meilleur équilibre entre coût, couverture et niveau de capital requis**.

Mieux connaître son risque pour mieux négocier

Challenger le coût de la réassurance

Les réassureurs basent souvent leurs modèles sur des données américaines, surestimant les pertes attendues en Europe. Disposer de modèles internes robustes permet de négocier des traités plus adaptés et de **réduire le coût de la réassurance** sans dégrader la protection.

Contexte :

- Les modèles des réassureurs reposent souvent sur des données américaines, beaucoup plus sinistrées que le marché européen.
- Cela conduit à une **surestimation du risque** et donc à des **traités coûteux** pour les assureurs européens.
- Les réassureurs disposent d'un fort pouvoir de négociation, surtout quand l'assureur n'a pas de modèle interne robuste.

Pourquoi ces modèles sont essentiels :

→ Mieux évaluer son risque réel

- Produire des analyses indépendantes basées sur ses propres données et hypothèses.
- Démontrer une meilleure maîtrise du risque pour **crédibiliser le dialogue technique** avec les réassureurs.

→ Négocier des conditions plus justes

- Argumenter sur des bases chiffrées locales pour **réduire le coût des traités** (prime de réassurance, seuils, priorités).
- Adapter les couvertures au profil réel du portefeuille plutôt qu'à un modèle générique "US-based".

Mieux connaître son risque pour mieux négocier

Structurer l'équipe de réponse à incident (CERT interne Dattak)

Les modèles prédictifs (comme les processus de Hawkes) aident à anticiper les vagues d'attaques et à prioriser les ressources internes. Ils permettent au CERT de Dattak d'adopter une approche proactive, en renforçant la veille, la prévention et la coordination lors des pics de sinistralité.

Contexte :

- En cas de vague d'attaques, la capacité de réponse rapide est clé pour limiter l'impact financier et réputationnel.
- Le CERT doit savoir **prioriser les incidents**, **anticiper les pics d'activité** et **coordonner la remédiation**.
- Les données de sinistres et de sécurité contiennent souvent des **signaux faibles** annonciateurs d'événements majeurs.

Pourquoi ces modèles sont essentiels :

→ Anticiper les vagues d'incidents

- Les processus de Hawkes permettent de détecter les dynamiques d'auto-contagion (une attaque en entraîne d'autres).
- Cela aide à **prévoir les périodes de tension** et à dimensionner les ressources du CERT en conséquence.

→ Prioriser et cibler les actions de réponse

- Identifier les **secteurs, technologies ou typologies d'entreprises** les plus exposés à court terme.
- Allouer les équipes techniques en priorité sur les incidents à fort potentiel de propagation.

→ Passer d'une posture réactive à proactive

- Exploiter les données de modélisation pour **déclencher des alertes préventives** auprès des assurés.
- Renforcer la communication et la prévention avant que les sinistres ne se multiplient.

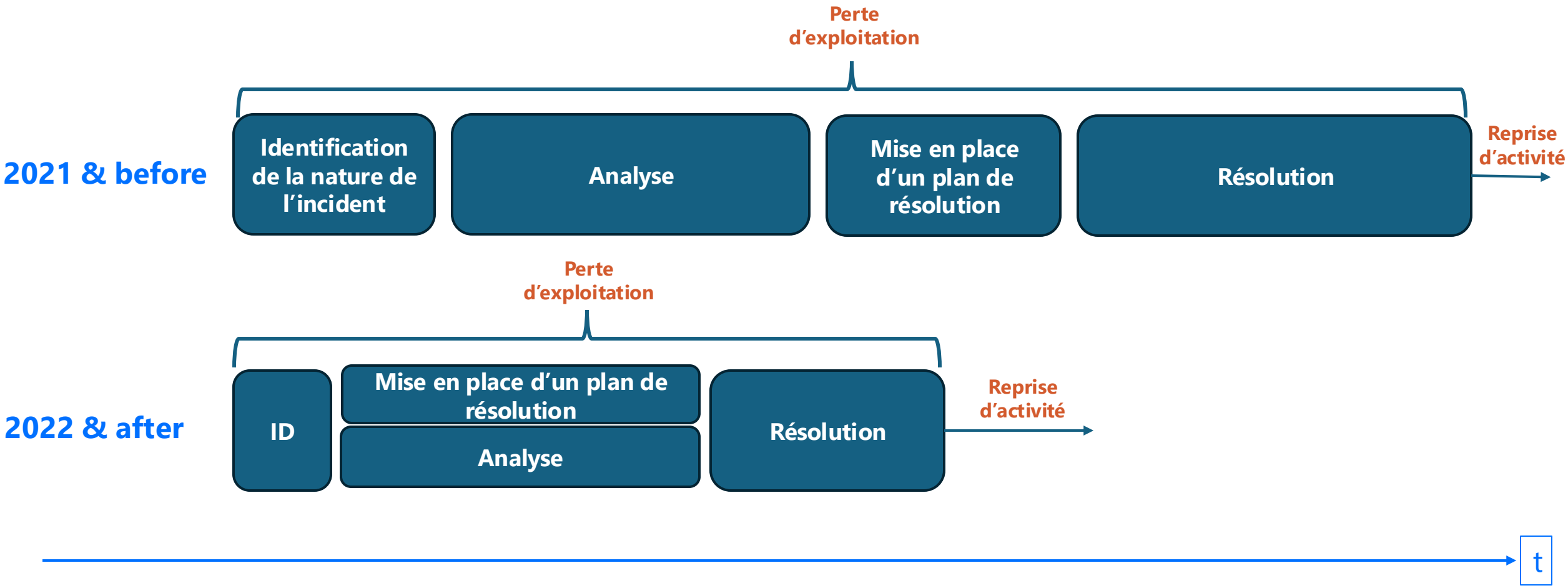
Merci pour votre attention!

