Mémoire présenté devant le jury de l'EURIA en vue de l'obtention du Diplôme d'Actuaire EURIA et de l'admission à l'Institut des Actuaires

le 8 Septembre 2023

Par : Agathe FERNANDES MACHADO
Titre : Contribution marginale des sites industriels au coût de réassurance d'un traité de type excédent de sinistre par évènement

Confidentialité : Oui, 2 ans

Les signataires s'engagent à respecter la confidentialité indiquée ci-dessus

Membres présents du jury de l'Institut des Actuaires :
Saman NEJAD
Stella JOVET
Christelle LACAZE
Signature :

Membre présent du jury de l'EURIA :
Franck VERMET

Entreprise :
Generali France
Signature :

Directeur de mémoire en entreprise :
Marco BRETTSCHNEIDER
Signature :

Invité :

Signature :

Autorisation de publication et de mise en ligne sur un site de diffusion de documents actuariels
(après expiration de l'éventuel délai de confidentialité)

Signature du responsable entreprise :

Signature du candidat :

EURIA
EURo Institut
d'Actuariat

6, avenue le Gorgeu
CS 93837
29238 Brest Cedex 3

T +33 (0)2 98 01 66 55
euria@univ-brest.fr

Résumé

Les épisodes de grêle et de tempête se sont multipliés en France en 2022, entre les mois de mai et juillet. A ces évènements s'ajoute également une longue période de sécheresse à partir du mois de mai. Ces catastrophes naturelles combinées à l'augmentation du taux d'inflation ont conduit les réassureurs à élever le prix des traités de réassurance au renouvellement 2023. Afin d'optimiser leur portefeuille, il est donc important pour les assureurs d'être en mesure d'identifier les sites assurés contribuant fortement à ce coût de réassurance. Les sites industriels représentant individuellement des engagements assurés importants, l'objectif de ce mémoire est d'évaluer leur coût marginal de réassurance pour un traité de type excédent de sinistre par évènement et pour le péril tempête.

La présente étude s'articulera autour de quatre axes. Il s'agira dans un premier temps de modéliser les pertes liées au péril tempête sur le portefeuille dommages aux biens de Generali France, à l'aide d'un logiciel de modélisation d'évènements climatiques, appelé RMS. Dans un deuxième temps, des simulations seront effectuées sur les tables en sortie de ce logiciel afin de calculer la prime de réassurance du traité excédent de sinistre par évènement, pour le péril tempête. L'obtention de cette prime permettra de calculer les coûts marginaux de réassurance des sites industriels du portefeuille existant. Pour finir, plusieurs algorithmes seront testés afin d'être en mesure de prédire le coût marginal de réassurance d'un nouveau site industriel, à l'aide des caractéristiques renseignées par le souscripteur.

Mots clefs: Réassurance, Excédent de sinistre par évènement, Tempête, RMS, Tarification, Coût marginal, Prédiction, Souscription

Abstract

Hailstorms and windstorms events have strongly increased in France in 2022, between May and July. These events were compounded by a long period of drought from May onwards. These natural disasters, combined with rising inflation, have led reinsurers to increase the price of reinsurance treaties for the 2023 renewal. In order to optimise their portfolio, it is therefore essential for insurers to be able to identify the insured risks that make a significant contribution to this reinsurance cost. As industrial risks individually represent a significant amount of sum insured, the aim of this report is to assess their marginal reinsurance cost for an excess of loss per event treaty, for storm events.

This study will be divided into four steps. Firstly, storm-related losses on Generali France's property damage portfolio will be modelled using CAT modelling software called RMS. Secondly, simulations will be carried out on the output tables of this software in order to calculate the reinsurance premium for the excess of loss treaty per event, for storms. Obtaining this premium will allow us to calculate marginal reinsurance costs of the industrial risks in the existing portfolio. Finally, several algorithms will be tested in order to be able to predict the marginal reinsurance cost of a new industrial risk, using the features provided by the underwriter.

Keywords: Reinsurance, Property per event treaty, Windstorm, RMS, Pricing, Marginal cost, Prediction, Underwriting

Note de synthèse

Contexte et problématique

L'année 2022 apparaît comme la deuxième année la plus coûteuse en terme de sinistralité climatique pour les assureurs selon France Assureurs : le coût est estimé à 10 milliards d'euros pour le marché français. En effet, le dérèglement climatique entraîne l'augmentation de la fréquence et de l'intensité des évènements naturels. Ce phénomène combiné à la hausse du taux d'inflation ont conduit les réassureurs à élever le prix des traités de réassurance pour l'année 2023.

Le type de traité de réassurance permettant généralement de couvrir l'assureur contre la sinistralité climatique est appelé excédent de sinistre par évènement et s'applique sur la somme des montants de sinistres dus à un même évènement. Ce traité est notamment utilisé chez Generali France pour couvrir les sinistres indemnisés par la garantie dommages aux biens, hors auto, également nommée DAB Domestique. Celle-ci couvre les produits MRH, MRC, MRA, RI, etc. Ce traité de réassurance, appelé CAT XL, est activé lorsqu'un évènement climatique, hors sécheresse, endommage un ensemble d'engagements assurés supérieur à un certain montant. Afin de réduire le coût de réassurance, il est aujourd'hui important pour les compagnies d'assurance de connaître les zones exposées à l'aléa climatique et comportant une concentration de sites assurés pouvant déclencher ce type de traité.

La connaissance de ces zones peut notamment s'obtenir en estimant la contribution individuelle au prix de réassurance des différents sites assurés composant le portefeuille DAB Domestique de Generali France. En effet, un site exposé à l'aléa tempête et étant situé dans une zone contenant de nombreux autres sites assurés aura tendance à participer fortement aux cessions du traité impactant directement son prix, se calculant lui-même comme une moyenne de cessions annuelles. Toutefois, à l'exception de produits comme le RI et le MRI, les engagements assurés des sites sont trop faibles pour que cette contribution individuelle soit significative pour un traité de réassurance s'appliquant au niveau évènement. De plus, il a été observé dans le passé que les dégâts causés sur un site industriel peuvent être très importants en raison des conséquences sur l'environnement et de la couverture perte d'exploitation (PE), permettant à une entreprise de compenser les effets d'une réduction d'activité.

L'objectif de cette étude est donc de calculer la contribution marginale des sites industriels, composant le produit RI, au prix d'un traité de réassurance de type excédent de sinistre par évènement, pour le péril tempête. L'utilisation opérationnelle de cette approche repose sur la mise à disposition d'un indicateur sur le coût de réassurance au niveau site pour la souscription RI. Cela implique d'être en mesure de prédire la contribution marginale d'un nouveau site industriel lors de son acquisition dans le portefeuille.

Périmètre de l'étude

L'étude porte sur la garantie DAB Domestique, dont les sinistres liés aux périls naturels (tels que la tempête, la grêle, la tornade, l'inondation, le gel et les ouragans) sont réassurés par le traité CAT XL. Les produits concernés sont : MRH, MRC, MRA, MRI, RI, *engineering* (branche construction) et *municipal* (branche collectivités). Seuls les évènements tempête sont considérés ici. En effet, les cessions historiques au traité CAT XL dues aux évènements tempête représentent environ 75% des cessions totales. De plus, les contributions individuelles à la tarification de ce traité sont calculées seulement pour les sites industriels à l'aide du coût marginal.

Démarche suivie

Le projet s'articule principalement autour de deux axes. La première étape consiste à calculer le coût marginal de réassurance, pour le traité CAT XL et le péril tempête, des sites industriels du portefeuille existant à l'aide d'un logiciel de modélisation CAT appelé RMS. Ainsi, la prime de réassurance du traité CAT XL, correspondant à une moyenne des cessions annuelles sur les pertes modélisées par le logiciel, est calculée. La deuxième étape consiste à mettre en place un proxy de calcul permettant de prédire le coût marginal d'un nouveau site industriel à l'aide de ses caractéristiques.

Mise en application sur les sites industriels du portefeuille existant

Le portefeuille RI de Generali France comporte plus de 10 000 sites assurés. Afin de calculer leur contribution marginale au prix de réassurance du traité CAT XL pour le péril tempête, il est nécessaire de calculer la prime de ce traité dans un premier temps. En effet, le coût marginal de réassurance du site industriel n°i se définit par la formule suivante :

$$CM(i) = Prime(N) - Prime(N \backslash i)$$

où $Prime(N)$ représente la prime de réassurance modélisée du traité, donc pour l'ensemble du portefeuille DAB Domestique, et $Prime(N \backslash i)$ correspond à cette même prime lorsque le site industriel n°i n'est pas présent dans le portefeuille. Ainsi, le coût marginal de réassurance du site n°i s'interprète comme étant le coût de réassurance supplémentaire engendré par l'acquisition du site n°i, en dernier, dans le portefeuille RI.

Tout d'abord il convient de calculer $Prime(N)$, pour les cessions liées au péril tempête. Pour cela, le logiciel RMS modélise les pertes causées par un catalogue d'évènements tempête historiques et stochastiques sur l'ensemble des sites de Generali couverts par la garantie DAB Domestique. En sortie de RMS, on dispose d'une table appelée *Event Loss Table* (ELT) pour chaque produit. Une ELT comporte un catalogue d'évènements tempête touchant le portefeuille considéré, ainsi que la perte moyenne et l'écart-type associés. En effet, le logiciel RMS considère que la perte modélisée pour chaque évènement correspond à une perte moyenne et non pas à une perte réelle. Cela est dû à l'incertitude secondaire causée par le manque de données, l'incertitude des modèles ainsi que

par les variations de résistance aux tempêtes des différentes régions françaises. De plus, une ELT contient la fréquence annuelle de survenance des évènements et l'exposition du portefeuille à chaque évènement. Ci-dessous est présenté un exemple d'ELT pour deux évènements tempête :

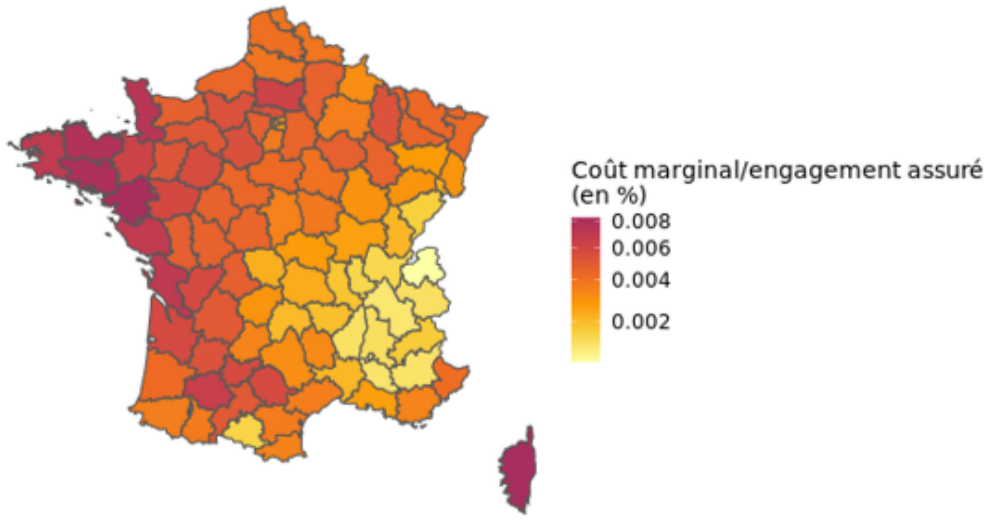
EVENT ID	FREQ	PERTE MOYENNE	ECART-TYPE	EXPOSITION
1	0.01	500	350	850
2	0.005	250	470	500

Dans un premier temps, les sept ELT, associées aux sept produits concernés, sont regroupées afin de disposer du catalogue d'évènements complet ainsi que des pertes associées sommées sur l'ensemble des sites assurés par la garantie DAB Domestique.
La prime de réassurance se calculant comme une moyenne de cessions annuelles, une méthode de simulation de type Monte Carlo est mise en place sur l'ELT regroupée afin de détenir une information sur l'année pour la survenance des évènements. Le choix d'années de simulation se porte sur 50 000 afin de fiabiliser les résultats. Pour chaque année de simulation, il s'agit dans un premier temps de tirer aléatoirement les évènements tempête grâce à des lois de Poisson dont les paramètres reposent sur les fréquences annuelles de survenance, puis de simuler les pertes associées à chaque évènement tiré. Deux méthodes différentes sont utilisées pour la simulation des pertes :

1. L'incertitude secondaire n'est pas prise en compte et on associe donc à chaque évènement sa perte moyenne, contenue dans l'ELT.
2. L'incertitude secondaire est prise en compte. Les taux de destruction de chaque évènement sont simulés à l'aide de lois Bêta, dont les paramètres sont calculés par la méthode des moments grâce à la perte moyenne et l'écart-type contenus dans l'ELT.

La table résultante des simulations se nomme *Year Loss Table* (YLT). Afin de tarifer le traité CAT XL pour le péril tempête, il s'agit de céder au traité le montant de perte supérieur à la priorité du CAT XL, pour chaque année et chaque évènement de l'YLT.
En comparaison à certaines valeurs directement calculées sur les ELT fournies par RMS, telles que les valeurs de la courbe OEP, la méthode de simulation des pertes avec incertitude secondaire est retenue. En effet, celle-ci permet de considérer toute l'étendue des pertes, autour de la valeur moyenne. Ainsi, des montants de perte plus élevés sont obtenus. La $Prime(N)$ sauvegardée est donc celle calculée grâce à la deuxième approche.

Deuxièmement, il convient de calculer $Prime(N \backslash i)$ pour l'ensemble des sites du portefeuille RI. Pour cela, il suffit de reproduire les étapes de l'approche exposée pour calculer $Prime(N)$ mais en supprimant, initialement, les lignes de l'ELT du portefeuille RI associées au site industriel n°i. De cette façon, les sept ELT des produits, dont la nouvelle ELT associée au portefeuille industriel, sont regroupées une nouvelle fois. Les coûts marginaux de réassurance, pour le traité CAT XL et le péril tempête, sont obtenus pour l'ensemble des sites assurés du produit RI. La carte des coûts marginaux divisés par l'engagement assuré RI par département est représentée en début de page suivante.



Les coûts marginaux de réassurance du traité CAT XL représentent correctement l'aléa tempête puisque celui-ci est également élevé à l'ouest de la France selon le zonier vent construit chez Generali à l'aide de RMS.

Prédiction pour un nouveau site industriel

L'objectif principal de cette étude est de mettre à disposition du souscripteur RI le coût marginal de réassurance, pour le traité CAT XL et le péril tempête, d'un site industriel lors de son acquisition. Après avoir développé une approche permettant de calculer les coûts marginaux des sites industriels du portefeuille existant de façon robuste, il s'agit de développer un proxy de calcul afin d'estimer ce coût marginal pour un nouveau site industriel lors de sa souscription. En effet, le souscripteur a besoin de cette information en un temps de calcul raisonnable afin de l'exploiter. Or, l'approche de calcul présentée ci-dessus ne le permet pas car elle nécessiterait de modéliser les pertes du portefeuille RI sur RMS en y ajoutant le nouveau site à souscrire. Par conséquent, des algorithmes de prédiction des coûts marginaux sont calibrés sur les sites industriels du portefeuille existant, à l'aide de leurs caractéristiques telles que leur somme assurée, leur activité, leur département, etc. Ces différentes caractéristiques sont renseignées dans un outil de tarification par le souscripteur RI lors de la sousciption d'un nouveau site. De plus, un zonier vent construit par département pour le produit MRC est ajouté comme variable explicative. Les sites industriels du portefeuille RI actuel sont donc séparés en une base d'apprentissage et une base de test.

Dans un contexte assurantiel, comme les coûts marginaux de réassurance appartiennent à $\mathbb{R}_+$, on aurait tendance à utiliser un GLM Gamma ou log-normal pour les prédire. Cependant, la distribution des coûts marginaux des sites assurés chez Generali présente une queue épaisse à droite. Les lois Gamma et log-normale ne s'ajustent donc pas correctement à nos données. Un seuil grave égal à 900 est ainsi défini sur la distribution à l'aide des méthodes d'estimation de seuil classiques de la théorie des valeurs

extrêmes. Le coût marginal se réécrit :

$$CM = CM_{écrêté} + P_{grave} \cdot Sur\text{-}écrêtement \tag{1}$$

avec :

- CM le coût marginal de réassurance,
- $CM_{écrêté} = \begin{cases} CM & \text{si } CM \leq 900 \\ 900 & \text{sinon} \end{cases}$, modélisé par un GLM Gamma puis log-normal,
- $P_{grave} = \begin{cases} 0 & \text{si } CM \leq 900 \\ 1 & \text{sinon} \end{cases}$, modélisée par un GLM Bernoulli,
- $Sur\text{-}écrêtement = \begin{cases} 0 & \text{si } CM \leq 900 \\ CM - 900 & \text{sinon} \end{cases}$, modélisé par un arbre de décision de profondeur égale à trois.

Par la suite, deux algorithmes de Machine Learning sont également calibrés sur la base d'apprentissage : un *Random Forest* ainsi qu'un *XGBoost*.

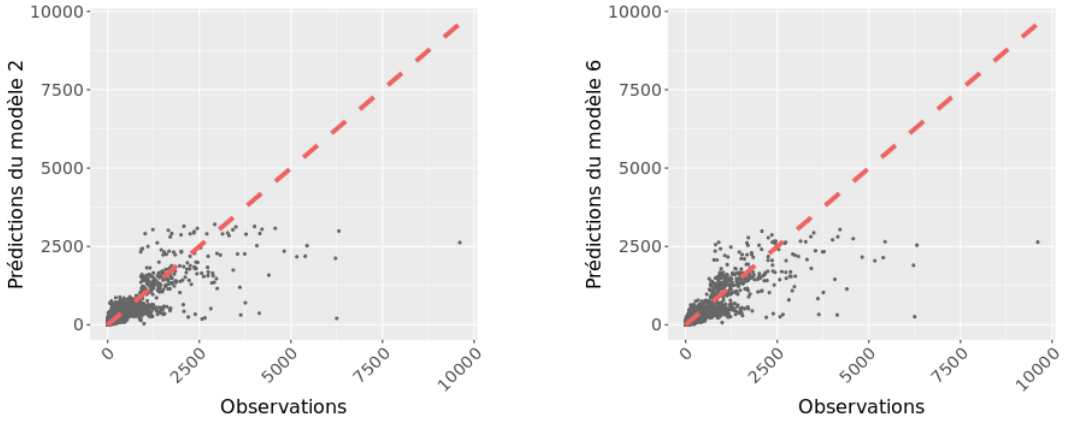
Pour finir, le coût marginal pouvant représenter la part de prime de réassurance pour l'année 2023 associée à un site, un modèle fréquence/coût moyen est également exploité. La fréquence annuelle d'un site RI se définit comme le nombre de fois où le site contribue à un évènement de l'YLT dont la perte est supérieure à la priorité du CAT XL. Ainsi, cette fréquence est déterministe pour les sites du portefeuille RI existant puisqu'elle se calcule simplement à partir de l'YLT construite à l'étape précédente. Afin d'être en mesure d'estimer la fréquence annuelle de dépassement de la priorité du CAT XL d'un nouveau site, une méthode d'interpolation spatiale est mise en place sur une grille de résolution de 200 mètres : l'*Inverse Distance Weighting*. Cette méthode permet d'obtenir une valeur de fréquence pour l'ensemble des coordonnées de la France métropolitaine.
Quant au coût moyen de cession de réassurance, il représente la contribution marginale moyenne d'un site aux cessions du CAT XL pour un évènement sur les 50 000 années de simulation. Ainsi, seul le coût moyen nécessite d'être modélisé par un algorithme de prédiction. On observe une nouvelle fois une queue épaisse à droite sur la distribution des coûts moyens de cession de réassurance des sites industriels du portefeuille de Generali. Par conséquent, un seuil grave égal à 4 800 est défini et le coût moyen se décompose de la même façon que le coût marginal de réassurance dans la formule 1, avec un seuil d'écrêtement de 4 800. La démarche de modélisation est similaire mais un arbre de décision de profondeur égale à deux est utilisé pour prédire la charge de sur-écrêtement.

