

# Apport de la Data Science en Provisionnement Non-Vie

Rémi BELLINA, IA  
Loup ORTIZ, IA



Le provisionnement est l'un des cœurs de métier de l'actuariat. Du fait des enjeux stratégiques et économiques majeurs qu'il représente pour les compagnies d'assurance, sa pratique fait l'objet de développements toujours plus avancés. Les techniques usuelles d'évaluation des provisions – dont la méthode **Chain Ladder** fait figure de proue – reposent sur des données sinistres agrégées.

Cependant, l'expansion récente des volumes de données collectées et des algorithmes associés permet aujourd'hui la conception de méthodologies d'estimation innovantes s'appuyant sur des informations à la maille individuelle. Ces méthodes, proposées en nombre par la communauté actuarielle, constituent des développements théoriques novateurs.

Toutefois à ce stade, et compte tenu des fortes contraintes opérationnelles (calendrier serré, disponibilité et qualité des données, etc.), celles-ci n'ont pas encore atteint un niveau de maturité suffisant au sein des compagnies pour s'imposer. En pratique, les méthodes classiques (telles que Chain Ladder, Bornhuetter-Ferguson ou pure exposition via l'utilisation de ratio Sinistres à Primes S/P) restent ainsi la référence pour l'évaluation des provisions.

Demeurent donc de **multiples enjeux**, que ces derniers soient liés à une connaissance approfondie du risque (obtention d'ultimes à la maille individuelle, détection anticipée des graves)

ou à des considérations plus pragmatiques (rationalisation des processus et automatisation des calculs, des analyses d'écart d'expérience, des sensibilités et des rapports).

## L'Intelligence Artificielle : une nouvelle boîte à outils

Pour accompagner l'actuaire dans ce contexte, une exploitation rationnelle des possibilités offertes par l'Analytics et la Data Science est possible : visualisations innovantes, automatisation de tâches répétitives, modélisation via machine learning de processus complexes, etc.

Historiquement, les pratiques d'évaluation des Provisions pour Sinistres A Payer (PSAP) reposent sur l'agrégation des données sous forme de triangles de développement et sur l'application de méthodes actuarielles d'estimation des provisions. Néanmoins, c'est bien **tout ce processus** complexe, de la prise en charge des données jusqu'à l'estimation d'un ultime, qui peut être considéré aujourd'hui sous un nouveau regard.

[Comment les nouvelles technologies et des éléments d'Intelligence Artificielle peuvent apporter un nouveau regard sur le provisionnement des compagnies d'assurance ?](#)

Qu'ils soient directeurs, actuaires, gestionnaires, souscripteurs, tous s'accordent sur la mutation des pratiques induites par

l'avènement du numérique, des techniques de Machine Learning et de l'exploitation des données en grande dimension.

C'est l'une des problématiques qui se pose à l'heure actuelle, dépassant le seul périmètre de calcul d'un montant de charge ultime. On attend en effet d'une Intelligence Artificielle en provisionnement qu'elle soit en mesure de manipuler efficacement des données de différentes natures, d'en extraire les informations pertinentes, d'automatiser certains calculs, d'intervenir en support auprès de l'actuaire en lui fournissant une aide à la décision de qualité, et enfin de restituer les résultats obtenus de façon claire et intelligible.

Comment également tirer parti des nombreuses informations récoltées à la maille individuelle des sinistres ?

Ces thématiques se cristallisent tout naturellement autour de la Data Science, en pleine émergence dans le provisionnement, et se concrétisent par l'utilisation de plateformes dédiées à la gestion et à la visualisation de données ainsi qu'à la mise en œuvre d'algorithmes de Machine Learning.

Des langages comme R ou Python répondent à ces différents prérequis. Libres d'accès, ils présentent une simplicité de prise en main qui leur permet d'être de plus en plus répandus au sein des compagnies d'assurance.