Des Questions?

Evaluez cet atelier

Annexes

Impact de la professionnalisation de la réponse à incident



La base PRC

Présentation de la base

- Base de données publique des *data breaches* aux Etats-Unis
- Contient 74000+ incidents documentés depuis 2005
- Permet de tracker et analyser les fuites de données par organisation, secteur et type d'attaque
- Référence académique et publique pour comprendre les tendances des violations de données
- **Une variable indicative de la sévérité : le « *number_of_records* »**

Type de variable	Variables	Utilisation modélisation
Qualitatives catégorielles	- organization_type, - organization_type_explanation - breach_type, - breach_type_explanation	Embeddings Or Co-variables
Textuelles	- incident_details, - information_affected, - information_affected_explanation, - impact_info_explanation	NLP → vectorisation
Numériques brutes	- total_affected, - residents_affected,	Variable cible, features
Temporelles	- breach_date, - reported_date, - end_breach_date, - reporting_delay= reported_date - breach_date - exposure_duration=end_breach_date - breach_date	Date encoding,

La base PRC

Une structure instable dans le temps

Les variables de la base PRC ont changé en 2024, que ça soit dans leur construction, dans le type de donnée collectée,…

Type de variable	Variables PRC 2025	Variables PRC 2019
Qualitatives catégorielles	organization_type, breach_type,	Typ_of_organization, Type_of_breach,
Textuelles	incident_details, pdf_contents_cleaned information_affected,	Description_of_incident,
Numériques brutes	total_affected, residents_affected,	Total_Records
Temporelles	breach_date, reported_date, end_breach_date,	Date_Made_Public, Year_of_Breach,

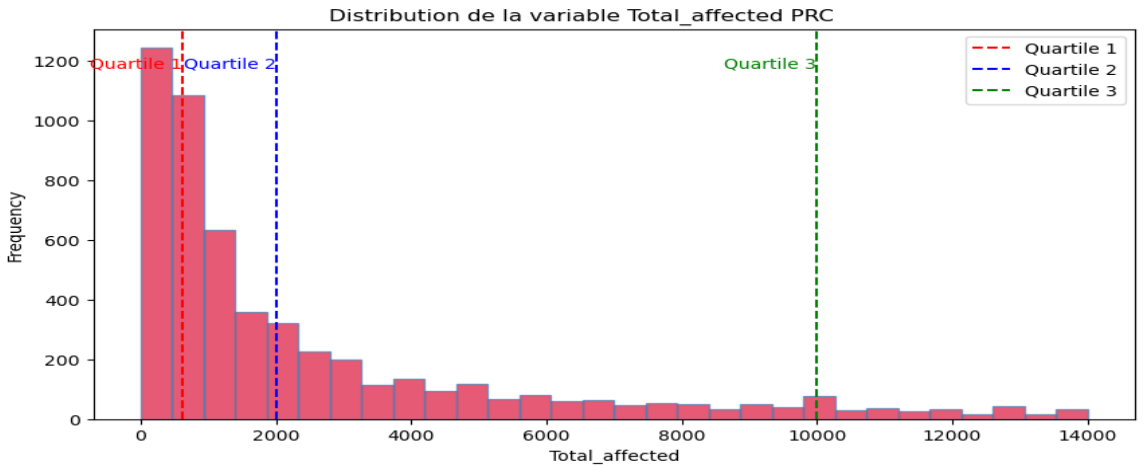
Conséquence directe : la distribution de la variable « number of records » (qui connote la sévérité) a également radicalement changé