Résultats obtenus

Les modèles utilisés dans le but de prédire les coûts marginaux des sites industriels pour le traité CAT XL et le péril tempête ont été comparés à l'aide de la RMSE et de la corrélation de Spearman, afin de s'assurer que l'ordre d'intensité des prédictions correspond bien à celui des observations. Les résultats sont présentés ci-dessous :

Numéro	Modèle	RMSE test	Corrélation de Spearman
1	Coût marginal écrêté/grave avec GLM Gamma	306	0.89
2	Coût marginal écrêté/grave avec GLM log-normal	295	0.87
3	*Random Forest*	414	0.77
4	*XGBoost*	317	0.90
5	Fréquence/coût avec GLM Gamma	306	0.90
6	Fréquence/coût avec GLM log-normal	298	0.90

Les modèles 2 et 6 font partie des modèles les plus performants, en terme de RMSE. Les graphiques des prédictions en fonction des observations des coûts marginaux de réassurance sont affichés pour ces deux modèles :



On observe que les coûts marginaux de réassurance élevés sont sous-estimés, à partir d'environ 2 500. La même observation est faite pour les autres modèles. Or, on cherche principalement à identifier les sites industriels contribuant fortement au prix de réassurance. Par conséquent, plutôt que de donner directement la valeur du coût marginal d'un site au souscripteur RI, le niveau d'intensité de ce coût sera plutôt fourni. Celui-ci pourra s'intégrer à un des outils de tarification du souscripteur RI sous forme d'alerte. Pour cela, trois catégories sont définies sur les coûts marginaux des sites industriels du portefeuille existant : faible, moyen et fort. Les prédictions des algorithmes testés sont alors classées et le pourcentage de bien classés par rapport aux observations est calculé :

Modèle	Pourcentage de bien classés		
	Faible	Moyen	Fort
Coût marginal écrêté/grave avec GLM Gamma	89	78	74
Coût marginal écrêté/grave avec GLM log-normal	86	85	70
Random Forest	88	67	53
XGBoost	89	77	59
Fréquence/coût avec GLM Gamma	89	79	74
Fréquence/coût avec GLM log-normal	87	84	74

D'après les résultats du tableau, l'algorithme de prédiction le plus approprié correspond au modèle fréquence/coût moyen pour lequel la charge écrêtée du coût moyen est modélisée par un GLM log-normal (modèle 6). Cependant, il est à noter que la fréquence utilisée dans le modèle est déterministe, puisqu'il est calibré sur les sites du portefeuille existant. Ainsi, l'erreur du modèle de prédiction est calculée seulement sur le coût moyen de cession de réassurance. Néanmoins, l'interpolation spatiale sur la fréquence appliquée à un nouveau site, non présent dans l'YLT, entraînera un biais supplémentaire. Ainsi, le modèle de séparation de la charge grave/attritionnelle du coût marginal pour lequel la charge écrêtée est modélisée par un GLM Gamma (modèle 1) pourra être préconisé.

Principale limite de l'approche

La principale limite de l'approche repose sur le manque de stabilité des coûts marginaux actuellement calculés sur le portefeuille RI existant, sur lesquels les modèles de prédiction sont construits. Tout d'abord, une variation significative de l'engagement assuré du portefeuille DAB Domestique à la maille département, constituant une variable explicative importante des modèles de prédiction, pourrait modifier la répartition de la prime de réassurance du CAT XL aux différents sites RI. De plus, si la prime de réassurance du traité CAT XL au renouvellement 2024 évolue de façon importante, l'ordre de grandeur des coûts marginaux évoluera également. Pour finir, si la version de RMS utilisée par Generali est améliorée et donc que la modélisation du péril tempête est différente, les deux changements évoqués précédemment pourraient également s'appliquer. Par conséquent, il sera nécessaire d'exécuter une nouvelle fois les modèles mis en place dans l'étude :

1. modélisation des pertes liées au péril tempête par le logiciel RMS,
2. construction de l'YLT via une méthode de simulation,
3. calibration du modèle de prédiction sur les coûts marginaux du portefeuille existant.

Toutefois, si l'engagement assuré du portefeuille DAB Domestique varie peu à la maille département dans les zones exposées à l'aléa tempête, il ne sera pas nécessaire de recalibrer le modèle de prédiction. Ainsi, ce modèle pourra être revu à des intervalles de temps plus élevés que les deux autres.

Continuité des travaux

Pour finir, il convient de revenir sur le périmètre d'étude. En effet, ce mémoire s'est concentré uniquement sur le traité CAT XL et le péril tempête. Or, la structure de réassurance du produit RI contient deux autres traités. En outre, le traité CAT XL couvre plusieurs périls naturels tels que les inondations et la grêle. Afin d'expliquer de manière plus globale la contribution d'un site industriel au prix de réassurance, il serait intéressant que ce type d'approche soit reproduit sur d'autres traités et/ou d'autres périls impactant le produit RI. Ainsi, la méthode développée dans ce mémoire pourrait servir de base aux approches applicables aux autres périls/traités de réassurance.

Synthesis note

Context and issues

According to France Assureurs, 2022 is the second most costly year in terms of climate-related claims for French insurers, with an estimated cost of 10 billion euros. Indeed, global warming is leading to an increase in the frequency and intensity of natural events. This phenomenon, combined with the rise in inflation rate, have led reinsurers to increase the price of reinsurance treaties for 2023.

The type of reinsurance treaty that generally enables the insurer to be covered against climatic claims is called excess of loss per event treaty and applies to the sum of claims due to the same event. This treaty is used by Generali France to cover claims compensated by property damage coverage, excluding motor, also known as Domestic Property (*DAB Domestique*). It covers products like MRH, MRC, MRA, RI, etc. This reinsurance treaty, called CAT XL, is triggered when a climatic event, other than drought, damages a group of sums insured in excess of a fixed amount. To reduce reinsurance cost, it is now important for insurance companies to know the areas exposed to climatic hazard and with a concentration of insured risks that can activate this type of treaty.

Identification of these areas can be obtained by estimating the individual contribution to the reinsurance price of the various insured risks contained in Generali France's Domestic Property portfolio. Indeed, a risk exposed to the storm hazard and located in an area containing many other insured risks will tend to participate heavily in the cessions of the treaty, directly impacting its price, which itself is calculated as an average of annual cessions. However, with the exception of the RI and MRI products, the sums insured of the risks are too low for this individual contribution to be significant for an event-level reinsurance treaty. Moreover, it has been observed that the damage caused to an industrial location can be very expensive, due to the consequences on the environment and the business interruption cover, which enables a company to offset the effects of a reduction in activity.

Then, the aim of this study is to calculate the marginal contribution of the industrial risks, that make up the RI product, to the price of an excess of loss per event reinsurance treaty, for storms. The operational use of this approach is based on providing an indicator of the reinsurance cost at risk level for RI underwriting. This means being able to predict the marginal contribution of a new industrial risk when it is acquired in the portfolio.

Scope of the study

The study relates to Domestic Property coverage, for which claims relating to natural perils (such as storms, hail, tornadoes, floods, frost and hurricanes) are reinsured by the CAT XL treaty. The products concerned are : MRH, MRC, MRA, MRI, RI, engineering and municipal. Only windstorm events are considered here. Indeed, historical cessions to

the CAT XL treaty due to windstorm events represent approximately 75% of total cessions. In addition, the individual contributions to the pricing of this treaty are calculated only for industrial risks, using the marginal cost.

Approach

The project has two main focuses. The first step involves calculating the marginal reinsurance cost, for the CAT XL treaty and storm events, of industrial risks in the existing portfolio, using CAT modelling software called RMS. Therefore, the reinsurance premium for the CAT XL treaty, corresponding to an average of the annual cessions on the losses modelled by the software, is calculated. The second step is to set up a calculation proxy that predicts the marginal cost of a new industrial risk based on its features.

Implementation on existing industrial risks

Generali France's RI portfolio contains more than 10 000 insured risks. In order to calculate their marginal contribution to the reinsurance cost of CAT XL treaty, for storm events, it is necessary to calculate the premium of this treaty in the first place. Indeed, the marginal reinsurance cost of the industrial risk n°i is defined by the following formula :

$$CM(i) = Premium(N) - Premium(N \backslash i)$$

where $Premium(N)$ represents the modelled reinsurance premium of the treaty, i.e. for the entire Domectic Property portfolio, and $Premium(N \backslash i)$ corresponds to this same premium when the industrial risk n°i is not present in the portfolio. Thus, the marginal reinsurance cost of risk n°i can be interpreted as being the additional reinsurance cost generated by acquiring risk n°i, in last place, in the RI portfolio.

First of all, we need to calculate $Premium(N)$, for cessions linked to storms. To do this, RMS software models the losses caused by a catalogue of historical and stochastic windstorm events on all Generali risks covered by Domestic Property insurance. The output of RMS is an *Event Loss Table* (ELT) for each product. An ELT includes a catalogue of windstorm events affecting the portfolio in question, as well as the associated average loss and standard deviation. Indeed, RMS software considers that the loss modelled for each event corresponds to an average loss and not to an actual loss. This is due to the secondary uncertainty caused by the lack of data, model uncertainty and variations in storm resistance in various French regions. In addition, an ELT contains the annual frequency of occurrence of events and the portfolio exposure to each event. Below is an example of an ELT with two windstorm events :

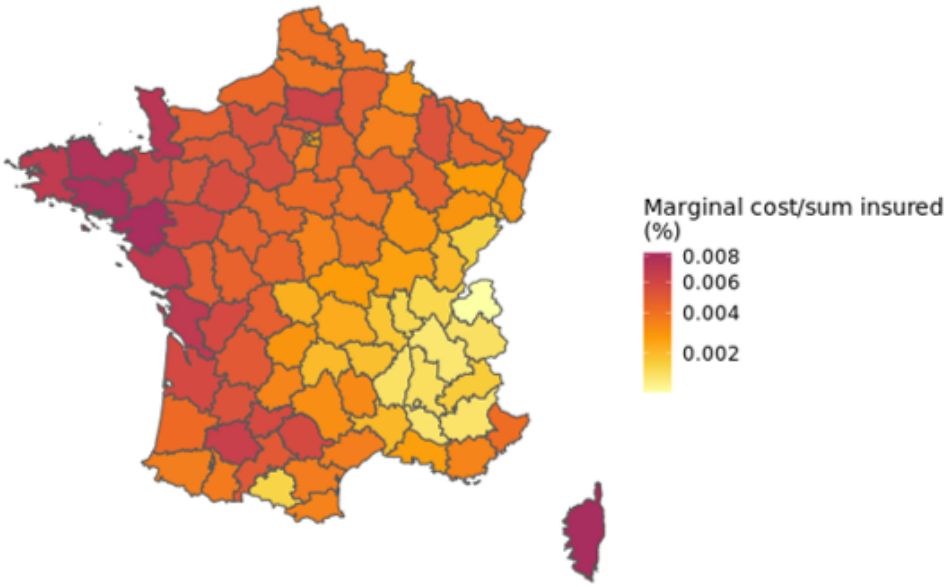
EVENT ID	FREQ	AVERAGE LOSS	STANDARD DEVIATION	EXPOSURE
1	0.01	500	350	850
2	0.005	250	470	500

Firstly, the seven ELT, associated with the seven products, are grouped together in order to obtain a complete catalogue of events and the associated losses summed over all the risks insured by the Domestic Property coverage.
As the reinsurance premium is calculated as an average of annual cessions, a Monte Carlo simulation method is used on the grouped ELT to obtain information on the year in which the events occur. To ensure reliable results, 50 000 years were selected for the simulation. For each year of simulation, in the first place, the windstorm events are randomly drawn using Poisson distributions whose parameters are based on their annual frequency of occurrence, and then the losses associated with each event drawn are simulated. Two different methods are used to simulate losses :

1. Secondary uncertainty is not taken into account, so each event is assigned its average loss, contained in the ELT.
2. Secondary uncertainty is taken into account and the destruction rates of each event are simulated with Beta distributions, whose parameters are calculated using the method of moments with the mean loss and standard deviation contained in the ELT.

The table resulting from the simulations is called the *Year Loss Table* (YLT). In order to price the CAT XL treaty for storms, we cede to the treaty the amount of loss greater than the CAT XL priority, for each year and each event in the YLT.
Compared to certain values calculated directly on the ELT supplied by RMS, such as the OEP curve values, the method of simulating losses with secondary uncertainty is chosen. Indeed, this method allows to consider the full range of losses, in addition to the average value. As a result, higher loss amounts are obtained. The $Premium(N)$ used is therefore the one calculated using the second approach.

Secondly, we need to calculate $Premium(N \backslash i)$ for all risks in the RI portfolio. To do this, we simply repeat the steps used in the approach described for calculating $Premium(N)$, but firstly, we remove the lines from the RI ELT associated with industrial risk n°i. In this way, the ELT of the seven products, including the new ELT associated with the industrial portfolio, are grouped together once again.The marginal reinsurance costs for the CAT XL treaty and storm events are obtained for all the risks insured under the RI product. The map of marginal costs divided by their sum insured per department is shown at the top of the next page.



The marginal reinsurance costs of the CAT XL treaty are a good representation of the storm hazard, since it is also high in the west of France according to the wind zonier constructed by Generali using RMS.

Prediction for a new industrial risk

The main objective of this study is to give the marginal reinsurance cost, for the CAT XL treaty and storm events, of an industrial risk to the RI underwriter during its acquisition. After having developed an approach for calculating the marginal costs of industrial risks in the existing portfolio in a robust manner, we now want to develop a calculation proxy for estimating this marginal cost for a new industrial risk when it is underwritten. Indeed, the underwriter needs this information in a reasonable computational time in order to use it. However, the approach presented above is not suitable as it would require to model the losses of the RI portfolio on RMS by adding the new risk to be underwritten. Consequently, marginal cost prediction algorithms are calibrated on the industrial locations in the existing portfolio, using their features such as their sum insured, their activity, their department, etc. These various features are entered into a pricing tool by the RI underwriter when a new risk is subscribed. In addition, a wind zoning plan constructed by department for the MRC product is added as a feature. The industrial risks in the current RI portfolio are therefore separated into a learning base and a test base.

In an insurance context, as the marginal costs of reinsurance belong to $\mathbb{R}_+$, we would tend to use a Gamma or log-normal GLM to predict them. However, the distribution of marginal costs for Generali's insured risks shows a heavy tail on the right. Therefore, the Gamma and lognormal distributions do not fit our data properly. Thus, a severe threshold equal to 900 is defined on the distribution using classical threshold estimation

methods from extreme value theory. The marginal cost is rewritten as :

$$CM = CM_{capped} + P_{severe} \cdot Over\text{-}capping \tag{2}$$

with :

— CM the marginal reinsurance cost,

— $CM_{capped} = \begin{cases} CM & \text{if } CM \leq 900 \\ 900 & \text{otherwise} \end{cases}$, modelled by a Gamma GLM then log-normal,

— $P_{severe} = \begin{cases} 0 & \text{if } CM \leq 900 \\ 1 & \text{otherwise} \end{cases}$, modelled by a Bernoulli GLM,

— $Over\text{-}capping = \begin{cases} 0 & \text{if } CM \leq 900 \\ CM - 900 & \text{otherwise} \end{cases}$, modelled by a decision tree of depth three.

Then, two Machine Learning algorithms are also calibrated on the learning base : a Random Forest and a XGBoost.

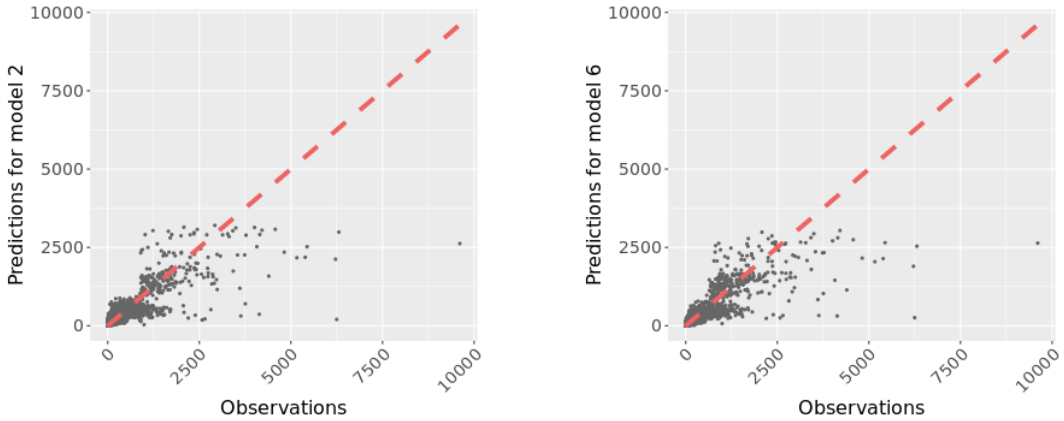
Finally, as the marginal cost can represent the contribution of a risk to the reinsurance premium for the year 2023, a cost/frequency model is also used.
The annual frequency of an industrial risk is defined as the number of times this risk contributes to an YLT event whose loss is greater than the CAT XL priority. This frequency is therefore deterministic for the risks contained in the existing RI portfolio, since it is calculated simply from the YLT constructed in the previous step. In order to be able to estimate the annual frequency with which a new risk exceeds the CAT XL priority, a spatial interpolation method is used on a grid with a resolution of 200 meters : Inverse Distance Weighting. This method makes it possible to obtain a frequency value for all the coordinates in metropolitan France.
The average reinsurance cession cost represents the average marginal contribution of a risk to CAT XL cessions for an event over the 50 000 years of simulation. Therefore, only the average cost needs to be modelled by a prediction algorithm. Once again, we observe a heavy tail on the right-hand side of the distribution of average reinsurance cession costs for industrial risks in Generali's portfolio. Thus, a severe threshold equal to 4 800 is defined and the average cost is decomposed in the same way as the marginal reinsurance cost in formula 2, with a capping threshold of 4 800. The modelling approach is similar but a decision tree of depth two is used to predict the over-capping losses.

Results achieved

The models used to predict the marginal costs of industrial risks for the CAT XL treaty and storms were compared using RMSE and Spearman correlation, to ensure that the order of intensity of the predictions correspond to that of the observations. The results are presented below :

Number	Model	RMSE test	Spearman correlation
1	Capped/severe marginal cost with Gamma GLM	306	0.89
2	Capped/severe marginal cost with lognormal GLM	295	0.87
3	Random Forest	414	0.77
4	XGBoost	317	0.90
5	Cost/Frequency with Gamma GLM	306	0.90
6	Cost/Frequency with lognormal GLM	298	0.90

Models number 2 and number 6 are among the best performing models according to its RMSE. Graphs of predictions as a function of observations of marginal reinsurance costs are displayed for these two models :



We note that the high marginal reinsurance costs are underestimated, from around 2 500. The same observation is made for other models. However, the main aim is to identify industrial risks that make a significant contribution to the reinsurance cost. Consequently, rather than giving directly the value of the marginal cost of a risk to the RI underwriter, the level of intensity of this cost will be provided instead. This could be integrated into one of the RI underwriter's pricing tools in the form of a warning. For this purpose, three categories are defined for the marginal costs of industrial risks in the existing portfolio : low, medium and high. The predictions of algorithms tested are then ranked and the percentage of well-ranked risks in relation to the observations is calculated :

Algorithm	Percentage of well-ranked		
	Low	Medium	High
Capped/severe marginal cost with Gamma GLM	89	78	74
Capped/severe marginal cost with lognormal GLM	86	85	70
Random Forest	88	67	53
XGBoost	89	77	59
Cost/Frequency with Gamma GLM	89	79	74
Cost/Frequency with lognormal GLM	87	84	74

According to the results in the table, the most appropriate prediction algorithm corresponds to the frequency/average cost model for which the loss capped of the average cost is modelled by a lognormal GLM (model 6). However, the frequency used in the model is deterministic, since it is calibrated on the locations in the existing portfolio. Thus, the error of the prediction model is calculated only on the average reinsurance cession cost. Nevertheless, the spatial interpolation on the frequency applied to a new location, not present in the YLT, will lead to an additional bias. Consequently, the model for separating severe/attritional load from marginal cost, for which the capped load is modelled by a GLM Gamma (model 1), may be recommended.