Les développements relatifs à ces technologies trouvent leur place dans l'évolution des pratiques usuelles. Ils constituent désormais des outils d'aide à la décision qui se veulent robustes et pragmatiques.

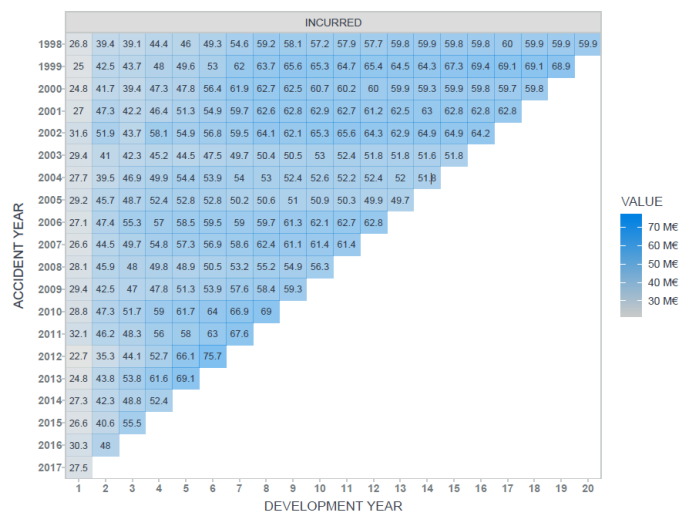
Diagnostic initial

La prise en charge des données et la réalisation d'analyses descriptives approfondies sont inhérentes à toute étude statistique. Elles s'inscrivent directement dans le processus de provisionnement en amont de la modélisation des montants de PSAP du fait de la quantité et de la richesse des caractéristiques de sinistres qu'il est possible de récolter.

Couplées aux bibliothèques de manipulation de données, les grammaires graphiques sont de plus en plus appréciées, véritables clés de voûte des visualisations, et peuvent permettre d'appréhender efficacement les formats agrégés comme individuels dans le cadre du provisionnement.

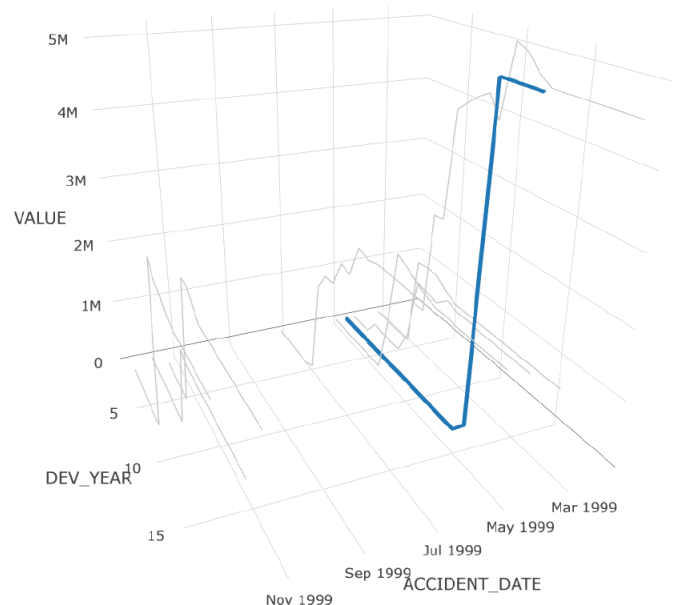
D'une part, la représentation des triangles de liquidation sous différentes formes rend compte de leur cadence de développement, du type d'activité concerné, ainsi que des spécificités du portefeuille notamment en matière de volume de sinistralité. Ce format de données standardisé est d'ailleurs d'autant plus à même d'être pris en charge de façon automatisée.

FIGURE 1 : TRIANGLE DE DÉVELOPPEMENT



D'autre part, la réalisation de représentations graphiques à l'échelle du sinistre permet d'exploiter la granularité des données individuelles et favorise l'extraction des dossiers atypiques du portefeuille.