Statistique	Sévérité Base PRC 2025 Vision 2005-2025	Sévérité Base PRC 2025 Vision 2005-2019	Sévérité Base PRC 2019 Vision 2005-2019
# sinistres	31 217	7 757	6 822
moyenne	434 789	781 721	1 522 632
Ecart-type	15 727 315	25 836 170	41 960 690
min	1	1	1
q25%	121	13	613
q50%	1 302	517	2 000
q60%	2 525	1 046	3 386
q75%	8 661	3 532	10 000
q90%	53 828	23 382	63 000
q95%	182 504	**80 000**	**259 123**
max	2 000 000 000	2 000 000 000	3 000 000 000

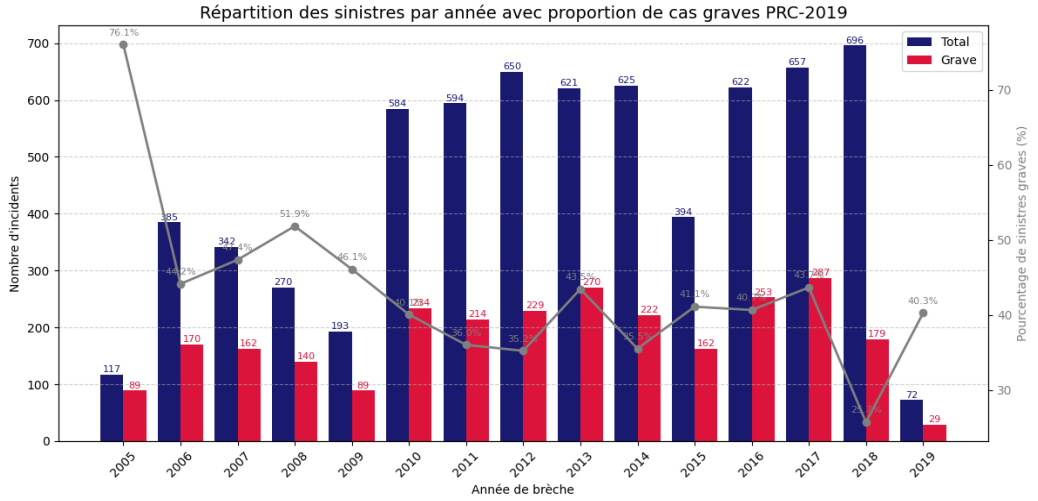
La base PRC

Visualisation du changement de distribution de la « sévérité »