Main limitation of the approach

The main limitation of the approach is due to the lack of stability of the marginal costs currently calculated on the existing RI portfolio, on which the prediction models are built. Firstly, a significant variation in the sum insured of the Domestic Property portfolio at the departmental level, which is an important feature in the prediction models, could modify the allocation of the CAT XL reinsurance premium to the various RI risks. In addition, if the CAT XL treaty reinsurance premium at renewal 2024 changes significantly, the order of magnitude of marginal costs will also change. Finally, if the version of RMS used by Generali is improved, and the windstorm modelling is therefore different, the two changes mentioned above could also apply. Consequently, it will be necessary to run the models set up in the study one more time :

1. modelling of storm-related losses using RMS software,
2. construction of the YLT using a simulation method,
3. calibration of the prediction model on the marginal costs of the existing portfolio.

However, if the variation of sums insured of the Domestic Property portfolio is little at department level in the areas exposed to storm hazard, it will not be necessary to recalibrate the prediction model. Thus, this model may be revised at higher time intervals than the other two.

Project continuity

Finally, it is worth returning to the scope of the study. This project has only focused on the CAT XL treaty and windstorm events. However, the reinsurance structure of the RI product contains two other reinsurance treaties. In addition, the CAT XL treaty covers several natural perils such as floods and hailstorms. In order to provide a more comprehensive explanation of the contribution of an industrial risk to the reinsurance premium, it would be interesting for this type of approach to be reproduced for other treaties and/or other perils. In this way, the method developed in this study could serve as a basis for approaches applicable to other perils/reinsurance treaties.

Remerciements

Tout d'abord, je tiens à remercier mon tuteur d'alternance, Marco Brettschneider, ainsi que ma manager, Amandine Gomez, pour leur suivi régulier tout au long de l'alternance ainsi que pour leur aide indispensable à la réalisation de ce mémoire. Je remercie également Thibaut Richard pour son implication et ses recommandations essentielles à la réalisation de l'étude ainsi que pour la rédaction de ce mémoire.

Je tiens également à exprimer ma gratitude envers mes collaborateurs, Michael Reis ainsi que Yassine Laghzali, pour leurs précieux conseils sur ce projet.

Je souhaite également remercier ma tutrice de l'Institut des Actuaires, Camille Henaff, pour son aide lors de la rédaction de ce mémoire ainsi que pour la relecture.

Je tiens à remercier mon tuteur académique, Franck Vermet, qui m'a suivi régulièrement durant cette année d'alternance et qui m'a apporté de nombreux conseils afin de mener à bien ce projet, mais également pour l'ensemble des connaissances apprises au sein de l'EURIA. De manière plus générale, je tiens à remercier l'ensemble du corps enseignant de l'EURIA pour la qualité de la formation.

Pour finir, je tiens à exprimer ma gratitude envers les collaborateurs m'ayant aidé pour la rédaction du mémoire.

Avertissement

Pour des raisons de confidentialité, certains résultats ont été modifiés ou supprimés. Les ordres de grandeur sont restés identiques de sorte à ne pas modifier les conclusions originales.

Acronymes

AAD *Annual Aggregate Deductible* (franchise annuelle globale).

AAL *Annual Aggregate Limit* (plafond annuel global).

AEP *Aggregate Exceedance Probability* (probabilité de dépassement sur un montant annuel).

ELT *Event Loss Table*.

GR *Gross Loss* (perte nette d'assurance et de coassurance).

GU *Ground-Up Loss* (perte brute d'assurance, de coassurance et de réassurance).

i.i.d indépendantes et identiquement distribuées.

LCI limite contractuelle d'indemnisation.

OEP *Occurrence Exceedance Probability* (probabilité de dépassement sur la survenance d'un évènement).

RMSE *Root Mean Square Error* (Erreur quadratique moyenne).

S/P Ratio Sinistres sur Primes.

VAE Variable à expliquer.

YLT *Year Loss Table*.

Glossaire

as-if Une mise en *as-if* consiste à adapter des montants passés aux conditions économiques actuelles (suivant un indice d'inflation par exemple).

BM La garantie Bris de Machine, contenue dans la garantie DAB pour les sites industriels, permet d'être couvert en cas de dégâts sur des outils de production utilisés par l'assuré.

CCR La Caisse Centrale de Réassurance (CCR) est habilitée à couvrir en réassurance, avec la garantie de l'Etat, les risques résultant des effets des catastrophes naturelles faisant partie du régime CatNat.

coassurance La coassurance correspond à l'assurance d'un même risque par plusieurs assureurs.

CRESTA Le CRESTA, en France, correspond au code du département.

DAB L'assurance dommages aux biens est une branche de l'assurance destinée à couvrir tous les dégâts matériels. Elle concerne à la fois les dommages auto et les dommages sur les biens mobiliers.

DAB Domestique Il s'agit d'une branche de la garantie DAB. Celle-ci ne prend pas en compte les dommages auto, mais seulement les sinistres sur les biens mobiliers.

modélisation CAT Logiciel (comme AIR ou RMS) permettant de modéliser les pertes liées à des évènements climatiques sur un portefeuille assuré.

MRA L'assurance multirisques agricole, MRA, est un contrat qui permet de couvrir une exploitation agricole.

MRC Une assurance multirisques commerciale, ou MRC, est un contrat d'assurance couvrant toutes sortes de risques professionnels.

MRH Il s'agit du contrat d'assurance multirisques habitation permettant de protéger le patrimoine familial (habitation et mobilier) lorsque l'on est responsable ou victime d'un sinistre.

MRI Le produit MRI est l'assurance multirisques immeubles et s'adresse aux copropriétaires d'immeubles réunis en syndicat de copropriété.

PE La garantie Pertes d'Exploitation, contenue dans la garantie DAB pour les sites professionnels, permet d'être couvert en cas de sinistre causant l'interruption de l'activité économique du site et ainsi une diminution du chiffre d'affaires.

plafond Le plafond d'un traité de type excédent de sinistre correspond à la limite d'intervention du réassureur, c'est-à-dire à la somme de la priorité et de la portée.

RI Il s'agit de l'acronyme de : risques industriels. Celui-ci représente l'assurance spécifique aux entreprises du secteur de l'industrie.

régime CatNat Dispositif français d'indemnisation des catastrophes naturelles telles que les inondations, les sécheresses, les mouvements de terrain, les ouragans, les cyclones, les séismes et les avalanches.

Table des matières

Acronymes **xx**

Glossaire **xxi**

1 Introduction **1**

2 Réassurance et catastrophes naturelles **4**
2.1 Introduction à la réassurance 4
2.1.1 Définition et objectifs de la réassurance 4
2.1.2 Les différents types de traités de réassurance 5
2.1.3 Tarification des traités de réassurance non proportionnels 7
2.2 Réassurance des risques climatiques au sein de Generali 8
2.2.1 Périmètre de l'étude 8
2.2.2 Structure de réassurance de la garantie DAB Domestique 9
2.2.3 Contribution du portefeuille RI au traité CAT XL 12

3 Modélisation du péril tempête par RMS **14**
3.1 Présentation du logiciel RMS 14
3.1.1 Modules exposition et aléa 15
3.1.2 Module vulnérabilité 16
3.1.3 Module financier 17
3.1.4 Sources d'incertitude du modèle 18
3.2 Présentation des données en entrée de RMS 20
3.3 *Event Loss Tables* 22

4 Calcul de la prime de réassurance du CAT XL **25**
4.1 Récupération et agrégation des données en sortie de RMS 25
4.2 Construction d'une *Year Loss Table* 27
4.2.1 Hypothèses de calcul 27
4.2.2 Tirage des évènements tempête 29
4.2.3 Utilisation des pertes moyennes 31
4.2.4 Application de la structure de réassurance 32
4.3 Prise en compte de l'incertitude secondaire dans l'YLT 33
4.3.1 Simulation des pertes 33

4.3.2 Application de la structure de réassurance . . . 34
4.3.3 Principale limite de l'approche . . . 36
4.4 Validation de la méthode de simulation . . . 37
4.4.1 Choix de la méthode de simulation . . . 38
4.4.2 Choix du nombre d'années de simulation . . . 41
4.4.3 Comparaison à la prime de marché du traité CAT XL . . . 42

5 Coût marginal de réassurance des sites industriels du portefeuille existant 44
5.1 Description théorique du coût marginal . . . 44
5.2 Calcul des coûts marginaux des sites industriels . . . 45
5.3 Présentation des résultats . . . 46
5.4 Limites de l'approche . . . 48

6 Prédiction du coût marginal de réassurance d'un site industriel 50
6.1 Présentation de la base de données . . . 50
6.2 Modèles et concepts théoriques utilisés . . . 52
6.2.1 Modèle linéaire gaussien . . . 52
6.2.2 GLM . . . 52
6.2.3 Sélection de variables . . . 53
6.2.4 Théorie des valeurs extrêmes . . . 54
6.2.5 Algorithmes de Machine Learning . . . 54
6.2.6 Interprétabilité des modèles de Machine Learning . . . 56
6.2.7 Métriques d'évaluation utilisées . . . 56
6.3 Prédiction des coûts marginaux de réassurance . . . 58
6.3.1 Séparation des coûts de réassurance graves et attritionnels . . . 58
6.3.2 Algorithmes de Machine Learning . . . 65
6.4 Modèle fréquence/coût moyen de cession de réassurance . . . 67
6.4.1 Présentation du modèle . . . 67
6.4.2 Utilisation d'une méthode d'interpolation spatiale sur la fréquence 69
6.4.3 Prédiction des coûts moyens de cession de réassurance . . . 72
6.5 Définition de catégories sur le coût marginal de réassurance . . . 76
6.6 Limites de l'approche de prédiction . . . 78

7 Conclusion 82

Annexes 85
A.1 Lois statistiques usuelles . . . 85
A.1.1 Loi de Bernoulli . . . 85
A.1.2 Loi de Poisson . . . 85
A.1.3 Loi uniforme continue . . . 85
A.1.4 Loi normale . . . 85
A.1.5 Loi log-normale . . . 85
A.1.6 Loi Gamma . . . 86

A.1.7 Loi Bêta . . . 86
A.1.8 Loi du χ^2 . . . 86
A.1.9 Loi de Student . . . 86
A.2 Corrélations . . . 87
A.2.1 Corrélation de Pearson . . . 87
A.2.2 V de Cramér . . . 87
A.2.3 Matrice de corrélation du *dataset* utilisé pour la prédiction des coûts marginaux de réassurance . . . 87
A.3 Ajustement de lois statistiques . . . 88
A.3.1 Méthode des moments . . . 88
A.3.2 Maximum de vraisemblance . . . 89
A.4 Détermination du seuil grave sur les coûts moyens de cession de réassurance 89

Liste des figures **92**

Liste des tableaux **94**

Bibliographie **96**

Chapitre 1

Introduction

Au mois d'août 2020, au Liban, a eu lieu un accident industriel causé par l'explosion d'un stock de nitrate d'ammonium à Beyrouth. En 2021, d'après le journal L'Argus de l'Assurance, l'estimation des pertes économiques s'élève à plus de 6 milliards d'euros dont 1.2 milliard de pertes assurées. Les sites industriels ont des montants de sinistres très élevés en raison des conséquences de leurs dégâts sur l'environnement mais également du fait de la couverture perte d'exploitation (PE) de la garantie dommages aux biens (DAB). En effet, celle-ci permet l'indemnisation des sinistres causant une interruption ou une réduction de l'activité économique. Les incendies et les catastrophes naturelles impactant des sites industriels, appartenant au produit RI, sont donc très coûteux pour les assureurs.

Toutefois, le réchauffement climatique tend à augmenter considérablement l'intensité et la fréquence des catastrophes naturelles. En effet, l'année 2022 apparaît comme la deuxième année la plus coûteuse en terme de sinistralité climatique pour les assureurs du marché français selon France Assureurs : le coût est estimé à 10 milliards d'euros. Afin d'être en mesure d'indemniser ce type de sinistres, les compagnies d'assurance font appel à des réassureurs. Parmi eux, un organisme public, nommé CCR, couvre une partie des dégâts dus aux périls reconnus par le régime CatNat. Cette couverture est obligatoire pour les assurés payant une prime DAB. En ce qui concerne les biens assurés par la garantie DAB Domestique, c'est-à-dire la garantie DAB hors auto, la surprime payée est fixée à 12% de la prime d'assurance correspondante. Ce régime d'indemnisation des catastrophes naturelles, créé en 1982, repose sur la mutualisation afin de couvrir les dégâts assurés lors de la survenance de risques naturels élevés. Le péril tempête n'étant pas considéré comme une catastrophe naturelle au titre du régime CatNat de la CCR, les assureurs font appel à des réassureurs privés afin qu'ils couvrent leurs sinistres climatiques au-dessus d'un certain montant. Ainsi, les assureurs paient à leur réassureur une prime de réassurance, prise en compte dans la prime d'assurance des sites assurés sous forme de taux de chargement.

Actuellement, au sein de Generali France, les chargements de réassurance sont ap-

pliqués globalement à l'ensemble des sites industriels. Il n'existe donc aucune politique de tarification au niveau site en terme de réassurance. La structure de réassurance des sinistres climatiques en vigueur s'applique à l'ensemble des produits couverts par la garantie DAB Domestique (tels que le RI, le MRC, le MRH, etc.) et la distinction la plus fine sur le prix de cette structure est réalisée à la maille produit. Les prix de réassurance de chaque produit sont calculés au prorata des assiettes de primes et sinistres historiques couverts par la garantie DAB Domestique. Cependant, en raison de la hauteur des sommes assurées du produit RI, la contribution individuelle des sites industriels au coût total de réassurance du portefeuille DAB Domestique est élevée. Ainsi, il serait intéressant de disposer d'un indicateur sur le prix de réassurance au niveau site pour le portefeuille RI, basé sur cette contribution. Dans un premier temps, le calcul de cet indicateur sur le portefeuille existant permettrait d'identifier les sites contribuant fortement au prix de réassurance actuellement en vigueur et notamment les zones dans lesquelles l'exposition devrait être limitée. L'objectif principal de cette étude est de mettre cet indicateur à disposition du souscripteur RI afin d'évaluer le risque de réassurance porté par une affaire nouvelle potentielle dans le portefeuille et ainsi le guider lors de l'étape de tarification.

La structure de réassurance actuellement en vigueur pour le produit RI, appartenant au portefeuille DAB Domestique de Generali, est présenté sur la figure 1.1.

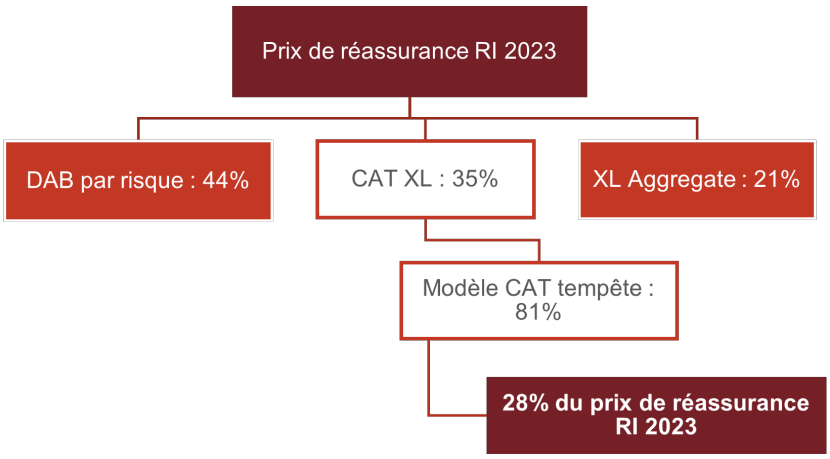


FIGURE 1.1 – Structure de réassurance 2023 de Generali France pour le produit RI

L'étude portera sur le traité de type excédent de sinistre par évènement, appelé CAT XL, pour les dégâts causés par le péril tempête. Ainsi, 28% du prix de réassurance du portefeuille RI sera expliqué. Ce traité s'applique à la maille évènement pour l'ensemble des contrats couverts par la garantie DAB Domestique, dont les contrats du produit RI. Le traité DAB par risque correspond à un traité de réassurance de type excédent de sinistre par risque. Quant au XL Aggregate, il regroupe l'ensemble des sinistres climatiques res-

tants pour différents produits d'assurance après application des deux traités précédents. Il fonctionne comme un traité stop-loss (2.1.2). La répartition du prix de réassurance aux sites industriels s'effectuera donc seulement pour le traité CAT XL et le péril tempête.

Afin de calculer la contribution des sites RI au traité CAT XL, il sera nécessaire de modéliser les pertes liées au péril tempête. Comme l'historique des sinistres climatiques est faible et que l'intensité des évènements naturels a tendance à augmenter, les logiciels de modélisation CAT sont de plus en plus utilisés par les assureurs. En effet, des évènements stochastiques sont construits et les pertes associées aux sites assurés du portefeuille de l'assureur sont estimées. Afin de calculer une prime de réassurance à partir des tables fournies par le logiciel, il est nécessaire de mettre en place une méthode de simulation de type Monte Carlo ([Vuong, 2011], [Wang, 2012]). L'objectif principal de l'étude étant l'utilisation d'un indicateur sur le prix de réassurance par la souscription RI, la contribution au coût de réassurance de chaque site industriel sera évaluée par le coût marginal. En effet, d'après les études réalisées auparavant ([Sermet, 2018], [Lair, 2011]), le coût marginal permet de quantifier l'apport de coût de réassurance qu'engendre l'acquisition d'un nouveau site à un portefeuille assuré.

Grâce au développement des logiciels de modélisation CAT, il est aujourd'hui possible d'obtenir le détail des pertes modélisées au niveau site. Néanmoins, les tables de données en sortie sont très volumineuses dès lors que le nombre de sites du portefeuille est élevé. Ainsi, le calcul des coûts marginaux de réassurance sur les sites RI du portefeuille existant s'effectuera grâce à un logiciel de modélisation CAT, afin d'obtenir des résultats robustes, mais un proxy de calcul devra être mis en place pour permettre l'utilisation à la souscription. Celui-ci pourra se baser sur les caractéristiques du nouveau site industriel renseignées par le souscripteur (somme assurée, localisation, domaine d'activité, etc.) et également sur les coûts marginaux déjà calculés sur le portefeuille existant. Le souscripteur pourra donc avoir une indication quant à l'apport en prime de réassurance qu'engendre la souscription d'un nouveau site industriel, en un temps de calcul raisonnable. Dans une précédente étude, des algorithmes de prédiction, tels qu'un GLM et certains algorithmes de Machine Learning, ont été mis en place afin d'estimer un taux de chargement de réassurance ainsi que des cessions à un traité de type excédent de sinistre par évènement pour un nouveau site ([Sermet, 2018]).

Malgré la mise en place de l'approche seulement sur le traité de réassurance CAT XL et le péril tempête, l'objectif prochain serait d'être en mesure de calculer la contribution marginale d'un site industriel au coût de réassurance total du produit RI. Les méthodes décrites dans ce mémoire pourront donc servir de base au calcul des coûts marginaux de réassurance des sites RI pour les autres traités et/ou les autres périls.

Chapitre 2

Réassurance et catastrophes naturelles

2.1 Introduction à la réassurance

2.1.1 Définition et objectifs de la réassurance

La réassurance se définit comme étant « l'assurance des institutions d'assurance » ([Braham, 2022]). Cela signifie que les assureurs peuvent céder une partie de leurs risques aux réassureurs. Afin de se réassurer, l'assureur, également appelé cédante, signe un contrat avec un réassureur, appelé traité de réassurance. A travers ce traité, les objectifs de l'assureur sont de différentes natures :

- protéger son compte de résultat contre une dérive de sinistralité (survenance d'évènements extrêmes ou sur-fréquence de sinistres),
- protéger ses fonds propres en réduisant l'exigence de capital et le risque de ruine,
- augmenter sa capacité de souscription,
- assurer des types de risques nouveaux en utilisant l'expertise du réassureur.

On trouve sur le marché assurantiel trois types de réassurance : la réassurance obligatoire, la réassurance facultative et la facultative-obligatoire. Lorsque la couverture est de type obligatoire, les deux parties ont l'obligation d'accepter ou de céder les risques identifiés dans un traité de réassurance (zone géographique, nature des risques, exclusions, etc.), signé au préalable. Au contraire, dans le cas de la réassurance facultative, l'assureur propose au réassureur de lui céder un risque ou un ensemble de risques et celui-ci peut décider de ne pas les accepter. Ce type de réassurance est généralement utilisée pour les grands risques industriels ou les risques atypiques. Pour finir, la réassurance facultative-obligatoire est une couverture de type facultative mais que le réassureur est obligé d'accepter, si la cédante souhaite lui céder certains risques importants.