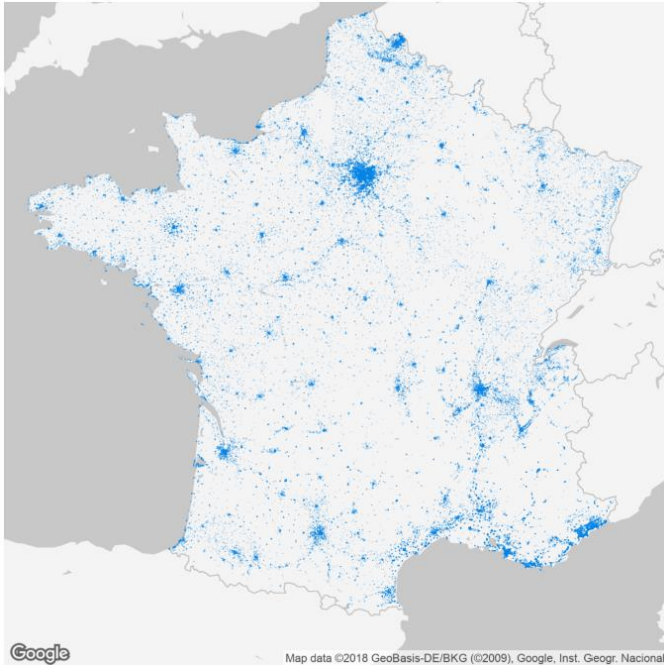
FIGURE 2 : SUIVI INDIVIDUEL DES SINISTRES



La description spatiale des données à la maille individuelle permet plus généralement d'élargir le spectre des discussions entre les différents acteurs : les techniques cartographiques illustrent la répartition de la sinistralité d'un portefeuille au sein du territoire considéré afin d'identifier les zones de forte concentration. A titre d'exemple, la sinistralité des branches Responsabilité Civile peut ainsi être reliée plus facilement aux

jurisprudences de tribunaux régionaux. De telles représentations peuvent en outre être couplées aux informations historiques d'événements catastrophiques, pour en mesurer l'impact.

FIGURE 3 : VISUALISATION SPATIALE DES SINISTRES



La superposition de l'ensemble des informations au sein d'une même carte fournit une synthèse intéressante aux actuaires en charge du provisionnement, mais aussi aux gestionnaires ainsi qu'aux souscripteurs dont la politique de souscription peut être ajustée en conséquence.

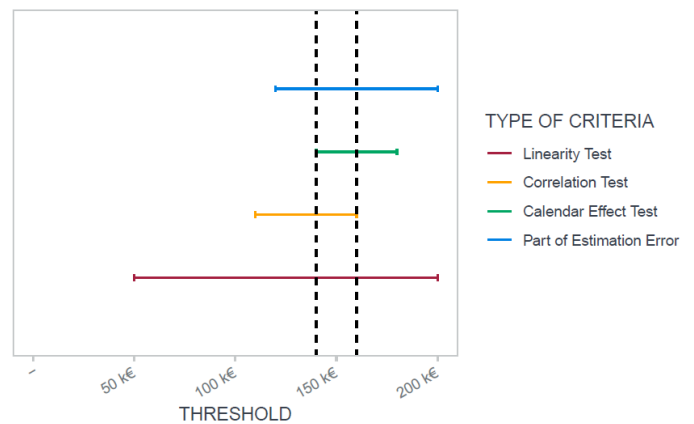
### Identification des sinistres atypiques

La segmentation des données a pour but de constituer des groupes de sinistres homogènes dans une optique de modélisation des montants de provisions. A ce titre, la classification des sinistres dans les catégories des attritionnels et des graves est une problématique centrale en provisionnement. Ces deux classes de sinistres présentent en effet des caractéristiques différentes et font généralement l'objet de traitements spécifiques. Les méthodes agrégées les plus usuelles reposent par exemple sur des hypothèses de stabilité des triangles, aussi est-il nécessaire de disposer d'un périmètre adapté pour garantir un bon ajustement des modèles.