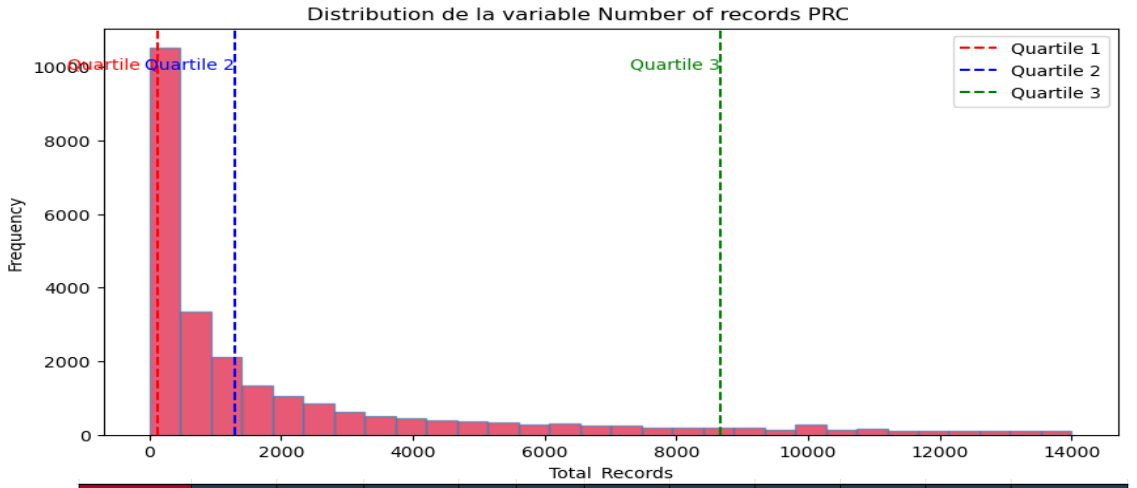
PRC 2019



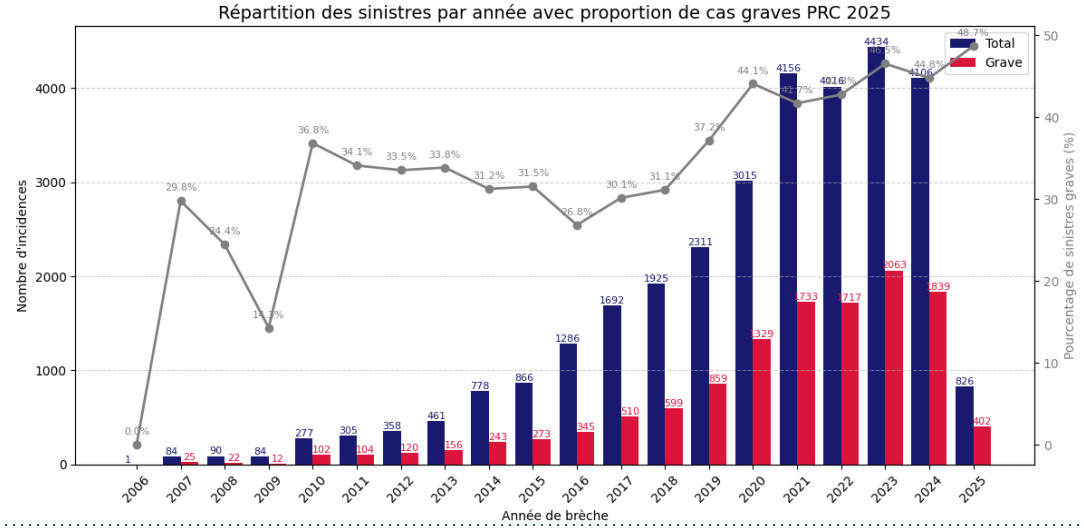
Statistiques	count	mean	std	min	25%	50%	60%	75%	90%	95%	max
Valeur	6 822	1 522 632	41 960 690	1	613	2 000	3 386	10 000	63 000	259 123	3 000 000 000



PRC 2025



Statistique	count	mean	std	min	25%	50%	60%	75%	90%	95%	max
Valeur	31 217	434 789	15 727 315	1	121	1 302	2 525	8 661	53 828	182 504	2 000 000 000