2.1.2 Les différents types de traités de réassurance

Au sein de la réassurance obligatoire, il existe différents types de traités de réassurance :

— la réassurance proportionnelle comprenant les traités excédent de plein et quote-part,
— la réassurance non proportionnelle comprenant les traités excédent de sinistre par risque, excédent de sinistre par évènement et stop-loss.

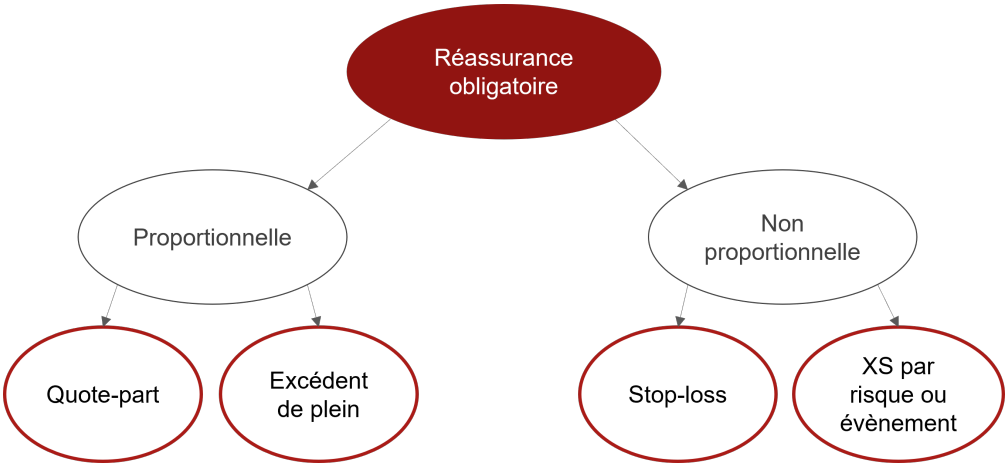


FIGURE 2.1 – Les différents types de traités de réassurance obligatoire

La réassurance proportionnelle permet à l'assureur de céder un même pourcentage de sinistres et de primes au réassureur. Pour les traités quote-part, le taux de cession est le même pour tous les assurés et est décidé dans le contrat alors que pour les traités en excédent de plein, le taux de cession est propre à chaque assuré et est déterminé en fonction de ses capitaux assurés. Par exemple, pour un traité quote-part, si le réassureur s'engage à couvrir 80% des sinistres de la cédante, alors celle-ci lui verse également 80% des primes qu'elle a reçues des assurés concernés par le traité. Pour un excédent de plein, le réassureur intervient uniquement sur les polices dépassant un certain montant de valeur assurée : les petits risques sont donc conservés. Le taux de cession des primes et sinistres est défini de la façon suivante : $\frac{VA - PR}{VA}$, avec VA la valeur assurée du risque et PR le plein de rétention déterminé par le traité. Le plein de rétention correspond au montant de risque assuré que l'assureur conserve, et donc, qu'il ne cède pas au réassureur. Ainsi, dans le cas d'un traité excédent de plein, le taux de cession correspond au pourcentage de somme assurée cédée au réassureur.

Dans le cas de la réassurance non proportionnelle, le montant de sinistre pris en charge par le réassureur est déterminé grâce à un seuil et une limite d'intervention, sans lien direct avec la prime ou la somme assurée du risque. Ainsi, la prime est tarifée par

le réassureur et n'est plus égale au taux de cession multiplié par les primes directes d'assurance, comme pour les traités de réassurance proportionnels. Dans ce mémoire, deux types de traité seront abordés : l'excédent de sinistre par risque et l'excédent de sinistre par évènement. Ces deux traités suivent la même méthodologie mais diffèrent au niveau de l'échelle sur laquelle sont appliquées les conditions financières de réassurance : le sinistre ou l'évènement. Ils se notent généralement « Portée XS Priorité ». La partie conservée par l'assureur correspond à la priorité. Au-dessus de la priorité, le réassureur ne s'engage à prendre qu'une partie du sinistre, appelée la portée. Le schéma 2.2 permet de comprendre leur fonctionnement.

FIGURE 2.2 – Exemple de traités excédent de sinistre par risque et excédent de sinistre par évènement

Le traité stop-loss, également appelé excédent de perte, concerne un ensemble d'affaires et s'applique lorsque le montant des sinistres est supérieur à un certain pourcentage, à déterminer lors de la signature du traité. Voici un exemple de traité stop-loss en fonction du ratio S/P : 50% XS 100%. Si la prime estimée pour l'année est égale à 10 M€, le réassureur couvrira dès que le montant total des sinistres encourus dépasse le seuil de 10 M€ = 100%× 10 M€ (et jusqu'à 15 M€ = (100% + 50%)× 10 M€).

En outre, il existe plusieurs clauses financières dans la définition des traités non proportionnels : l'AAD, l'AAL et les reconstitutions. L'AAD correspond à une franchise annuelle à atteindre pour que l'assureur puisse céder ses sinistres au réassureur. L'AAL correspond à un montant limite que le réassureur s'engage à payer au maximum sur une année. Par exemple, dans le traité 8 XS 12 AAD 15 AAL 36, le réassureur commencera à payer les sinistres lorsque la cession annuelle atteindra 15 M€ (AAD). Celui-ci arrêtera le remboursement des sinistres quand la cession annuelle dépassera les 36 M€ (AAL). L'AAL peut également être exprimée sous la forme de reconstitutions. Une reconstitution signifie que l'assureur a cédé au réassureur une portée entière sur l'année. Ainsi, dans les traités de réassurance, il est souvent inscrit un nombre de reconstitutions maximal. En effet, le traité 8 XS 12 avec 3 reconstitutions signifie que le réassureur remboursera au maximum quatre fois (3 + 1 reconstitutions) la portée dans l'année suivant la signature du traité.

2.1.3 Tarification des traités de réassurance non proportionnels

Contrairement aux traités proportionnels, le réassureur tarifie un prix pour la signature du traité. Il s'agit du montant que paie la cédante au réassureur pour l'application du traité. La prime de réassurance peut être déterminée de différentes façons : par tarification à l'expérience, par tarification à l'exposition ou en utilisant une méthode stochastique ([Buntun, 2016]).

La tarification à l'expérience se base sur les sinistres historiques et est appelée méthode *Burning Cost*. Elle est utilisée pour tarifer les tranches travaillantes des traités de réassurance, c'est-à-dire les tranches dont la priorité est régulièrement atteinte. La méthode *Burning Cost* repose sur les étapes suivantes :

1. Le réassureur met en *as-if* les sinistres projetés à l'ultime de l'historique du portefeuille concerné.
2. Il calcule ensuite les cessions conformément aux conditions du traité dont le prix doit être déterminé.
3. Enfin, celui-ci calcule les cessions annuelles pour les différentes années composant l'historique.
4. Le prix du traité s'obtient en effectuant la moyenne de ces cessions annuelles.
5. Le taux de réassurance s'obtient en divisant la moyenne des cessions annuelles par

les primes d'assurance pour l'année à venir.

La modélisation stochastique s'articule autour des axes suivants :

1. Le réassureur met en *as-if* les sinistres projetés à l'ultime de l'historique du portefeuille concerné.
2. Il ajuste ensuite des lois statistiques de coût et fréquence sur les sinistres historiques.
3. Il simule des sinistres sur de nombreuses années à l'aide des lois adaptées.
4. Enfin, celui-ci calcule les cessions annuelles des sinistres simulés.
5. La prime de réassurance s'obtient en calculant la moyenne des cessions annuelles.

Cette méthode permet ainsi de pallier aux limites de la première. En effet, suivant la loi de sévérité utilisée, des sinistres extrêmes pourront être simulés, bien qu'ils n'aient pas été observés dans le passé. Le réassureur sera ainsi en mesure de tarifer des tranches peu ou pas travaillantes du traité. Cette méthode se base encore une fois sur les sinistres historiques de la cédante.

Au contraire, en appliquant la méthode de tarification à l'exposition, le réassureur se base sur le profil du portefeuille actuel de la cédante. Le portefeuille concerné par la prime de réassurance à déterminer est divisé en plusieurs *band profiles*, aussi appelées bandes de risques. Celles-ci dépendent de la somme assurée du risque ou de sa perte maximale possible. L'assureur doit déterminer, pour chacune de ces bandes, le pourcentage de primes de risque à céder au réassureur grâce à une courbe d'exposition ([Bernegger, 1997]). La prime de risque correspond à l'espérance du montant des sinistres survenus sur une certaine période de temps. Des courbes d'exposition de marché sont fournies par des réassureurs tels que Swiss Re et Lloyd's. En général, le modèle d'exposition est utilisé en complément de la tarification à l'expérience, dans le cas où la cédante dispose de peu de sinistres historiques pour les tranches hautes du traité.

Dans ce mémoire, seule la prime d'un traité de type excédent de sinistre par évènement sera calculée. La tarification consistera à calculer la moyenne des cessions annuelles au traité. La méthode sera expliquée en détails par la suite. Celle-ci combine à la fois les méthodes de tarification stochastique et de tarification à l'exposition. En effet, les données prises en entrée correspondent au portefeuille actuel de Generali dont les pertes associées sont d'abord estimées par un logiciel de modélisation CAT puis simulées grâce à des lois de probabilité.

2.2 Réassurance des risques climatiques au sein de Generali

2.2.1 Périmètre de l'étude

Dans la présente étude, on s'intéressera à la réassurance des risques climatiques pour les sinistres de la garantie DAB Domestique du portefeuille de Generali. Celle-ci com-

porte trois types de couverture : le bâtiment, le contenu et la perte d'exploitation (PE). Cette dernière concerne les entreprises et leur permet de compenser une diminution du chiffre d'affaires et/ou de faire face à des charges fixes suite à un sinistre. Les principaux périls couverts par la garantie DAB sont l'incendie, le vol et les risques climatiques, parmi lesquels se trouve la tempête. Seul ce péril sera modélisé ici. Au sein de Generali, sept branches de produits sont concernées par la garantie DAB Domestique : MRA, MRC, MRH, MRI, *engineering* (branche construction) et *municipal* (branche collectivités) ainsi que le RI. On prendra en compte seulement les polices d'assurance de Generali situées en France métropolitaine.

Au cours de ce projet, on s'intéressera plus particulièrement aux sites appartenant au portefeuille industriel (RI), dont on calculera la contribution au coût de réassurance. En effet, les risques industriels correspondent à une somme assurée globale moins importante que le MRH dans la garantie DAB mais représentent individuellement des engagements assurés très élevés. En effet, l'engagement assuré moyen d'une habitation MRH est environ 100 fois moins élevée que celui d'un site RI. Cela s'explique par le fait que les sites industriels disposent de bâtiments plus grands que les habitations, de contenus plus chers (machines, etc.) ainsi que de la couverture PE qui n'existe pas pour les particuliers. Par conséquent, si un site industriel est touché par une catastrophe naturelle, les pertes engendrées pour l'assureur seront importantes. Le RI, chez Generali, se composent des entreprises suivantes :

— les entreprises industrielles et commerciales,
— les prestataires de services (bureaux, formation, santé, etc.),
— les centres commerciaux,
— les bâtiments abritant des activités industrielles,
— etc.

Le portefeuille est actuellement composé de plus de 10 000 sites. La particularité d'un contrat RI est que celui-ci peut couvrir plusieurs sites, situés à des endroits différents et appartenant à la même entreprise. En effet, une police RI peut contenir aussi bien un seul site que 100 sites différents.

2.2.2 Structure de réassurance de la garantie DAB Domestique

La structure de réassurance utilisée par Generali contre les risques climatiques et l'incendie impactant le portefeuille DAB Domestique est décrite sur la figure 2.3. Les périls naturels concernés sont : la tempête, la grêle, la tornade, l'inondation, le gel et les ouragans. Les évènements sécheresse sont gérés par une autre structure de réassurance chez Generali, et notamment par le régime CatNat de la CCR. Pour rappel, cette étude ne traitera que des dommages liés aux évènements tempête. Le traité DAB par risque est un excédent de sinistre par risque et le traité CAT XL est un excédent de sinistre par évènement (2.1.2).

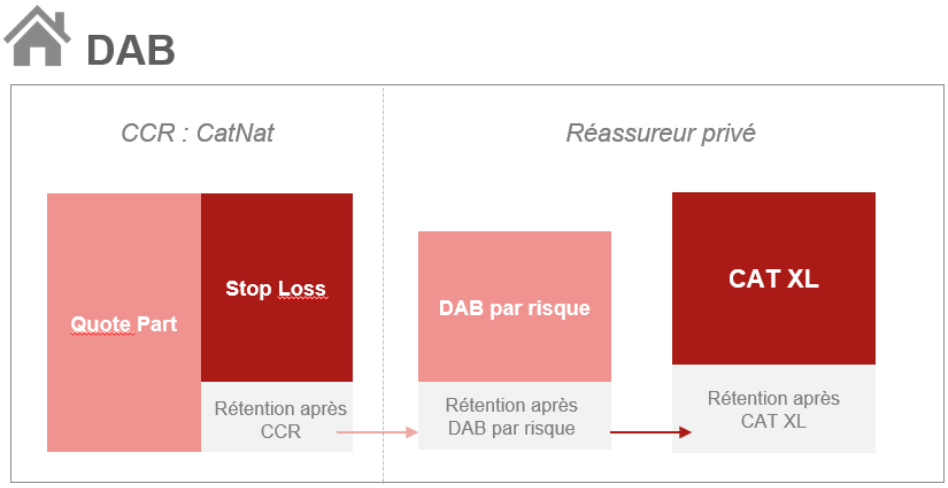


FIGURE 2.3 – Structure de réassurance des risques climatiques du portefeuille DAB Domestique

Tous les périls ne sont pas cessibles au titre du régime CatNat de la CCR. En effet, la grêle, la tornade, le gel et la tempête ne sont pas considérés comme des catastrophes naturelles. Par conséquent, les assureurs ne bénéficieront pas des traités quote-part et stop-loss du schéma 2.3 pour les sinistres liés à ces périls.

Pour résumer, les sinistres du portefeuille DAB Domestique liés au péril tempête sont réassurés par la structure suivante :

1. application d'un traité excédent de sinistre par risque, appelé traité DAB par risque,
2. agrégation des sinistres nets du DAB par risque au niveau évènement,
3. application d'un traité excédent de sinistre par évènement, appelé CAT XL, aux sinistres nets du DAB par risque.

Les deux traités cités précédemment sont divisés en plusieurs tranches, c'est-à-dire en plusieurs « sous-traités ». Les tranches des traités excédent de sinistre par risque et excédent de sinistre par évènement utilisés chez Generali France sont décrites sur les figures 2.4 et 2.5. Chaque tranche du traité CAT XL dispose d'un nombre de reconstitutions suffisant à la couverture de la compagnie d'assurance. De plus, pour ce traité, la priorité et la portée changent tous les ans et sont déterminées grâce à la sinistralité historique du portefeuille DAB Domestique ainsi qu'à la modélisation stochastique du principal péril touchant l'exposition de ce portefeuille, la tempête. Ce dernier est modélisé à l'aide d'un logiciel de modélisation CAT appelé RMS.

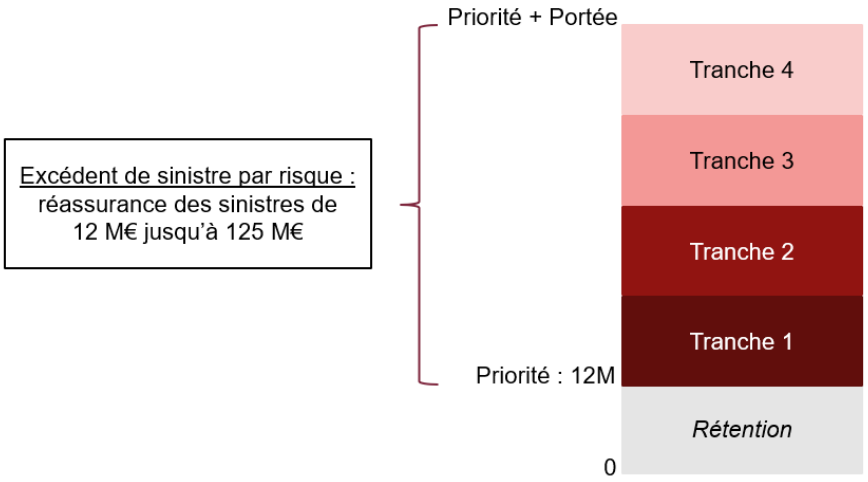


FIGURE 2.4 – Tranches du traité excédent de sinistre par risque, le DAB par risque

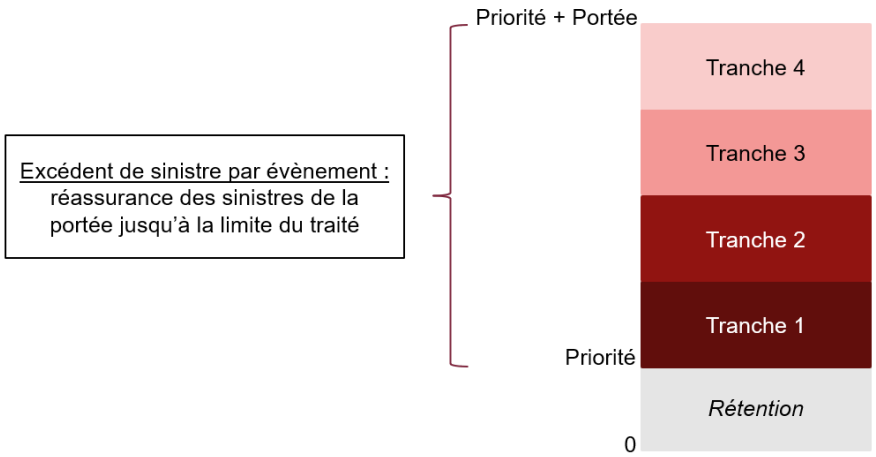


FIGURE 2.5 – Tranches du traité excédent de sinistre par évènement, le CAT XL

Afin de calculer les contributions individuelles des sites RI au prix du traité CAT XL, la première étape de ce mémoire consiste à calculer cette prime de réassurance, c'est-à-dire le paiement que Generali effectuera à son réassureur afin d'être couvert par le traité 2.5. Cependant, seul le péril tempête sera pris en compte. Ce risque naturel contribue à 74% des cessions historiques du CAT XL. Or, les cessions historiques n'ont atteint que la première tranche du traité. Concernant les tranches plus hautes, aucune cession historique n'a été observée. On émet alors l'hypothèse que seuls les évènements tempête peuvent atteindre ces tranches. En effet, le péril inondation étant considéré comme une

catastrophe naturelle par la CCR, contrairement à la tempête et à la grêle, il donnera lieu à des cessions au CAT XL limitées puisqu'une partie des sinistres aura d'abord été transférée à la CCR. Ainsi, il sera plus difficile pour un évènement inondation d'atteindre les tranches hautes du traité CAT XL. De plus, on considèrera qu'un évènement grêle ne pourra atteindre 100 M€, puisque ce type de catastophe est très localisé. Finalement, en considérant l'ensemble des tranches du CAT XL, les évènements tempête contribuent à environ 91% de ce traité de réassurance.

2.2.3 Contribution du portefeuille RI au traité CAT XL

L'objectif de cette étude est de quantifier la contribution marginale des sites industriels assurés chez Generali au traité CAT XL, pour le péril tempête. Ainsi, dans cette partie, on observera la répartition des cessions historiques du CAT XL entre différents produits, et notamment pour le portefeuille RI.