La définition d'un **seuil de gravité** pour la catégorisation des sinistres constitue un premier levier de segmentation. Plutôt que de considérer un niveau de seuil parfois arbitraire ou imposé, il est intéressant d'objectiver ce montant en se basant sur des indicateurs représentatifs de la stabilité des données résultantes. Dans ce cadre, l'application de l'approche Chain Ladder et de sa variante stochastique le modèle de Mack sur le triangle d'attritionnels issu de la segmentation constitue un outil intéressant : la simplicité de mise en œuvre de ces méthodes

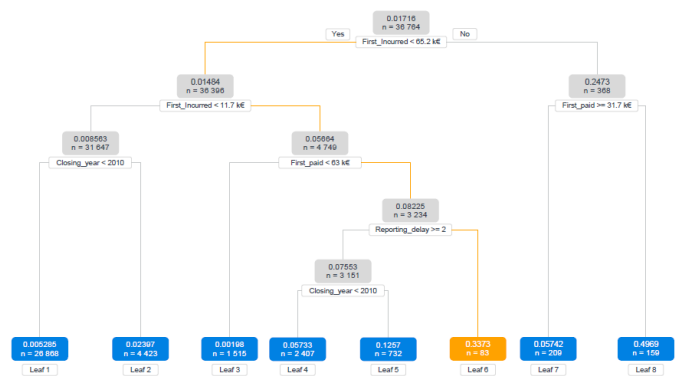
rend envisageable leurs ajustements successifs pour une large gamme de valeurs de seuils de gravité. En conséquence, la validité des hypothèses des modèles ainsi que certains indicateurs peuvent être analysés au regard de ce paramètre : hypothèse de linéarité, hypothèse de non-corrélation des facteurs de développement successifs, absence d'effet calendaire, évolution de l'erreur de prédiction de Mack (MSEP), etc. L'identification et l'intersection des intervalles de seuils pour lesquels chaque critère est validé permet de définir un ensemble de valeurs admissibles à comparer aux montants retenus en pratique. Cette méthode se veut objective et a pour but d'apporter de l'information pertinente aux actuaires.

FIGURE 4 : INTERVALLES DE SEUILS



La segmentation en fonction d'un niveau de seuil de gravité a néanmoins pour effet l'introduction d'une hétérogénéité relative aux sinistres attritionnels les plus récents. Ces derniers sont en effet susceptibles de dévier dans la classe des graves au cours de leur développement futur. La mise en œuvre de modèles prédictifs de Machine Learning visant à identifier les sinistres les plus coûteux peut permettre d'affiner la segmentation.

FIGURE 5 : ARBRE DE DECISION POUR ANTICIPER LES ATYPIQUES



Dans ce contexte, l'algorithme CART, calibré sur la partie la plus développée de l'historique de sinistralité, constitue un outil flexible en matière de prise en charge des données et aboutit à des résultats facilement interprétables. Plus généralement, le fait de disposer d'une vision synthétique de l'ensemble des

sinistres a priori sensibles constitue un apport intéressant pour les différents acteurs.

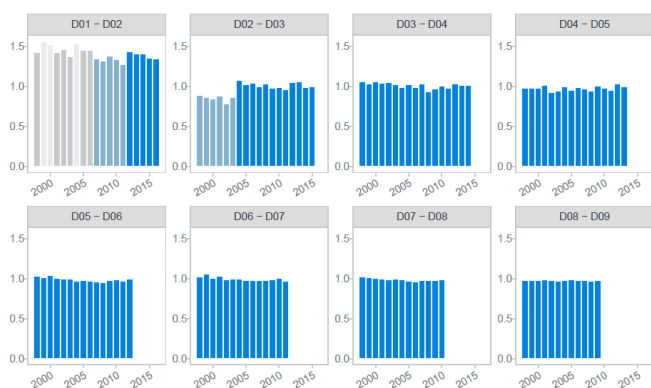
En définitive, la combinaison des deux approches de segmentation extrait des renseignements supplémentaires sur les profils de risque du portefeuille en tirant parti des données individuelles de sinistralité.