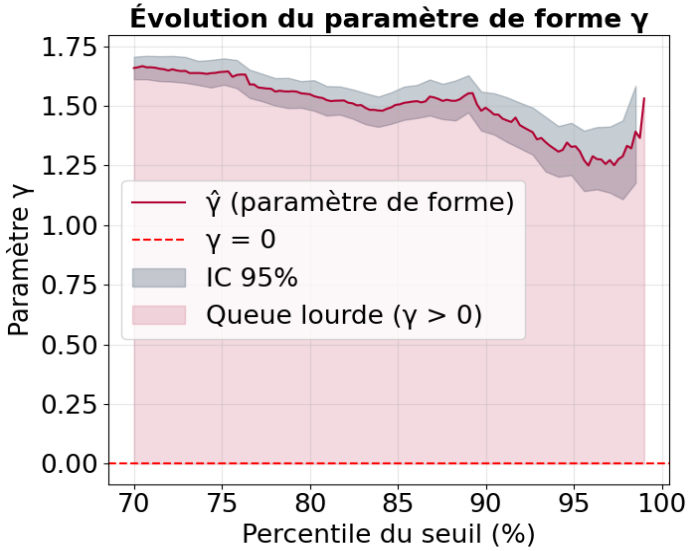
Un peu de Théorie des Valeurs Extrêmes

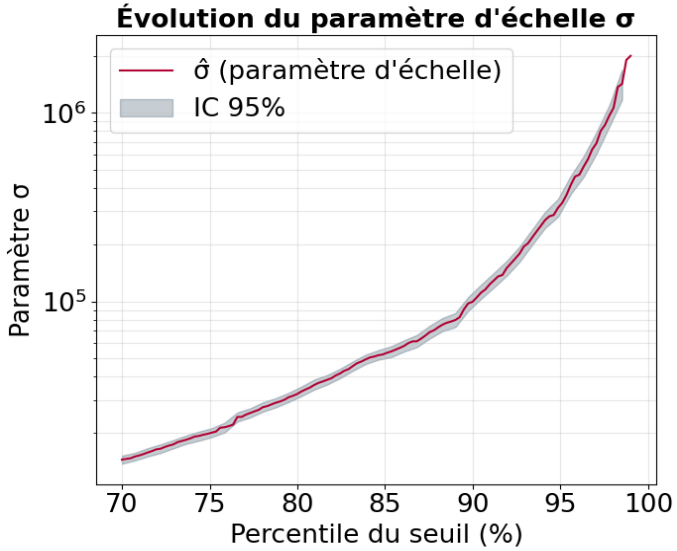
Approche Peaks Over Threshold (POT) afin de déterminer notre GPD

- Approche Peaks Over Threshold (POT) et loi de Pareto généralisée : théorème de Pickands- Balkema - de Haan

Objectif : calibrer les paramètres d'une loi de Pareto Généralisée sur les excès (le paramètre de position) à partir d'un seuil donné

On considère que le seuil « optimal » de graves est atteint lorsque le paramètre de forme cesse d'évoluer





Un peu de Théorie des Valeurs Extrêmes

Définition d'un seuil de « graves » optimal

- La distribution de la sévérité de la base est excessivement attritionnelle, encore plus que sur la version de 2019
  -> Nécessité de bien / mieux définir un seuil de « graves »

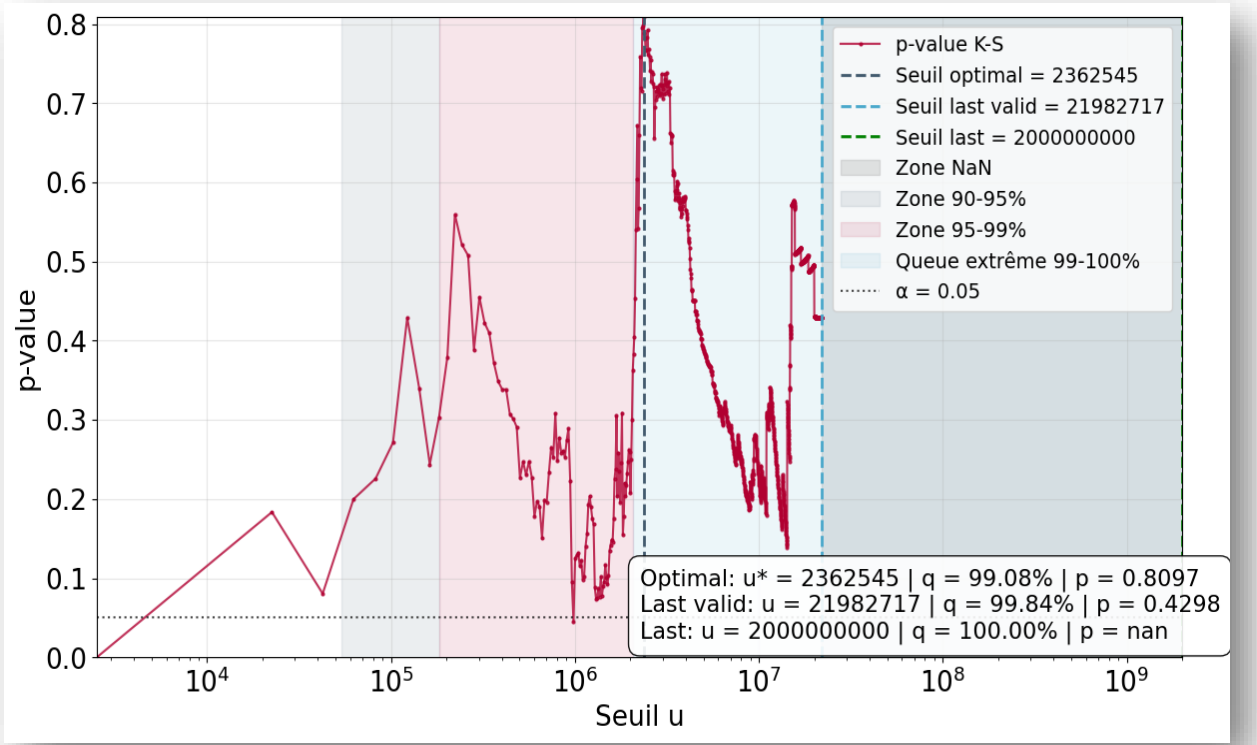
Idée : Calibrer un seuil de graves tel que la « vraisemblance » qui en résulterait soit maximale