Péril	Tempête	Grêle	Inondations (après CCR)
Contribution totale	74%	15%	4%
MRH	44%	8%	2%
MRC	9%	4%	1%
RI	3%	2%	1%

TABLE 2.1 – Contribution des périls et produits aux cessions historiques du CAT XL

D'après la table 2.1, la tempête est le péril contribuant le plus aux cessions historiques du CAT XL. Il y a également d'autres périls intervenant dans le traité de réassurance, autres que ceux cités ci-dessus, tels que le gel. En s'intéressant aux produits, on remarque que les cessions historiques des sites industriels liées aux évènements tempête ne représentent que 3% des cessions historiques totales. Cependant, le montant moyen d'un sinistre sur un site industriel pour le péril tempête est beaucoup plus élevé que pour une habitation ou un bien commercial :

Péril	Tempête
Montant moyen d'un sinistre	x
MRH	$0.3\ x$
MRC	$0.6\ x$
RI	$1.9\ x$

TABLE 2.2 – Montant moyen des sinistres historiques tempête

Ainsi, on peut voir qu'un sinistre tempête sur un site industriel a un coût environ six fois plus élevé que sur un bien MRH. Cela est dû au fait que les biens immobiliers du produit RI ont une surface plus élevée que ceux des particuliers et disposent de la couverture PE en plus. Il est donc pertinent d'étudier la contribution marginale de chaque

site industriel (individuellement) au coût de réassurance. Il aurait été difficile d'appliquer cette approche pour chaque site du produit MRH en vue de leur exposition. De plus, en terme d'assiettes de primes, le RI contribue à 17% au prix de marché du traité CAT XL 2023.

Pour conclure, cette étude se focalisera sur la répartition du coût de réassurance d'un traité de type excédent de sinistre par évènement, appelé CAT XL, aux sites industriels, pour le péril tempête. Ce traité de réassurance s'appliquant à l'ensemble des portefeuilles couverts par la garantie DAB Domestique, seule une partie de son prix sera alloué car la contribution ne sera calculée que pour les sites du portefeuille RI. Cela s'explique par le fait que le coût d'un sinistre RI est généralement beaucoup plus élevé que les dommages causés sur d'autres portefeuilles couverts par la garantie DAB Domestique, autres que le MRI, en cas de péril climatique. Ainsi, les polices du portefeuille industriel participent davantage aux cessions de réassurance. En outre, en vue des cessions historiques cédées au traité CAT XL, les évènements tempête sont les plus contributeurs. Il est donc pertinent, dans un premier temps, de développer l'approche sur ce péril afin d'expliquer une grande partie du coût de réassurance total du traité CAT XL. Toutefois, les sinistres historiques n'ont permis d'atteindre que les premières tranches du traité. Il est donc nécessaire de modéliser des pertes pouvant potentiellement atteindre l'ensemble des tranches. Deux solutions existent : la modélisation stochastique (2.1.3) ou l'utilisation des logiciels de modélisation CAT, plus adaptés à l'estimation des sinistres appartenant au domaine climatique.

Chapitre 3

Modélisation du péril tempête à partir du logiciel RMS

Dans ce chapitre, un logiciel de modélisation CAT, appelé RMS (*Risk Management Solutions*), est présenté ([de la Cruz, 2009]). La version 18 de ce logiciel est utilisée pour l'étude ([RMS, 2019]). Il existe également un autre logiciel très utilisé sur le marché assurantiel, dans le but de modéliser les pertes liés à des évènements climatiques : *AIR Worldwide*. Ce type de logiciel est de plus en plus utilisé par les assureurs et réassureurs pour la modélisation des catastrophes naturelles. En effet, en raison du réchauffement climatique et donc de l'augmentation potentielle de l'intensité et de la fréquence des périls naturels non encore observée dans l'historique, un modèle de coût et fréquence ne permet plus de représenter une sinistralité assez extrême, notamment pour la garantie tempête. En outre, le logiciel RMS est utilisé pour cette étude puisqu'il a été validé en interne au sein de Generali. Il s'agit notamment d'un modèle de référence sur le marché pour modéliser le péril tempête en Europe.

3.1 Présentation du logiciel RMS

Le logiciel RMS est un outil de modélisation CAT permettant d'évaluer l'impact des catastrophes naturelles telles que les inondations ou les tempêtes sur les sites assurés du portefeuille d'une compagnie d'assurance. Ce type de logiciel modélise les pertes associées à un catalogue d'évènements climatiques dont l'intensité est fixée par celui-ci. Concernant le péril tempête, RMS dispose d'un catalogue d'environ 20 000 évènements historiques et stochastiques, modélisés à partir des tempêtes historiques (Lothar, Martin, etc.) touchant le portefeuille DAB de Generali France. Ce dernier calcule les pertes engendrées par ces évènements sur un ou plusieurs portefeuilles, en fonction de la vitesse du vent. Plus généralement, la modélisation des risques de catastrophes naturelles se décompose en quatre modules : exposition, aléa, vulnérabilité et financier.

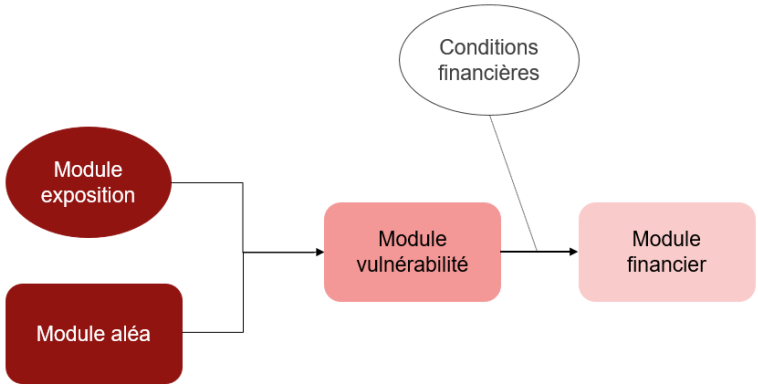


FIGURE 3.1 – Schéma de modélisation des catastrophes naturelles

3.1.1 Modules exposition et aléa

Le module exposition contient les données de la compagnie d'assurance et plus particulièrement les informations relatives aux sites assurés du portefeuille considéré. Les trois paramètres primordiaux pour la modélisation sont : la valeur assurée, la localisation du site ainsi que la typologie du bâti (maison, appartement, commerce, etc.). D'autres caractéristiques viennent se rajouter afin d'affiner les analyses d'impact des évènements : les matériaux de construction, l'année de construction, le nombre d'étages, les détails sur la toiture, etc. En effet, un bâtiment en béton sera moins sensible à des tempêtes qu'un bâtiment en bois. Plus ces données en entrée seront précises et meilleurs seront les résultats du modèle.

Le module aléa contient un catalogue d'évènements à la fois historiques et stochastiques. Les évènements stochastiques sont simulés à partir des évènements historiques, dans lesquels différents paramètres sont modifiés : intensité de l'évènement et zone géographique parcourue. Par exemple, les évènements du péril tempête se distinguent par la vitesse de vent, généralement exprimée en km/h. La carte 3.2 répertorie les principales tempêtes historiques survenues en France, et sur lesquelles RMS se base pour construire son catalogue d'évènements tempête.

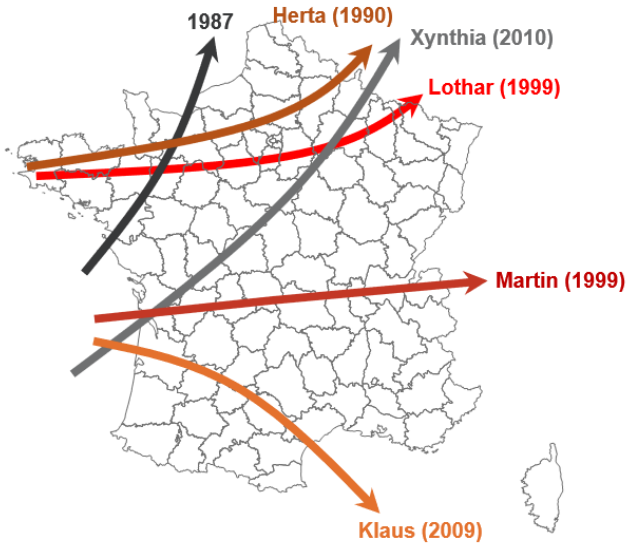


FIGURE 3.2 – Principaux évènements tempête historiques en France

L'intérêt des logiciels de modélisation CAT par rapport à la modélisation basée sur les sinistres historiques est que ceux-ci permettent de simuler des évènements extrêmes non encore observés, et de tenir compte de l'exposition actuelle du portefeuille d'assurés. En effet, l'historique des sinistres d'assurance liés aux évènements climatiques est faible en France. Ainsi, les modèles de coût et fréquence généralement utilisés pour estimer la sinistralité ne sont pas assez robustes et n'ont pas une vision assez long terme : ce type de modèles ne prend pas en considération les évènements naturels ayant une période de retour élevée. En faisant varier les paramètres (trajectoire, intensité, durée) de quelques scénarios de tempêtes historiques, RMS parvient à modéliser un grand nombre de tempêtes stochastiques.

3.1.2 Module vulnérabilité

Le module vulnérabilité peut se définir comme la fragilité des enjeux (module exposition) vis-à-vis de l'aléa. Cette étape consiste donc à reunir les deux modules précédents. Dans ce module, des courbes de vulnérabilité sont construites. Il s'agit d'estimer le taux de destruction moyen des sites assurés en fonction de l'intensité des évènements. Les courbes de vulnérabilité sont différenciées selon certaines caractéristiques des sites (matériau de construction, activité professionnelle, etc.). Le taux de destruction représente le pourcentage de somme assurée impactée par l'évènement climatique. Afin d'obtenir la perte économique d'un site assuré impacté par un évènement, il suffit de multiplier le taux de destruction associé aux caractéristiques du site et à l'intensité de l'évènement considéré par l'exposition du site assuré à cet évènement. Il est à noter que cette perte est, dans un premier temps, brute d'assurance et de réassurance.

Les courbes de vulnérabilité de RMS sont construites à l'aide de plusieurs variables :

1. les fonctions primaires de vulnérabilité : la classe d'occupation (locataire ou propriétaire), les matériaux de construction ainsi que les caractéristiques du toit (la principale source de dommages dus aux tempêtes en Europe), la hauteur du bâtiment et l'âge de construction ;
2. les caractéristiques régionales au sein de la France : celles-ci comprennent les différences de vulnérabilité entre plusieurs zones en France (normes de construction, type de toiture, cartes du vent de référence) ;
3. les caractéristiques secondaires des bâtiments : cette composante est utilisée s'il y a des données manquantes dans les fonctions primaires, ce sont notamment des précisions sur le toit du site.

La figure 3.3 présente quelques exemples de courbes de vulnérabilité suivant l'intensité de la tempête et les matériaux de construction d'un bien assuré.

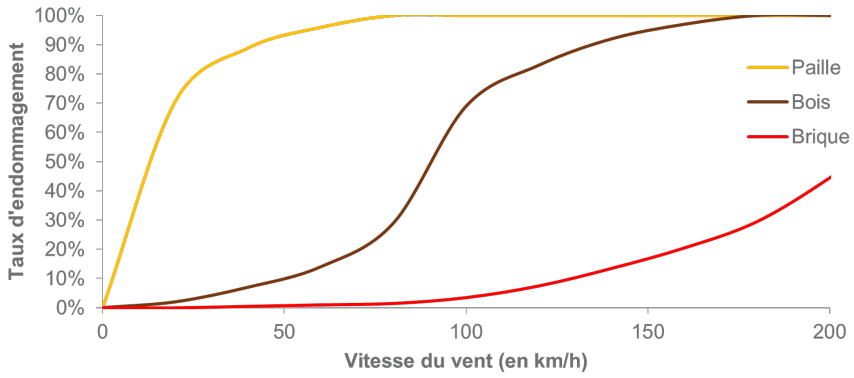


FIGURE 3.3 – Courbes de vulnérabilité pour la tempête

3.1.3 Module financier

Cette dernière étape de la modélisation CAT consiste à appliquer aux pertes économiques l'ensemble des conditions financières des contrats d'assurance et de réassurance afin d'obtenir la perte réellement payée par l'assureur. Par exemple, on peut appliquer à la perte modélisée d'une police des conditions d'assurance telles que la franchise et des conditions de coassurance. Sur RMS, une perte brute d'assurance et de réassurance se nomme *Ground-Up Loss* (GU) et une perte à laquelle on a appliqué les conditions d'assurance s'écrit *Gross Loss* (GR). Différentes métriques sont utilisées afin de mettre en forme les pertes obtenues. On retrouve principalement les courbes OEP et AEP. La courbe OEP représente la probabilité que l'évènement maximal sur une période d'un an dépasse un certain montant de perte. Quant à l'AEP, elle consiste à sommer les pertes

par année puis à évaluer la probabilité que cette somme de pertes annuelle dépasse un certain montant. Ainsi la courbe AEP est toujours supérieure ou égale à la courbe OEP. Ces deux courbes sont tracées suivant deux axes : la probabilité de dépassement annuelle et le montant de pertes associé. La probabilité de dépassement est généralement exprimée en période de retour. Si on note RP la période de retour et EP la probabilité de dépassement annuelle, on a l'égalité suivante : $RP = {}^{1}/_{EP}$. La période de retour d'un évènement correspond à la durée moyenne avant qu'un évènement de même intensité se reproduise. Par la suite, la valeur de courbe OEP pour une période de retour RP, en années, sera notée « OEP RP Y ». Les courbes OEP et AEP du portefeuille DAB Domestique de Generali France, calculées par le logiciel RMS pour le péril tempête, sont décrites par la figure 3.4.

FIGURE 3.4 – Courbes OEP et AEP du portefeuille DAB Domestique calculées par RMS pour la tempête

La courbe OEP de la figure 3.4 fournie par RMS permet notamment de fixer la priorité ainsi que la portée du traité CAT XL chaque année. Ainsi, sur le schéma 2.5, le plafond du traité CAT XL correspond à l'OEP 250Y. Seul le péril tempête est pris en compte dans la détermination des conditions du traité de réassurance puisque ce dernier représente le péril climatique causant le plus de dégâts sur le portefeuille DAB Domestique à la maille évènement.

3.1.4 Sources d'incertitude du modèle

Il existe deux principales sources d'incertitude dans la modélisation de RMS : l'incertitude primaire et l'incertitude secondaire.

Tout d'abord, l'incertitude primaire porte sur la survenance d'un évènement ou non. Cette dernière concerne également le type d'évènement se produisant, dans le cas de survenance. Par exemple, il existe une incertitude sur le nombre de survenance d'évènements tempête l'année prochaine en France. En outre, il existe également une incertitude concernant l'événement qui pourrait se produire, en terme d'intensité et de localisation.

Deuxièmement, l'incertitude secondaire porte sur le montant de perte réelle sachant qu'un certain évènement du catalogue RMS est survenu. Malgré un ensemble de connaissances scientifiques sur l'évènement et l'exposition, on ne peut prédire les pertes réelles qui se produiront. Ainsi, le logiciel RMS fournit une meilleure estimation appelée perte moyenne. Par conséquent, pour un même évènement (c'est-à-dire une intensité fixée) et un site déterminé, le taux de destruction ne correspond pas à une seule valeur mais est en réalité une distribution. Celui-ci suit une loi Bêta (A.1.7). La loi Bêta est pertinente pour modéliser un taux de destruction car :

— la distribution est définie sur le domaine $[0; 1]$, correspondant à l'intervalle de valeurs possibles pour un taux de destruction (le sinistre lié à un site assuré ne peut être négatif ou supérieur à l'exposition du site),
— suivant les valeurs des paramètres de loi α et β, la forme de la distribution peut être très différente : cela permet ainsi de couvrir une grande diversité de taux de destruction.

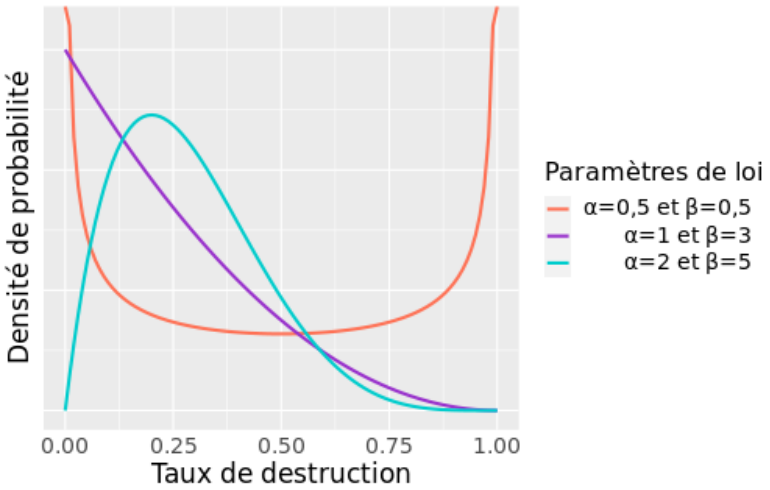


FIGURE 3.5 – Densité de lois Bêta

L'étendue de la perte autour de la moyenne est mesurée par le coefficient de variation, correspondant à l'écart-type de la perte divisée par la perte moyenne. Plus ce coefficient est élevé et plus l'incertitude secondaire est importante.
Ci-dessous sont énumérées les principales sources d'incertitude secondaire :

— l'aléa : cela correspond à l'incertitude quant au phénomène de variation de la vitesse du vent lors de la traversée de certaines zones,
— la vulnérabilité : des bâtiments présentant les mêmes caractéristiques peuvent être affectés différemment par un événement spécifique (qualité de construction différente, caractéristiques inconnues, etc.),
— le modèle : ce type d'incertitude est due au niveau de détails des paramètres du modèle (résolution géographique, etc.),
— les données en entrée : cela correspond à la complétude des données en entrée de RMS (année de construction, etc.) qui sont fournies par Generali sur le portefeuille assuré.

3.2 Présentation des données en entrée de RMS

Afin de calculer les pertes dues à la tempête sur le portefeuille DAB Domestique de Generali France, plusieurs variables sont rentrées dans RMS afin que le logiciel puisse procéder à la modélisation. Elles sont de différents types :

1. les sommes assurées bâtiment, contenu et PE dans le cas des professionnels,
2. des données géographiques telles que la longitude et la latitude, le code postal, le code CRESTA (correspondant au département en France) ainsi que le pays,
3. des caractéristiques du bâtiment comme le nombre d'étages, le type de bâtiment (entrepôt, usine, etc.) ainsi que l'année de construction,
4. et enfin, le type d'occupation (propriétaire ou locataire).

Plus les données en entrée de RMS seront précises et meilleurs seront les résultats de la modélisation des pertes.

L'objectif de la présente étude étant de calculer l'impact des sites RI sur le coût de réassurance du traité CAT XL, les pertes modélisées pour ce produit devront donc être fiables à la maille site. Les taux de remplissage des différentes variables citées ci-dessus sont affichés dans le tableau 3.1.

Variable	Taux de remplissage (en %)
Sommes assurées	100
Longitude et latitude	100
Code postal	100
CRESTA	100
Nombre d'étages	0
Type de bâtiment	27
Année de construction	0

TABLE 3.1 – Taux de remplissage des variables pour la modélisation RMS sur les sites du portefeuille RI

On remarque ainsi qu'il y a peu d'informations disponibles sur les caractéristiques des bâtiments industriels. De plus, malgré le taux de remplissage à 100% des variables longitude et latitude, les sites industriels ne sont pas parfaitement géocodés. En effet, la précision de géocodage est soit au numéro, soit à la voie ou soit au code postal. Une précision à la voie ou au code postal augmente l'incertitude du modèle RMS. De plus, même si certains sites industriels sont annoncés géocodés au numéro, leur adresse n'est pas toujours correctement renseignée. Cela est dû au fait qu'une même police RI peut contenir plusieurs sites à des endroits différents et il se peut que tous les sites aient l'adresse du siège social. Ainsi, il a été possible d'identifier certains sites dont l'adresse était incorrecte afin de les supprimer de l'étude.

Précision de géocodage	Répartition avant suppression	Répartition après suppression
Numéro	67	64
Voie	25	27
Code postal	8	9

TABLE 3.2 – Précision de géocodage des sites du portefeuille RI avant et après suppression des sites mal géocodés

Afin de connaître l'exposition aux évènements tempête de l'ensemble du portefeuille RI de Generali, une carte montrant l'engagement assuré tempête par département est présentée sur la figure 3.6. Cet engagement correspond à la somme assurée des sites pour la garantie tempête nette de coassurance et de limite contractuelle d'indemnisation (LCI).