## L'Intelligence Artificielle : une aide à la décision pour la modélisation des charges ultimes

Les problématiques de choix et de mise en œuvre des modèles d'estimation des PSAP sans marge de prudence sont bien évidemment centrales dans un processus de provisionnement. La phase de modélisation est ici considérée sous l'angle de la Data Science et se décline en différentes thématiques autour de l'automatisation et de l'exploitation des données individuelles de sinistralité.

La réalisation de diagnostics via l'extraction d'informations ainsi que l'obtention d'un ultime sont des pratiques quotidiennement réalisées au sein des compagnies d'assurance. Dans ce contexte, l'automatisation de ces deux procédés sous la forme d'une Intelligence Artificielle façonnée comme une aide à la décision pour les actuaires représente une valeur ajoutée certaine. La standardisation des données nécessaires à son exécution est telle qu'il est envisageable de l'appliquer à tout type de triangle et de bénéficiaire d'un gain de performance intéressant : la réduction du temps nécessaire au traitement de l'information peut permettre de concentrer les analyses sur les aspects métiers les plus essentiels.

FIGURE 6 : ANALYSE AUTOMATIQUE DES COEFFICIENTS DE DEVELOPPEMENT (LINK RATIOS) PAR ANNEE DE SURVENANCE



Les techniques issues de la Data Science ont énormément à apporter aux méthodes classiques d'estimation, que ce soit de façon totalement automatique pour une seconde opinion, ou bien en support à la modélisation des segments à développements longs les plus complexes.

Au-delà, les méthodes populaires de Machine Learning permettent également une prise en charge flexible des données individuelles en vue d'en extraire des informations inexploitées

jusqu'alors, du moins de façon efficace et standardisée. La segmentation des données permet l'application de méthodes dédiées tirant parti de la nature et des spécificités des groupes homogènes constitués ainsi que des données collectées.

FIGURE 7 : EXPLOITATION DES COMMENTAIRES



La modélisation du processus d'occurrence des paiements, basée sur l'exploitation des développements individuels de la masse des sinistres attritionnels, constitue une première étape dans la démarche d'estimation des montants de charges ultimes. Le cadre structurel qui découle d'une telle approche ligne à ligne peut être étendu à d'autres variables d'intérêt telles que les coefficients individuels de passage dont la construction n'est pas sans rappeler l'approche de référence Chain Ladder. De même, l'exploitation systématique de commentaires, enregistrés sous forme de texte par les gestionnaires à l'attention des sinistres les plus déviants, permet une appréhension nouvelle des sinistres graves dont le nombre est généralement plus restreint. Les méthodes sophistiquées de **text mining**, ou plus simplement l'emploi pragmatique d'expressions régulières, conduisent à l'extraction d'informations pertinentes et favorisent le dialogue entre actuaires et gestionnaires.

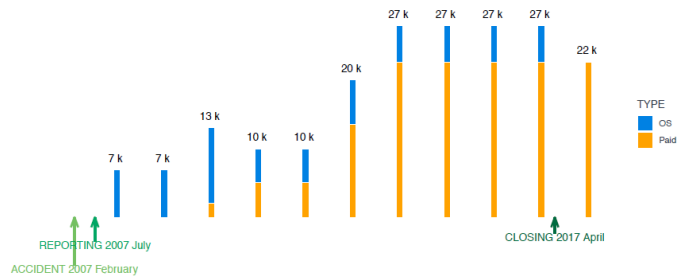
## Restitution automatisée des résultats

La mise en forme et l'interprétation intelligentes des résultats obtenus à l'issue de la réalisation des diverses analyses en matière de visualisation, d'extraction d'informations et de modélisation, représentent une composante supplémentaire à intégrer au concept d'Intelligence Artificielle en provisionnement. La plupart d'entre elles viennent en effet enrichir un rapport dont la constitution fait partie du quotidien des compagnies d'assurance. La génération automatisée de commentaires descriptifs ou conclusifs sous forme de langage naturel représente un gain de temps important et participe à la rationalisation de la pratique de description et d'interprétation des résultats.