Approche retenue : Test de Kolmogorov-Smirnov

On teste l'adéquation entre les données simulées à partir des paramètres calculés et la distribution empirique H0

Plus la p-value est forte, plus cela signifie qu'on pourra être confiant de garder le seuil comme seuil de séparation de graves

On voit ici une maximisation de la p-value pour un **seuil de ~2,3M**, ce qui équivaut à un **quantile de 99%**

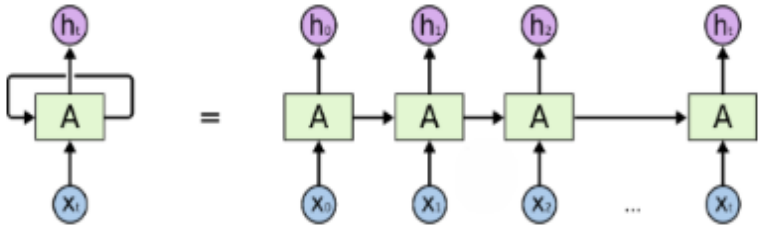


Percentile	n	$\hat{\gamma}$	IC 95 %(γ)	$\hat{\sigma}$	IC 95 %(σ)
95e	1561	1.3247	[1.2035 ; 1.4503]	326 668	[291 961 ; 365 415]
99e	313	1.5307	[1.2671 ; 1.7671]	2 005 313	[1 658 467 ; 2 498 293]

Table Estimation des paramètres de la loi GPD pour différents seuils

Le Deep Learning
Réseaux LSTM (Long Short-Term Memory)

Schématisation d'une Mémoire séquentielle (RNN)



Module répétitif d'un RNN standard avec une seule couche

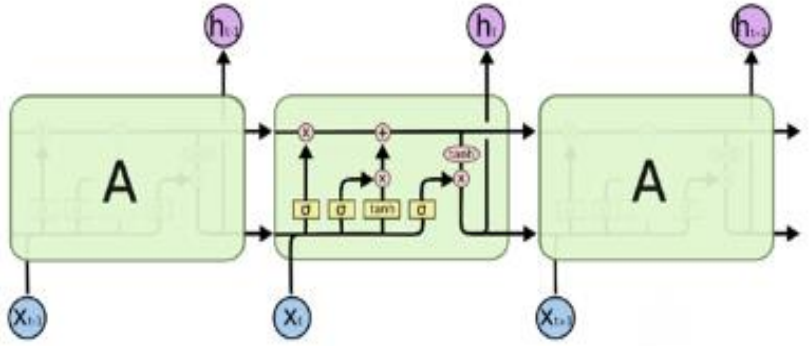
Réseau de Neurones Récurrent (RNN) standard

- Chaque sortie dépend de l'entrée courante et de l'état précédent.
- Oublie les dépendances et informations lointaines lointaines *(vanishing gradient)*

LSTM amélioré (LSTM initial : Hochreiter & Schmidhuber, 1997)

- Introduit une **cellule mémoire** C_t et **trois portes** (*gates*) :
  - **Forget gate** : décide ce qu'on oublie.
  - **Input gate** : décide ce qu'on ajoute à la mémoire.
  - **Output gate** : décide ce qu'on transmet à la sortie.
- Permet d'**apprendre le contexte long terme** sans perte d'information.

Le LSTM : une solution au problème de mémoire



Structure simplifiée d'une cellule LSTM (Olah 2015)

En résumé

Le LSTM « se souvient » là où le RNN « oublie » une avancée clé pour le traitement de textes séquentiels comme les **rapports d'incidents cyber** ou les **logs de sécurité**.

Références

- On contagion scenarii :

  Hillairet C., Lopez O. (2021), ***Propagation of cyber incidents in an insurance portfolio: counting processes combined with compartmental epidemiological models***,
  Scandinavian Actuarial Journal.

  Hillairet C., Lopez O., D'Outremont L., Spoorenberg B. (2022), ***Cyber contagion: impact of the network structure on the losses of an insurance portfolio***,
  Insurance: Mathematics and Economics.

- Frequency modeling and clustering effect

  Bessy-Roland Y., Boumezoued A., Hillairet C. (2020), ***Multivariate Hawkes process for Cyber Risk Insurance***,
  Annals of Actuarial Science, Volume 15 Issue 1.

  Boumezoued A., Cherkaoui Y., Hillairet C. (2025), ***Cyber Risk Frequency Modelling Using Hawkes Processes: Calibration on Attack and Vulnerability*** *Data*, hal-05305048.