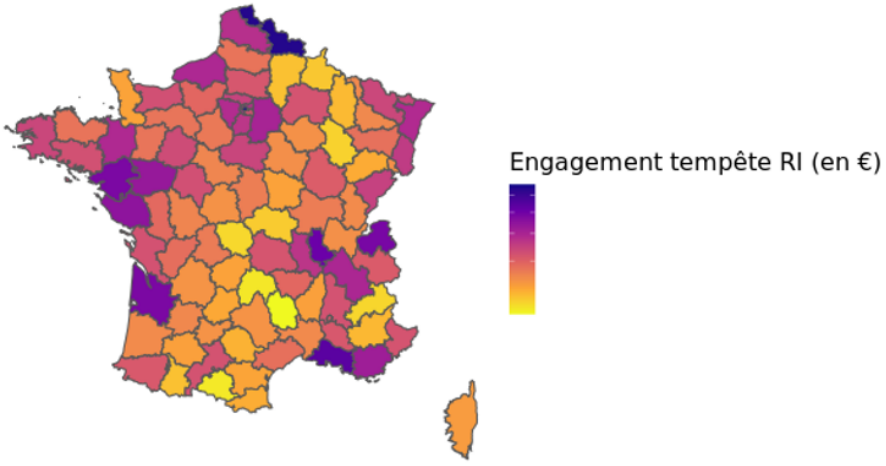


FIGURE 3.6 – Engagement assuré tempête par département du produit RI

3.3 *Event Loss Tables*

En sortie du logiciel RMS, on dispose premièrement des courbes AEP et OEP présentées sur la figure 3.4, qui sont calculées de façon analytique. On dispose également d'une *Event Loss Table* (ELT) au niveau d'agrégation choisi : portefeuille, police, site ou niveau de couverture du site (bâtiment, contenu, PE). En effet, une même police peut contenir plusieurs sites. Une ELT présente, pour l'ensemble du catalogue d'évènements impactant notre portefeuille en entrée, les pertes moyennes associées ainsi que d'autres informations complémentaires. L'agrégation choisie définit le niveau de détails sur les pertes. Une ELT au niveau portefeuille comporte les pertes cumulées sur l'ensemble des sites composant le portefeuille, par évènement. En revanche, une ELT au niveau site contient les pertes par évènement site par site. La table 3.3 présente un exemple d'ELT au niveau site, pour un portefeuille étant impacté par deux évènements et comportant deux sites.

EVENT ID	RATE	LOCATION ID	MEAN	SDC	SDI	EXPOSURE
1	0.01	A	80	50	15	120
1	0.01	B	100	70	80	110
2	0.005	A	105	45	60	130
2	0.005	B	100	15	60	135

TABLE 3.3 – Exemple d'ELT au niveau site

Les différents éléments présents dans le tableau 3.3 sont décrits ci-dessous :
— EVENT ID : l'identifiant unique de l'évènement,
— RATE : la fréquence annuelle de survenance de l'évènement,
— LOCATION ID : l'identifiant unique du site assuré,
— MEAN : la perte moyenne associée à l'évènement pour le site considéré,
— SDC : l'écart-type corrélé de la perte pour le site et pour l'évènement,
— SDI : l'écart-type indépendant de la perte pour le site et pour l'évènement,
— EXPOSURE : l'exposition du site assuré à l'évènement.

Il convient de décrire plus précisément la séparation de l'écart-type de la perte en une partie corrélée (« SDC ») et une partie indépendante (« SDI »). Il est à noter que cette séparation apparaît dans les ELT à partir du niveau site. Celle-ci est donc absente dans les ELT au niveau couverture. De plus, l'écart-type total de la perte associée à un événement est la somme de ces deux éléments. Cette séparation permet de considérer la corrélation spatiale entre les différents sites assurés. « La corrélation et l'indépendance spatiale des pertes entre plusieurs sites pour un même évènement correspondent au degré de corrélation ou d'absence de corrélation entre les dommages subis par ces sites » ([RMS, 2019]).
Afin d'expliquer ces termes, considérons l'« EVENTID » n°1 et les sites A et B du tableau 3.3. L'objectif est de réunir l'ELT contenant les deux sites au niveau portefeuille, et ainsi d'obtenir une seule ligne par évènement. Afin d'obtenir la perte moyenne, il s'agira simplement de sommer les pertes moyennes des deux sites. En revanche, sommer également

les écarts-types totaux signifient que les pertes associées à l'évènement sont entièrement corrélées. Si les pertes des deux sites sont corrélées à 100%, les dégâts subis par un site pour un évènement tempête sont proportionnels à ceux subis par l'autre site, pour ce même évènement. Supposons les deux hypothèses suivantes séparément :

1. Les sites sont entièrement corrélées pour l'évènement considéré.
2. Le sites sont indépendants pour l'évènement considéré.

On note σ_A (respectivement σ_B) l'écart-type de la perte pour le site A (respectivement pour le site B) et σ l'écart-type de la perte totale du portefeuille associée à l'évènement n°1. Ainsi, $\sigma_A = SDC_A + SDI_A = 65$ et $\sigma_B = SDC_B + SDI_B = 150$. Dans le cas général, on a la formule suivante :

$$\sigma^2 = \sigma_A^2 + \sigma_B^2 + 2\ \mathrm{Cov}\,(A, B)$$

En considérant le coefficient de corrélation $\rho_{AB} = \dfrac{\mathrm{Cov}\,(A, B)}{\sigma_A\,\sigma_B}$, la formule devient :

$$\sigma^2 = \sigma_A^2 + \sigma_B^2 + 2\,\rho_{AB}\,\sigma_A\,\sigma_B$$

Les deux hypothèses précédentes impliquent ainsi les résultats suivants :

1. $\sigma^2 = \sigma_A^2 + \sigma_B^2 + 2\,\sigma_A\,\sigma_B = (\sigma_A + \sigma_B)^2$ avec $\rho_{AB} = 1$
2. $\sigma^2 = \sigma_A^2 + \sigma_B^2$ avec $\rho_{AB} = 0$

En réalité, RMS donne un poids de corrélation aux pertes par région et péril. Cela signifie que les pertes dues à un même évènement sont en partie corrélées. Pour la tempête en Europe, considérons le poids de corrélation w. Ainsi, on a la formule suivante :

$$\sigma = w\,(\sigma_A + \sigma_B) + (1 - w)\,\sqrt{\sigma_A^2 + \sigma_B^2}$$

Afin d'obtenir les « SDC » et « SDI » au niveau portefeuille pour l'évènement 1, on a : SDC $= w\,(\sigma_A + \sigma_B)$ et SDI $= (1 - w)\,\sqrt{\sigma_A^2 + \sigma_B^2}$. Dans le but d'agréger l'ELT 3.3 au niveau portefeuille, la méthode de calcul présentée ci-dessus pourrait être appliquée pour connaître l'écart-type total associé à chaque évènement. Cependant, comme l'ELT au niveau site dispose de la séparation de l'écart-type en une partie corrélée et une partie indépendante, une méthode d'agrégation plus simple existe ([Homer et Li, 2017]). Si on disposait en sortie de RMS d'une ELT au niveau couverture, il aurait été nécessaire d'appliquer l'approche de calcul de l'écart-type ci-dessus (ne disposant pas de la séparation sur l'écart-type à cette maille). Les différents calculs d'agrégation effectués pour l'« EVENTID » n°1 sont :

— $MEAN = MEAN_A + MEAN_B = 80 + 100$,
— $SDC = SDC_A + SDC_B = 50 + 70$,
— $SDI = \sqrt{SDI_A^2 + SDI_B^2} = \sqrt{15^2 + 80^2}$,
— $EXPOSURE = EXPOSURE_A + EXPOSURE_B = 120 + 110$.

Ainsi, l'ELT 3.3 est agrégée au niveau portefeuille :

EVENT ID	RATE	MEAN	SDC	SDI	EXPOSURE
1	0.01	180	120	81	230
2	0.005	205	60	85	265

TABLE 3.4 – ELT 3.3 au niveau portefeuille

Finalement, RMS permet la modélisation des pertes sur un portefeuille assuré pour différents périls climatiques, et notamment pour le péril tempête. Ce logiciel met à disposition des tables contenant un certain niveau de détails sur les pertes pour un catalogue d'évènements historiques et stochastiques, ainsi que certains résultats comme les courbes AEP et OEP, qu'il calcule de façon analytique directement sur les ELT. Par la suite, l'unité des ELT présentées dans l'étude est l'euro.

Chapitre 4

Calcul de la prime de réassurance du CAT XL à l'aide de méthodes de simulation

Afin de calculer la contribution des sites RI au coût de réassurance du traité CAT XL, il convient de calculer la prime de réassurance de ce traité. Le logiciel RMS a permis de modéliser les pertes sur le portefeuille DAB Domestique pour le catalogue d'évènements tempête survenant en France métropolitaine. Ce chapitre présente l'approche permettant d'obtenir le prix modélisé que Generali paie à son réassureur pour l'application du traité CAT XL à ses sinistres. Il est à noter que la prime de réassurance se calcule comme une moyenne de cessions annuelles. Or, en sortie de RMS, il est mis à disposition des *Event Loss Table* (ELT) ne contenant aucune information sur l'année de survenance des évènements. Il sera donc nécessaire d'effectuer des simulations à l'aide de lois statistiques afin d'obtenir des pertes réelles par évènement et par année. La table contenant ces informations s'appelle la *Year Loss Table* (YLT). La première étape consistera donc à construire cette YLT. Ensuite, il s'agira de calculer les cessions annuelles sur cette table de pertes afin d'en obtenir la moyenne. Les résultats obtenus dans cette partie seront comparés à la prime de réassurance de marché ainsi qu'à d'autres grandeurs calculées analytiquement par le logiciel RMS, c'est-à-dire directement sur l'ELT et n'utilisant donc aucune méthode de simulation. La méthode de construction de l'YLT employée s'inspire du document [Homer et Li, 2017]. Il est à noter que la prime du traité calculée dans ce chapitre est une prime calculée à l'aide des modélisations RMS, et ne correspond pas à la prime de réassurance de marché réellement payée par Generali.

4.1 Récupération et agrégation des données en sortie de RMS

Le calcul de la prime de réassurance s'effectue sur les pertes de l'ensemble des polices couvertes par la garantie DAB Domestique puisque les cessions au traité CAT XL sont

réalisées à la maille évènement. Ainsi, le traité CAT XL est activé si la somme des pertes dues à un évènement tempête sur le portefeuille DAB Domestique entier est supérieure à la priorité. Si seules les pertes sur le portefeuille RI sont comptabilisées, l'activation du traité par évènement est extrêmement rare puisqu'il manque une partie importante des produits concernés. Par conséquent, en sortie de RMS, sept ELT sont mises à disposition. En effet, la garantie DAB Domestique concerne les portefeuilles suivants : MRH, MRC, MRI, MRA, *engineering* (branche construction), *municipal* (branche collectivités) et RI. L'objectif de l'étude étant de calculer le coût marginal des sites industriels au traité, voici les niveaux de détail des différentes ELT récupérées sur RMS :

— maille portefeuille pour les produits : MRH, MRC, MRI, MRA, *engineering* et *municipal*,
— maille site pour le produit RI.

Le niveau de détail choisi pour le portefeuille RI ne sera pas utile dans le calcul de la prime de réassurance globale du portefeuille DAB Domestique mais seulement dans le calcul des contributions des sites assurés. Ainsi, l'ELT niveau site RI a été agrégée au niveau portefeuille d'après la méthode décrite dans la partie 3.3.

L'objectif de cette partie est d'obtenir l'ELT au niveau portefeuille contenant les pertes cumulées de l'ensemble des produits couverts par la garantie DAB Domestique afin d'être en mesure de calculer la prime de réassurance du CAT XL. Il est ainsi nécessaire de regrouper les ELT au niveau portefeuille des sept différents produits en une seule table. Pour cela, la même méthode que celle décrite dans la partie 3.3 est utilisée. Cependant, au lieu de passer du niveau site au niveau portefeuille, il s'agit d'un regroupement de plusieurs portefeuilles. Ci-dessous un exemple de deux ELT (correspondant à deux produits différents et donc à deux portefeuilles assurés différents) décrivant la perte due à un même évènement (EVENTID n°1). On prendra l'exemple des produits MRH et MRC.

EVENT ID	RATE	MEAN	SDC	SDI	EXPOSURE
1	0.01	800	300	400	1 200

Table 4.1 – Exemple d'ELT au niveau portefeuille du produit MRH

EVENT ID	RATE	MEAN	SDC	SDI	EXPOSURE
1	0.01	900	200	500	1 300

Table 4.2 – Exemple d'ELT au niveau portefeuille du produit MRC

D'après la méthode décrite en 3.3, l'agrégation de ces deux ELT donne la table 4.3. Ainsi, dans cette ELT, on retrouve à la fois les sites assurés du produit MRH et les sites assurés du produit MRC, regroupés dans un seul portefeuille.

Finalement, en appliquant cette méthode aux sept ELT récupérées sur RMS, on obtient l'ELT de l'ensemble du portefeuille DAB Domestique. Une ligne correspond à

EVENT ID	RATE	MEAN	SDC	SDI	EXPOSURE
1	0.01	1 700	500	640	2 500

TABLE 4.3 – Regroupement au niveau portefeuille des ELT MRH 4.1 et MRC 4.2

un évènement tempête du catalogue RMS et contient la perte *Gross Loss* (perte nette d'assurance et de coassurance) associée pour l'ensemble du portefeuille DAB Domestique. Cette table comporte 19 912 évènements tempête. Par la suite, cette dernière est appelée « ELT-DABDOM ».

EVENT ID	RATE	MEAN	SDC	SDI	EXPOSURE
3180656	0.00028	16M	10M	966k	330Mrd
3180668	0.00005	12M	7M	1 620k	195Mrd
3180700	0.00007	22M	10M	960k	184Mrd
3180717	0.00038	46M	21M	3 064k	326Mrd
3180675	0.00010	21M	11M	1 088k	239Mrd

TABLE 4.4 – Premières lignes de l'ELT au niveau portefeuille de la garantie DAB Domestique après agrégation des sept produits

4.2 Construction d'une *Year Loss Table*

4.2.1 Hypothèses de calcul

Avant d'expliquer la méthode de calcul de la prime de réassurance du CAT XL, il convient de présenter les différentes hypothèses conduites par le logiciel de modélisation CAT, RMS, sur les pertes modélisées :

1. La survenance annuelle de l'évènement n°i suit une loi de Poisson (A.1.2) de paramètre λ_i correspondant aux valeurs de la colonne « RATE » de l'ELT-DABDOM.
2. La survenance d'un évènement est indépendante de la survenance d'un autre évènement. Cette hypothèse peut s'appliquer dans le cadre du péril tempête mais n'est pas totalement vérifiée pour les séismes notamment en cas de répliques sismiques.
3. La survenance de l'évènement n°i ne dépend pas des survenances passées de cet évènement.
4. Le nombre annuel d'évènements suit une loi de poisson de paramètre $\lambda = \sum_{i=1}^{n} \lambda_i$ où n est le nombre d'évènements présents dans l'ELT-DABDOM ($n = 19312$). Cette hypothèse découle des précédentes. Ainsi, pour connaître le nombre annuel moyen d'évènements tempête, il suffit de sommer les valeurs de la colonne « RATE » de l'ELT-DABDOM puisque dans le cas d'une loi de Poisson de paramètre λ, l'espérance est égale à λ. On compte en moyenne environ 47.8 tempêtes par an.

5. Le taux de destruction associé à chaque évènement suit une loi Bêta (3.1.4). Pour l'évènement n°i, les paramètres de loi Bêta sont :

$$\alpha_i = \frac{\mu_i^2\,(1-\mu_i)}{\sigma_i^2} - \mu_i \text{ et } \beta_i = \frac{\alpha_i\,(1-\mu_i)}{\mu_i}$$

où $\mu_i = {}^{MEAN_i}/_{EXPOSURE_i}$ et $\sigma_i = {}^{(SDI_i + SDC_i)}/_{EXPOSURE_i}$ sont la moyenne et l'écart-type de la distribution. Les paramètres de loi Bêta peuvent donc être calculés à l'aide de la méthode des moments (A.3.1).

Ces hypothèses de calcul permettent de déterminer les lois statistiques suivies par la survenance annuelle des évènements et le taux de destruction associé à chaque évènement. Ces dernières sont utilisées pour effectuer des simulations sur l'ELT-DABDOM afin d'appliquer aux pertes simulées les traités de réassurance intervenant pour le portefeuille DAB Domestique et ensuite, calculer la prime de réassurance du CAT XL.

De plus, afin de pouvoir appliquer correctement la structure de réassurance du portefeuille DAB Domestique, il convient de faire une hypothèse sur le périmètre d'application du traité DAB par risque. En effet, ce traité intervient avant le traité CAT XL et est à la maille sinistre (2.2.2). Comme nous ne disposons de l'ELT au niveau site que pour le produit RI (c'est-à-dire que les pertes sont détaillées par site par évènement), nous émettons l'hypothèse que la priorité du DAB par risque ne pourra être atteinte que par des sinistres sur les sites assurés constituant le portefeuille industriel, et non pas par des sinistres sur les sites assurés des six autres portefeuilles de la garantie DAB Domestique. Le tableau 4.5 présente les coûts maximaux de sinistres dus au péril tempête dans l'historique de Generali France.

Produit	Montant maximal de sinistre (en k€)
Municipal	520
Engineering	591
MRH	669
MRC	797
MRI	920
MRA	2 711
RI	3 673

TABLE 4.5 – Montant maximal des sinistres historiques tempête

Ainsi, le produit RI détient le sinistre dû au péril tempête le plus élevé du portefeuille DAB Domestique et il est donc plus probable qu'un site assuré industriel atteigne la priorité du DAB par risque, en vue de l'historique de Generali France. Il faut également remarquer que la priorité du traité DAB par risque est beaucoup plus élevée que le montant maximal d'un sinistre industriel dû à une tempête.

4.2.2 Tirage des évènements tempête

Cette étape consiste à choisir un nombre d'années de simulations sur lesquelles seront calculées les cessions au traité CAT XL et de simuler les évènements survenant chaque année.

Premièrement, le choix du nombre d'années de simulation porte sur la comparaison entre les valeurs de la courbe OEP du portefeuille DAB Domestique fournies par le logiciel RMS, et celles calculées sur l'YLT, donc après avoir effectué des simulations. Plus le nombre de simulations sera élevée, et plus l'erreur constatée devrait être faible. On regardera également certaines données disponibles directement via l'ELT-DABDOM et des données recalculées à partir de l'YLT. Au final, il a été décidé d'effectuer 50 000 années de simulation. La robustesse de ce paramètre est expliquée dans la section 4.4.2.

Deuxièmement, il convient de calculer le nombre d'évènements survenant chaque année. Pour cela, on tire aléatoirement, 50 000 fois, dans une loi de Poisson de paramètre $\lambda = \sum_{i=1}^{n} RATE_i = 47.8$. Le nombre d'évènements survenant pour les cinq premières années de simulation est présenté dans la table 4.6.

Année de simulation	Nombre d'évènements tempête
1	49
2	55
3	47
4	48
5	30

TABLE 4.6 – Nombre d'évènements tempête des cinq premières années de simulation tirés dans une loi de Poisson

En effectuant la moyenne du nombre d'évènements tempête sur l'ensemble des 50 000 années de simulation, on retrouve bien 47.8 soit l'espérance de notre loi de Poisson, λ.

Pour finir, comme le nombre d'évènements survenant chaque année est connu, il suffit à présent de tirer aléatoirement dans les « EVENTID » de l'ELT-DABDOM. Cette étape se divise en trois sous-étapes. Afin d'expliquer clairement la méthode, on se basera sur un exemple d'ELT au niveau portefeuille contenant deux évènements (tableau 4.7), pour laquelle on souhaite construire une YLT de deux ans. En réalité, on dispose d'une ELT contenant 19 912 évènements et on souhaite construire une YLT de 50 000 ans.