De nombreux travaux sont en cours sur ce sujet et peuvent être appliqués à l'échelle globale du portefeuille, comme à la maille individuelle des sinistres. Le développement d'**applications web interactives** est par ailleurs de plus en plus répandu du fait

de l'accessibilité de certaines technologies dédiées à leur réalisation (à titre d'exemple, **Shiny** et **Dash** sont respectivement adaptées aux langages R et Python).

FIGURE 8 : RAPPORT AUTOMATIQUE INDIVIDUEL



Le sinistre numéro SIN\_33021 est survenu en début d'année 2007 dans le département de l'Essonne et a été déclaré 5 mois plus tard. Il présente une charge à date de plus de 22 k€ et a fait l'objet de 4 ajustements de paiements pour un versement cumulé total de 22 k€. Par rapport à la dernière extraction, nous observons un boni dû à une diminution de la charge d'environ 5 k€. Ce sinistre a été clôturé en 2017, 10 ans après sa déclaration. Du fait des montants mis en jeu et de ses caractéristiques, il a été classé dans la catégorie des sinistres attritionnels. Les différentes méthodes mises en œuvre en conséquence ont conduit à des estimations d'ultimes comprises entre 20 k€ et 27 k€.

De telles applications sont d'autant plus appropriées dans le cadre du provisionnement qu'elles permettent d'unifier l'ensemble des analyses du processus au sein d'une même interface tout en permettant à l'actuaire d'agir directement sur les choix de modélisation.



Milliman is among the world's largest providers of actuarial and related products and services. The firm has consulting practices in life insurance and financial services, property & casualty insurance, healthcare, and employee benefits. Founded in 1947, Milliman is an independent firm with offices in major cities around the globe.



[milliman.com](http://milliman.com)

© 2019 Milliman, Inc. All Rights Reserved. The materials in this document represent the opinion of the authors and are not representative of the views of Milliman, Inc. Milliman does not certify the information, nor does it guarantee the accuracy and completeness of such information. Use of such information is voluntary and should not be relied upon unless an independent review of its accuracy and completeness has been performed. Materials may not be reproduced without the express consent of Milliman.

## Perspectives

Le provisionnement en assurance IARD est impacté par cette mutation technologique progressive. L'avènement de la Data Science regroupe un grand nombre de thématiques qui s'articulent en une succession de modules, de la prise en charge des données jusqu'à la restitution des résultats. Deux axes complémentaires émergent naturellement : la rationalisation des processus et la connaissance approfondie des risques. Plus d'informations sont disponibles dans le mémoire d'actuaire « Eléments d'Intelligence Artificielle faible en Provisionnement Non-Vie » (Milliman, ORTIZ, février 2019).

Chaque étape peut être repensée sous l'angle de l'Intelligence Artificielle, en support des techniques actuarielles existantes et en intégrant l'indispensable expertise métier.

Les futures ambitions sont d'ores et déjà soulevées : approfondissement des méthodes d'estimation des PSAP, élaboration de sensibilités intelligentes, évaluation et allocation individuelles pertinentes des montants de charges ultimes, etc. Ces diverses perspectives sont d'autant plus prometteuses que les volumes de données collectées sont de plus en plus importants et de qualité.

## CONTACTS

Remi BELLINA, IA  
[remi.bellina@milliman.com](mailto:remi.bellina@milliman.com)

Loup ORTIZ, IA  
[loup.ortiz@milliman.com](mailto:loup.ortiz@milliman.com)