EVENT ID	RATE	MEAN	SDC	SDI	EXPOSURE
1	0.01	1 700	500	640	2 500
2	0.05	900	200	320	2 000

TABLE 4.7 – Exemple d'ELT au niveau portefeuille pour la garantie DAB Domestique

Dans l'exemple décrit par le tableau 4.7, $\lambda = 0.01 + 0.05 = 0.06$. En tirant aléatoirement deux fois (nombre d'années souhaitées dans l'YLT) dans une loi de Poisson de paramètre $\lambda = 0.06$, deux évènements surviennent la première année et un seul la deuxième.

Afin de connaître les évènements se déclarant pendant les deux années, il convient de calculer dans un premier temps la probabilité conditionnelle de chaque évènement, c'est-à-dire la probabilité que ce soit l'évènement n°i qui survienne sachant qu'un évènement survient pendant l'année de simulation. Cette probabilité correspond à $prob_i = {}^{RATE_i}/\lambda$ avec $\lambda = 0.06$. En effet, en notant :

— A l'évènement : « l'EVENTID n°i survient »,

— B l'évènement : « un évènement survient »,

la formule suivante est obtenue :

$$\mathrm{P}(A \mid B) = \frac{\mathrm{P}(A \cap B)}{\mathrm{P}(B)} = \frac{\mathrm{P}(A)}{\mathrm{P}(B)}$$

car $A \subset B$. Ensuite, les probabilités conditionnelles associées aux « EVENTID » sont cumulées ligne à ligne. Le résultat obtenu pour l'ELT exemple 4.7 est présenté dans le tableau 4.8.

EVENT ID	RATE	MEAN	Probabilité conditionelle	Probabilité cumulée
1	0.01	1 700	0.17	0.17
2	0.05	900	0.83	1.00

TABLE 4.8 – Probabilités conditionnelles de survenance des évènements de l'ELT 4.7

Par la suite, on tire autant de fois qu'il y a d'évènements sur l'ensemble des années de simulation dans une loi uniforme continue sur [0,1] (A.1.3). Pour l'exemple, trois nombres sont tirés aléatoirement : 0.28, 0.12 et 0.77. Ensuite, pour chaque nombre tiré, on sélectionne la première ligne de l'ELT dont la probabilité cumulée est supérieure à ce nombre. En considérant T évènements sur l'ensemble des années de simulation, l'approche peut être résumée par l'algorithme suivant :

Pour j allant de 1 à T :

1. Tirer aléatoirement dans une loi uniforme continue sur [0,1], notons u_j le résultat du tirage.
2. Sélectionner la ligne R de l'ELT telle que $R = \min\left\{r : u_j \leq \sum_{i=1}^{r} \frac{RATE_i}{\lambda}\right\}$.

Cette méthode consiste finalement à diviser l'intervalle [0,1] par le nombre d'évènements contenus dans l'ELT initiale de telle sorte que la part de l'intervalle associée à un évènement soit représentative de sa fréquence annuelle de survenance, soit du pourcentage de chance qu'il soit tiré. En utilisant l'ELT 4.8 servant d'exemple, les tirages aléatoires suivants sont effectués :

Loi uniforme	Ligne de l'ELT	EVENTID
0.28	2	2
0.12	1	1
0.77	2	2

TABLE 4.9 – Tirage aléatoire dans les évènements de l'ELT 4.8

La dernière étape consiste à attribuer à chaque année de l'YLT les évènements tirés à l'étape précédente, dans l'ordre et en fonction du nombre d'évènements survenant lors de chaque année de simulation. Ainsi, l'ELT exemple donne le résultat suivant :

Année	Nombre d'évènements	EVENTID
1	2	2 puis 1
2	1	2

TABLE 4.10 – Attribution des évènements tempête aux années de simulation pour l'exemple d'ELT 4.8

En repassant sur le cas réel, voici les deux premières années de l'YLT pour lesquelles plusieurs évènements tempête RMS ont été tirés dans l'ELT-DABDOM à l'aide de la méthode décrite ci-dessus :

Année	Nombre d'évènements	EVENTID
1	49	3198285, 3194241, 3178791, etc.
2	55	3200910, 3200112, 3200272, etc.

TABLE 4.11 – Evènements de l'ELT-DABDOM survenant durant les deux premières années de simulation

Après avoir tiré les évènements survenant chaque année de simulation, la finalisation de la construction de l'YLT consiste à déterminer les pertes associées à chaque évènement pour chaque année de simulation. Plusieurs approches peuvent être utilisées et deux d'entre elles seront présentées dans l'étude. L'approche la plus simple à mettre en place est présentée à la suite, dans la section 4.2.3.

4.2.3 Utilisation des pertes moyennes

La première approche consiste à attribuer à chaque évènement tiré aléatoirement la perte moyenne associée, contenue dans l'ELT-DABDOM. Cette méthode ne prend donc pas en compte l'incertitude secondaire et suppose que la perte moyenne correspond à la perte réelle. Il s'agit de l'approche la plus simple puisqu'aucune simulation n'est nécessaire sur les pertes. Les premières lignes de l'YLT associée au portefeuille DAB Domestique sont présentées dans le tableau 4.12.

Année	EVENTID	MEAN (Perte en €)
1	3198285	306k
1	3194241	429k
1	3178791	576k
1	3200165	167k
1	3200008	6k

TABLE 4.12 – Premières lignes de l'YLT au niveau portefeuille pour la garantie DAB Domestique sans incertitude secondaire

L'YLT 4.12 permet ainsi d'obtenir l'ensemble des pertes du portefeuille DAB Domestique par évènement pour les 50 000 années de simulation, sans prendre en compte l'incertitude secondaire.

4.2.4 Application de la structure de réassurance

Avant d'appliquer le traité CAT XL à ces pertes par évènement, un autre traité de réassurance intervient : le DAB par risque (2.2.2). Cependant, celui-ci ne s'applique pas sur les pertes au niveau évènement mais sinistre par sinistre. Or, l'YLT 4.12 ne comprend pas ce niveau de détail. Grâce à l'hypothèse émise dans la section 4.2.1, il suffit de vérifier que la contribution des sinistres du portefeuille RI aux différents évènements de l'YLT ne dépasse pas la priorité du DAB par risque. Comme les pertes de l'YLT 4.12 sont la somme des pertes moyennes de chaque ELT initiale des sept produits du portefeuille DAB Domestique, il suffit alors de vérifier qu'aucune perte moyenne de l'ELT du portefeuille RI n'est supérieure à la priorité du traité, c'est-à-dire supérieure à 12 M€.

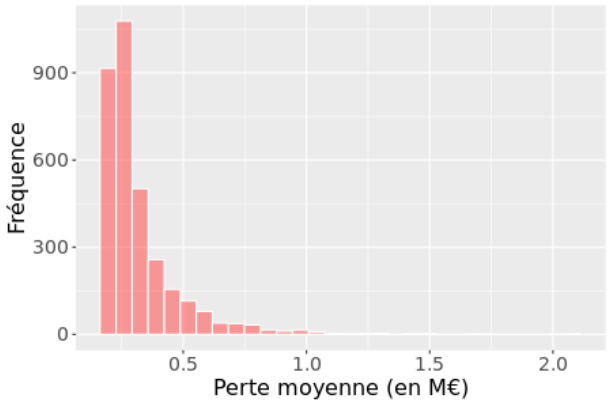


FIGURE 4.1 – Histogramme des pertes moyennes des sites RI modélisées par RMS supérieures à 200k€

D'après la figure 4.1, aucune valeur de la colonne « MEAN » de l'ELT-DABDOM n'est

supérieure à 12 M€. En effet, la perte moyenne maximale des sites industriels est égale environ à 2 M€. Ainsi, aucun sinistre n'est cédé au titre du traité DAB par risque et il est donc possible d'appliquer directement le traité CAT XL aux pertes de l'YLT 4.12. Une fois avoir calculé les cessions au traité CAT XL pour chaque évènement, il suffit de sommer ces cessions par année de simulation. Les cinq premières années de l'YLT activant le traité CAT XL sont les suivantes :

Année	Cessions au CAT XL (en M€)
1	185
6	104
11	7
15	11
23	29

TABLE 4.13 – Cinq cessions annuelles de l'YLT 4.12 au traité de réassurance CAT XL

La prime de réassurance du CAT XL modélisée s'obtient en effectuant la moyenne des cessions annuelles sur les 50 000 années de simulation et en ajoutant 20% de l'écart-type de ces cessions annuelles. Le taux de 20% correspond à un taux de chargement standard utilisé en interne chez Generali. Celui-ci ne correspond pas forcément au taux appliqué par le réassureur dans la prime de réassurance réellement payée par Generali. Ces derniers définissent leur taux dans le but de couvrir leurs frais et améliorer leur rentabilité. Sur l'ensemble de l'YLT, la moyenne des cessions annuelles est égale à 13.1 M€ ce qui donne une prime de réassurance modélisée égale à 26.1 M€.

4.3 Prise en compte de l'incertitude secondaire dans la *Year Loss Table*

Cette section permet de mettre en place une approche de simulation des pertes de l'YLT prenant en compte l'incertitude secondaire calculée par RMS. Ainsi, les pertes moyennes de l'ELT-DABDOM ne sont pas directement utilisées comme pertes réelles de l'YLT. Afin de pouvoir comparer les résultats des méthodes de construction d'YLT avec et sans incertitude secondaire, les évènements survenant durant les 50 000 années de simulation seront similaires pour les deux approches (4.2.2).

4.3.1 Simulation des pertes

La deuxième approche consiste à prendre en compte l'incertitude secondaire et donc à considérer toute l'étendue des pertes associées à chaque évènement. Ainsi, pour chaque année de simulation et chaque évènement, le taux de destruction réel est tiré aléatoirement dans une loi Bêta. Les paramètres de loi Bêta associés à l'évènement n°i, α_i et β_i, sont calculés via les informations associées à cet évènement dans l'ELT-DABDOM.

En effet, à chaque évènement tempête de l'ELT-DABDOM est associé une loi Bêta. En considérant T évènements survenant sur l'ensemble des années de simulation, la méthode utilisée afin de tirer aléatoirement un taux de destruction dans une loi Bêta est décrite ci-dessous :

Pour j allant de 1 à T :

1. Tirer aléatoirement dans une loi uniforme continue sur [0,1], notons u_j le résultat du tirage.
2. Calculer le quantile du taux de destruction $q_i^{u_j}$ tel que $u_j = \mathrm{P}\left(TD_i \leq q_i^{u_j}\right)$ où TD_i suit une loi Bêta et représente la distribution du taux de destruction pour l'évènement i de l'ELT-DABDOM.

Ainsi, pour chaque évènement tiré, le taux de destruction est calculé grâce à l'inverse de la fonction de répartition de la loi Bêta (A.1.7) de l'évènement associé. Une fois que les taux de destruction ont été déterminés pour chaque année et chaque évènement de l'YLT, la perte réelle associée s'obtient en multipliant ce taux de destruction par l'exposition du portefeuille DAB Domestique (colonne de l'ELT-DABDOM « EXPOSURE »). Les premières lignes de la deuxième YLT associée au portefeuille DAB Domestique sont présentées dans le tableau 4.14.

Année	EVENTID	α_i	β_i	u_j	Taux de destruction simulé	EXPOSURE	Perte simulée (€)
1	3198285	0.47	123k	0.78	5e-6	80Mrd	461k
1	3194241	0.59	365k	0.17	1e-7	268Mrd	31k
1	3178791	0.53	27k	0.17	1e-6	29Mrd	32k
1	3200165	0.38	381k	0.35	1e-7	166Mrd	22k
1	3200008	0.17	276k	0.93	4e-5	9Mrd	379k

TABLE 4.14 – Premières lignes de l'YLT au niveau portefeuille pour la garantie DAB Domestique avec incertitude secondaire

La colonne u_j correspond à la valeur tirée dans l'intervalle [0,1], décrite dans l'algorithme ci-dessus. En quelque sorte, u_j représente l'intensité des pertes pour chaque évènement de l'YLT. En effet, plus la valeur u_j associée à l'évènement i est élevée et plus le taux de destruction tiré est élevé.

4.3.2 Application de la structure de réassurance

Ici, les pertes ont été modélisées par évènement à l'aide de lois Bêta et la contribution du portefeuille RI n'est plus égale à sa perte moyenne pour l'évènement. Afin de vérifier que les sites industriels RI ne contribuent pas au traité DAB par risque en cas de survenance des évènements tempête de l'ELT-DABDOM, il s'agira de vérifier que des pertes

simulées, seulement sur les sites assurés du portefeuille RI, avec la même intensité que dans l'YLT 4.14 (correspondant à la colonne u_j), ne dépassent pas la priorité du traité de réassurance DAB par risque. Pour cela, dans l'ELT initiale du portefeuille RI, sont sélectionnés les évènements et sites pour lesquels l'exposition (colonne « EXPOSURE ») est supérieure à 12 M€. Cette sélection permet de réduire les calculs à effectuer : 17% des lignes de l'ELT RI sont conservées. Ensuite, les paramètres des lois Bêta, α_i^s et β_i^s, sont calculés pour chacun des sites (s) et évènements sélectionnés (i). Il est à noter que jusqu'à présent, les lois Bêta étaient calculées par évènement. Ici, elles sont calculées par site RI et par évènement. Pour chacune de ces lois Bêta va être calculée la probabilité à partir de laquelle la perte simulée est supérieure à 12 M€. Cela revient à calculer la fonction de répartition de chaque loi Bêta des sites du portefeuille RI et des évènements sélectionnés, dont le quantile est égal à 12 M€ divisé par l'exposition de chaque site aux évènements :

$$\mathrm{P}(X_i^s \leq 12M) = p_i^s \Leftrightarrow \mathrm{P}(X_i^s > 12M) = 1 - p_i^s = q_i^s$$

où $X_i^s = TD_i^s \cdot EXPOSURE_i^s$ avec TD_i^s le taux de destruction suivant une loi Bêta pour le site industriel s et l'évènement i. $EXPOSURE_i^s$ est la valeur d'exposition du site s à l'évènement i (correspondant à la colonne « EXPOSURE » de l'ELT à la maille site du portefeuille RI). Ainsi, on calcule q_i^s et donc p_i^s pour chaque site industriel et chaque évènement. Les premières lignes des résultats obtenus sont affichées dans le tableau 4.15.

EVENTID	LOCID	EXPOSURE	Taux de destruction pour 12M de perte	α_i^s	β_i^s	p_i^s	q_i^s
3184563	13371	13M	0.91	0.015	0.176	0.95	0.05
3175249	13371	13M	0.91	0.006	0.096	0.95	0.05
3172308	13371	13M	0.91	0.004	0.082	0.96	0.04
3163796	13371	13M	0.91	0.014	0.236	0.97	0.03
3168094	759	33M	0.36	0.109	2.313	0.97	0.03

TABLE 4.15 – Paramètres de lois Bêta des sites RI

Désormais, il s'agit de comparer la valeur u_j tirée pour la construction de l'YLT 4.14, pour les évènements touchant le portefeuille RI sélectionnés ci-dessus, aux probabilités p_i^s obtenues pour les sites assurés du produit RI. En effet, si $u_j \geq p_i^s$ pour un même évènement, alors la perte simulée pour le site industriel considéré sera supérieure à 12 M€ pour cet évènement. En notant x_j le quantile de la perte associée à la probabilité u_j, c'est-à-dire $\mathrm{P}(X_i^s \leq x_j) = u_j$, et si $p_i^s \leq u_j$, alors on a :

$$\mathrm{P}(X_i^s \leq 12M) = p_i^s \leq u_j = \mathrm{P}(X_i^s \leq x_j) \implies 12M \leq x_j$$

car la fonction de répartion de X_i^s est strictement croissante et continue sur l'intervalle $]0, EXPOSURE_i^s[$. Cela est dû au fait que la fonction de répartition d'une loi Bêta est

strictement croissante et continue sur]0,1[. Après comparaison, aucune valeur de fonction de répartition u_j associée aux taux de destruction simulés pour l'ensemble de l'YLT du portefeuille DAB Domestique n'est supérieure à celles calculées sur les sites assurés du portefeuille RI, p_i^s, pour lesquelles le sinistre serait supérieur à 12 M€. Par conséquent, aucune cession n'est réalisée au titre du traité DAB par risque. Si nous avions observé un site s pour lequel la perte causée par l'évènement i était supérieure à 12 M€, et pour une intensité u_j pour cet évènement dans l'YLT du tableau 4.14, on aurait pu retiré le montant suivant à la ligne de l'YLT associée : $\min\{q_i^s - 12M, portée\}$ tel que $q_i^s = P(X_i^s \leq u_j)$, correspondant au sinistre net du traité DAB par risque du site industriel. L'inconvénient de cette approche est que l'incertitude secondaire des évènements tempête n'est pas entièrement exploitée. En effet, il est supposé ici qu'un évènement tempête engendre le même niveau de perte au global (sur tout le portefeuille DAB Domestique) et pour chaque site RI puisque les probabilités associées aux taux de destruction de l'YLT 4.14, u_j, n'ont pas été resimulées pour l'étude sur le produit RI. Néanmoins, les probabilités liées aux taux d'endommagement doivent variées suivant les localisations et donc les sites considérés, puisque l'intensité d'une tempête peut évoluer suivant la zone parcourue.

Comme aucune perte n'est cédée au traité de réassurance DAB par risque, les conditions du CAT XL sont appliquées directement aux pertes présentées dans l'YLT 4.14, construite à l'aide de la deuxième méthode de simulation. Ainsi, pour chaque perte associée à un évènement, on calcule la cession au traité CAT XL au titre de cet évènement.

Année	Cessions au CAT XL (en M€)
1	177
6	186
12	9
15	20
23	13

TABLE 4.16 – Cinq cessions annuelles de l'YLT 4.14 au traité de réassurance CAT XL

Sur l'ensemble des 50 000 années de simulation, la moyenne des cessions annuelles est égale à 14M€ et ainsi, la prime de réassurance modélisée est égale à 27.8M€. Il est normal que cette prime soit supérieure à celle obtenue avec la première approche puisque la simulation de lois Bêta a permis d'obtenir des taux de destruction plus élevés que les taux de destruction moyens.

4.3.3 Principale limite de l'approche

Les ELT récupérées en sortie de RMS contiennent des pertes GR. Cela signifie qu'elles sont nettes des conditions d'assurance et de coassurance. La méthode de simulation avec incertitude secondaire consistait à calculer les paramètres de lois Bêta pour chaque évènement contenu dans l'ELT-DABDOM à l'aide de la perte moyenne, de l'écart-type et

de l'exposition, puis de simuler des taux de destruction pour chaque ligne de l'YLT du portefeuille DAB Domestique à l'aide des lois obtenues pour les évènements associés. Or, les grandeurs contenues dans l'ELT-DABDOM ne sont pas des pertes brutes puisqu'il a été retiré des franchises, des limites ainsi que des conditions de coassurance. Les lois Bêta sont donc en réalité tronquées. La méthode exacte nécessiterait de prendre en compte les grandeurs associées à la troncature des lois telles que la franchise et la LCI.

Il existe deux solutions permettant de tenir compte de la troncature des lois Bêta associées à chaque évènement de l'ELT-DABDOM :

— La première méthode consisterait à récupérer les sept ELT *Ground-Up Loss* (perte brute d'assurance, de coassurance et de réassurance) au niveau site du portefeuille DAB Domestique. Ensuite, il s'agirait de calculer les paramètres de lois Bêta non plus pour chaque évènement mais pour chaque site et chaque évènement. Ainsi, les pertes seraient également simulées par site et par évènement. La dernière étape consisterait à appliquer les conditions d'assurance et de coassurance à chaque perte simulée. Cependant, en vue du nombre de simulations, du nombre d'évènements survenant chaque année de simulation et de sites appartenant au portefeuille DAB Domestique, la simulation des pertes au niveau site demanderait un temps et une complexité de calcul trop importants.

— La deuxième méthode consisterait à récupérer les ELT en sortie de RMS également au niveau site mais GR. Cependant, la simulation des taux de destruction devrait s'effectuer à l'aide de lois Bêta tronquées, par site et par évènement. Il faudrait donc une nouvelle fois avoir l'information des conditions d'assurance et de coassurance de chaque site couvert par la garantie DAB Domestique. Comme la simulation des pertes s'effectue également au niveau site et par évènement, la complexité de calcul serait similaire à la première solution.

Ainsi, la méthode de simulation présentée pour la construction de l'YLT au niveau portefeuille de la garantie DAB Domestique avec prise en compte de l'incertitude secondaire n'est pas totalement exacte mais représente une alternative afin de pallier à la complexité de calcul des méthodes décrites ci-dessus. En effet, celles-ci ne sont pas réalisables opérationnellement. De plus, il serait difficile de stocker la nouvelle YLT au niveau site, au moins pour le produit RI, afin de calculer les coûts individuels de réassurance pour plus de 10 000 sites industriels.

4.4 Validation de la méthode de simulation

Dans cette section, il s'agira de comparer les résultats obtenus à l'aide des simulations (et donc à partir de l'YLT) et ceux affichés et calculés par le logiciel RMS sur les ELT de façon analytique, c'est-à-dire sans effectuer de simulations. Cette comparaison pourra confirmer que le nombre de simulations choisi permet d'obtenir des résultats semblables à ceux disponibles sous RMS et que la méthode de simulation utilisée est suffisamment robuste. Pour finir, on comparera également la prime de réassurance obtenue à l'aide de

la méthode de simulation retenue et celle actuellement utilisée par le réassureur pour le renouvellement 2023. L'erreur relative absolue sera utilisée et se définit par la formule suivante :

$$\frac{|Valeur\ approchée - Valeur\ théorique|}{|Valeur\ théorique|}$$

4.4.1 Choix de la méthode de simulation

Afin de déterminer la méthode de simulation à utiliser pour que l'erreur entre les valeurs proposées par RMS sur la base de l'ELT et celles calculées sur les YLT soit minimale, quatre métriques seront comparées : la perte annuelle moyenne, l'écart-type des pertes annuelles σ, l'OEP 50Y et l'OEP 250Y. L'objectif de la méthode de simulation étant de calculer une prime de réassurance et non pas une prime d'assurance, il est important de vérifier que la méthode de simulation choisie permet de reproduire correctement la queue de distribution des pertes modélisées par RMS, d'où l'utilisation des valeurs OEP en plus de la perte annuelle moyenne. A présent, il convient alors définir les quatre grandeurs, qui sont disponibles directement sur RMS après modélisation de l'ELT.

Premièrement, la perte annuelle moyenne se calcule simplement sur l'ELT et l'YLT :

1. Sur l'ELT, la perte annuelle moyenne est égale à : $\sum_{i=1}^{n} RATE_i \cdot MEAN_i$ où n est le nombre d'évènements tempête touchant le portefeuille DAB Domestique dans le catalogue proposé par RMS (19 912 évènements tempête).
2. Sur l'YLT, la perte annuelle moyenne est égale à : $\sum_{i=1}^{50000} PA_i / 50000$ où PA_i correspond à la somme des pertes sur l'année i.

Deuxièmement σ correspond à l'écart-type des pertes annuelles sur les 50 000 années de simulation :

1. Sur l'ELT, $\sigma = \sqrt{\sum_{i=1}^{n} RATE_i \cdot ((SDC_i + SDI_i)^2 + MEAN_i^2)}$.
2. Quant à l'YLT, $\sigma = \sqrt{\sum_{i=1}^{50000} (PA_i - m)^2 / (50000 - 1)}$ où m représente la perte annuelle moyenne.

Pour finir, l'OEP 50Y (respectivement l'OEP 250Y) correspond à la valeur de la courbe OEP pour une période de retour de 50 ans (respectivement 250 ans) soit une probabilité de dépassement égale à 2% (respectivement 0.4%) :

1. Sur l'ELT, cette valeur se calcule grâce à la connaissance de la distribution du taux de destruction, suivant une loi Bêta, de la distribution de la survenance annuelle de chaque évènement ainsi que de la survenance annuelle globale, suivant toutes deux une loi de Poisson.
2. Sur l'YLT, la valeur d'OEP souhaitée pour un certain niveau de pertes se définit comme étant :

$$OEP(l) = \mathrm{P}(ML > l) = \frac{\#\{i \in [\![1\ ;\ 50000]\!] : ML_i > l\}}{50000}$$

où l correspond au montant de pertes souhaité, ML_i représente la perte maximale de l'année i et ML est la distribution des maximas annuels des pertes sur les années de simulation. Ainsi, dans le cas où la période de retour est connue et le montant de pertes est recherché, $OEP(RP)$ correspond cette fois-ci à la fonction quantile de la distribution des maximas annuels de pertes dont la probabilité en argument est égale à $1/RP$ (ici 2% et 0.4%).

Il est à noter que l'ensemble de ces grandeurs sont calculées avant application des conditions de réassurance du traité CAT XL. Les valeurs disponibles sous RMS pour le portefeuille DAB Domestique sont :

— $OEP(50Y) \approx 250$ M€,
— $OEP(250Y) \approx 510$ M€,
— la perte annuelle moyenne est égale environ à 54 M€,
— l'écart-type des pertes annuelles est égal environ à 85 M€.

L'erreur relative absolue sur les quatre paramètres cités précédemment est calculée pour les deux YLT construites à l'étape précédente, en fonction du nombre de simulations effectuées. Nous souhaitons obtenir l'erreur relative la plus basse possible et choisissons un seuil de 1%.

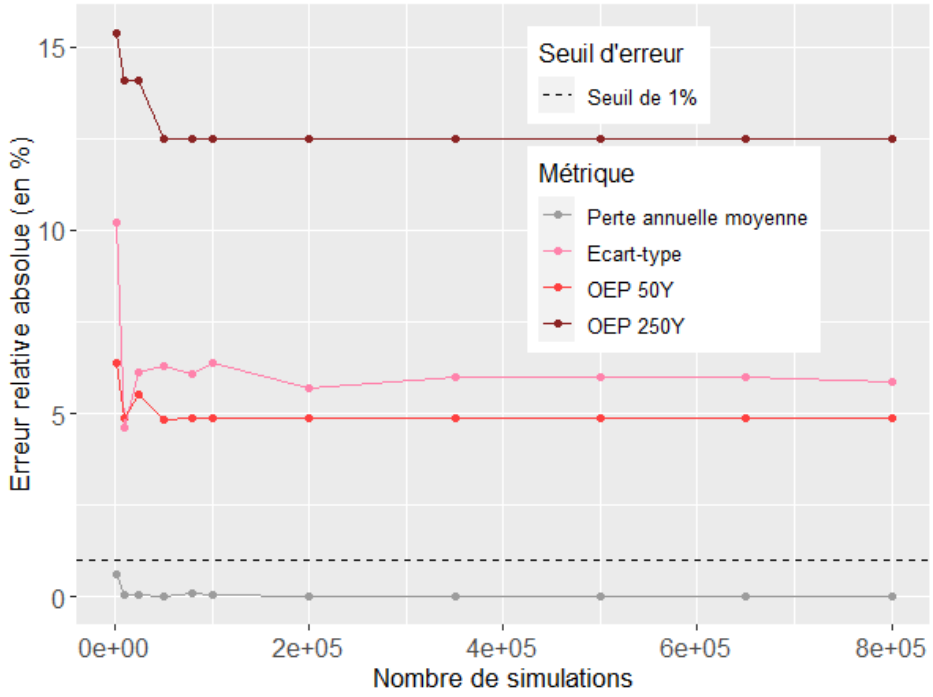


FIGURE 4.2 – Erreur relative absolue en fonction du nombre de simulations pour la méthode de simulation des pertes sans incertitude secondaire

Le graphique 4.2 décrit l'erreur sur la méthode de simulation ne prenant pas en compte l'incertitude secondaire et donc se basant seulement sur les pertes moyennes associées aux différents évènements. L'erreur sur la moyenne des pertes annuelles est toujours inférieure à 1% peu importe le nombre de simulations. Ainsi, utiliser les pertes moyennes pour la construction de l'YLT permet de reproduire convenablement la moyenne de la distribution des pertes de l'ELT. Cependant, on peut voir que l'erreur sur l'écart-type ne descend pas en-dessous de 5% : cela est dû au fait que les étendues de pertes ne sont pas prises en compte par la méthode, ainsi l'écart-type initial des pertes annuelles ne peut être correctement reproduit. Quant aux valeurs OEP, augmenter le nombre de simulations ne permet pas non plus d'obtenir une erreur inférieure à 1%. Pour conclure, la première méthode de simulation ne reproduit pas correctement les pertes au-delà de la valeur moyenne, notamment les valeurs extrêmes de la distribution.

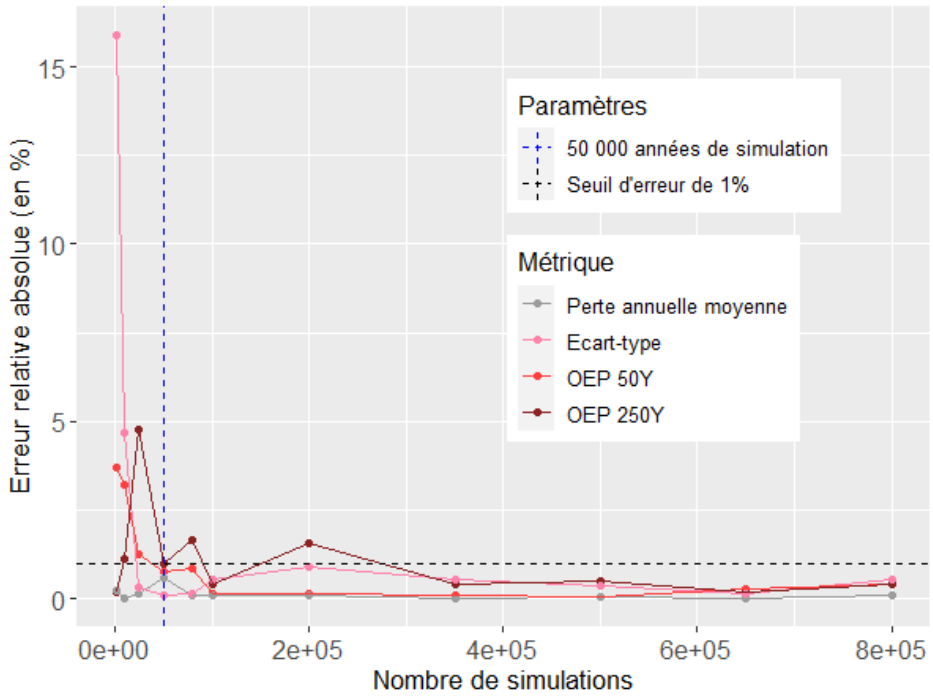


FIGURE 4.3 – Erreur relative absolue en fonction du nombre de simulations pour la méthode de simulation des pertes avec incertitude secondaire

Sur la figure 4.3, à partir de 50 000 simulations, les erreurs sur l'écart-type des pertes annuelles et l'OEP 50Y sont toujours inférieures à 1%. Cela est dû à la considération de la perte par évènement comme une distribution et non plus comme la valeur moyenne. L'incertitude secondaire permet de prendre en compte toute l'étendue des pertes et ainsi simuler des valeurs plus élevées que la valeur moyenne. Il en est de même pour l'OEP 250Y à partir de 350 000 simulations. Le nombre de simulations minimal à partir duquel

l'erreur est inférieure à 1% est plus important pour la période de retour 250 ans que pour 50 ans puisque l'OEP 250Y correspond à un évènement tempête plus intense et nécessite ainsi un grand nombre de simulations avant de pouvoir être atteint. Quant à l'erreur sur la moyenne des pertes annuelles, celle-ci est toujours inférieure à 1%. Ainsi, la méthode de construction de l'YLT utilisée pour la suite de l'étude est celle prenant en compte l'incertitude secondaire, c'est-à-dire pour laquelle les taux de destruction par évènement sont modélisés par des lois Bêta. La figure 4.3 permet également de démontrer la robustesse de cette méthode malgré que celle-ci ne tienne pas compte de la troncature des lois Bêta (section 4.3.3). En effet, l'erreur entre les valeurs calculées sur l'ELT et l'YLT est très faible alors que la méthode analytique utilisée par RMS afin de calculer la perte annuelle moyenne, l'écart-type des pertes annuelles ainsi que la courbe OEP tient compte de la troncature des pertes contenues dans les ELT GR. Par conséquent, l'erreur sur la deuxième méthode de simulation étant rapidement inférieure à 1%, l'approche est validée pour cette étude.

4.4.2 Choix du nombre d'années de simulation

En vue du graphique précédent, on aurait tendance à choisir un nombre de simulations supérieur ou égal à 350 000 afin que l'erreur sur l'OEP 250Y soit toujours inférieure à 1%. Cependant, il est important de choisir le nombre minimal de simulations permettant d'avoir des bons résultats afin de diminuer le temps de calcul. En effet, on observe également sur le graphique 4.3 qu'il est possible d'obtenir une erreur relative absolue sur l'OEP 250Y inférieure à 1% avec 50 000 simulations. Afin de confirmer ces propos, l'erreur sur la valeur de l'OEP 250Y a été calculée sur 1 000 simulations différentes et pour plusieurs nombres d'années de simulation. Ainsi, on retiendra le nombre de fois où l'erreur est inférieure à 1% et 2% dans le tableau 4.17.

Nombre de simulations	Pourcentage d'erreur $\leq 1\%$	Pourcentage d'erreur $\leq 2\%$
50 000	41	72
100 000	60	89
200 000	76	98
300 000	86	100
500 000	93	100
600 000	96	100

TABLE 4.17 – Pourcentage d'erreur sur l'OEP 250Y inférieur à 1% et 2% pour un nombre de simulations différent

Sur 1 000 tests effectués, dans 41% des cas, l'erreur sur l'OEP 250Y est inférieure à 1%. On a ainsi fait le choix d'utiliser 50 000 années de simulation dans l'YLT afin de diminuer le temps de calcul de la prime de réassurance mais également, le temps de calcul des coûts de réassurance individuels des sites industriels. Il suffit de choisir, pour la méthode de simulation avec incertitude secondaire, une graine aléatoire permettant d'obtenir une

valeur d'OEP 250Y pour laquelle l'erreur avec le résultat directement fourni par RMS est inférieure à 1%.

4.4.3 Comparaison à la prime de marché du traité CAT XL

La prime de marché du traité CAT XL correspond au montant que paie Generali France au réassureur afin de lui céder ses sinistres couverts par la garantie DAB Domestique au titre de ce traité de réassurance. Dans cette prime sont compris plusieurs périls : la tempête, l'inondation, la grêle, la tornade, le gel et les ouragans. Or, l'étude réalisée jusqu'à présent ne concerne que les pertes liées aux évènements tempête. Afin de connaître l'écart entre la prime obtenue grâce à la méthode de simulation des pertes avec incertitude secondaire et la prime de réassurance de marché, il est nécessaire de calculer la contribution du péril tempête à cette dernière. Les cessions historiques au CAT XL ne touchent que la première tranche du traité et 74% de ces cessions sont dues à des évènements tempête. Comme précisé dans la section 2.2.2, on émet l'hypothèse que les tranches plus hautes du traité ne peuvent être touchées que par le péril tempête. La prime de réassurance de marché étant détaillée par tranche, on obtient une prime de marché du CAT XL pour le péril tempête représentant 91% de la prime globale du traité. A partir de l'YLT du tableau 4.14, la prime de réassurance du traité CAT XL est également calculée par tranche. L'écart relatif de la prime obtenue à partir de l'YLT 4.14 par rapport à la prime de marché est décrit dans la table 4.18.

Tranche	Contribution du péril tempête	RMS + Simulations	Ecart relatif
T1	74%	8M€	7%
T2	100%	9M€	13%
T3	100%	5M€	2%
T4	100%	6M€	5%
Total	91%	28M€	1%

TABLE 4.18 – Ecart entre la prime de réassurance de marché et la prime issue d'un logiciel de modélisation CAT

Ainsi, la prime de réassurance de marché et celle obtenue à l'aide de la méthode de simulation et du logiciel RMS ont un écart inférieur à 1%. Cependant, il est à noter que les données en entrée du logiciel RMS comprennent l'inflation jusque juin 2022 pour les sommes assurées alors que dans sa prime de réassurance, le réassureur prend en compte en plus l'inflation entre juin et décembre 2022, et projette également l'inflation pour l'année 2023. L'écart relatif par rapport à la prime de marché étant de +1%, le logiciel de modélisation CAT produit une prime de réassurance pour le traité CAT XL plus élevée que celle du réassureur. Cela peut s'interpréter par le fait que RMS dispose d'un catalogue d'évènements stochastiques comprenant des évènements extrêmes peu observés dans l'historique de Generali France. Par conséquent, si le réassureur ne se base que sur l'historique des sinistres de Generali, les cessions historiques ou celles simulées en

se basant sur l'historique auront tendance à être plus faibles que les cessions issues des modélisations réalisées par des logiciels tels que RMS.

Pour conclure, la prime de réassurance du CAT XL calculée en tenant compte de l'incertitude secondaire modélisée par RMS, permet de capter les valeurs extrêmes de la distribution des pertes dues aux évènements tempête. Par conséquent, les simulations de pertes réalisées sur l'ELT-DABDOM grâce à cette approche permettent d'obtenir des résultats semblables à ceux fournis par le logiciel RMS, ce qui montre que la méthode de simulation prenant en compte l'incertitude secondaire est suffisamment robuste. Pour rappel, l'objectif de l'étude est la mise à disposition d'un indicateur sur le coût de réassurance d'un nouveau site industriel pour le souscripteur RI. De ce fait, il est important que les cessions au traité CAT XL soient correctement simulées afin que ce coût de réassurance soit fiable pour pouvoir calculer les contributions marginales des sites du portefeuille RI en se basant sur ce dernier.

Chapitre 5

Calcul de la contribution marginale des sites industriels du portefeuille existant de Generali à la prime du traité CAT XL

Le premier objectif de l'étude correspond à la mise en place du calcul de la répartition de la prime de réassurance du CAT XL, pour le péril tempête, aux sites industriels du portefeuille existant de Generali France. Cela permet, par la suite, le développement d'un proxy de calcul pour des potentielles affaires nouvelles du portefeuille RI. Ainsi, il est nécessaire que l'indicateur sur le coût de réassurance au niveau site soit adapté à son utilisation à l'étape de souscription. Dans cette section sera présentée le calcul du coût marginal des sites assurés RI par Generali à partir de la prime de réassurance du CAT XL. Il est à noter que seule une partie de cette prime sera allouée aux sites assurés composant le produit RI puisqu'en réalité, le traité CAT XL couvre l'ensemble des produits protégés par la garantie DAB Domestique.

5.1 Description théorique du coût marginal

Le coût marginal est une grandeur provenant de la théorie des jeux. Tout d'abord, il est nécessaire de définir un groupe de joueurs, c'est-à-dire l'ensemble des participants pour lesquels une fonction de coût est calculée. Ici, les sites assurés de l'ensemble du portefeuille DAB Domestique existant correspondent aux joueurs. L'ensemble des joueurs est noté N. En considérant qu'il y a n sites concernés par la garantie DAB Domestique de Generali France, $N = [\![1, n]\!]$. On appelle coalition tout sous-ensemble de N, souvent notée S. Ici, le coût marginal ne sera calculé que pour les joueurs correspondant aux sites industriels, c'est-à-dire pour plus de 10 000 joueurs. On note n_i le nombre de sites industriels de Generali France. Le coût marginal est défini par la formule suivante :

$$\forall i \in [\![1, n_i]\!],\ CM(i) = c(N) - c(N \backslash i) \tag{5.1}$$

avec :

— $CM(i)$ le coût marginal du joueur n°i,
— $c(N)$ la fonction de coût de l'ensemble des joueurs,
— $c(N\backslash i)$ la fonction de coût de la coalition $S = N\backslash i$.

La fonction de coût c peut être calculée pour n'importe quelle coalition de l'ensemble N et représente le coût associé à un ou plusieurs joueurs. Ainsi, le coût marginal du joueur n°i peut être défini comme le coût engendré par l'entrée du joueur i dans la coalition $S = N\backslash i$. Ce dernier est donc adapté à l'objectif d'utilisation à la souscription d'un indicateur sur le coût de réassurance, calculé lors de l'acquisition d'un nouveau site industriel.

5.2 Calcul des coûts marginaux des sites industriels

Afin de calculer la contribution marginale des sites RI du portefeuille existant au coût du traité CAT XL, la fonction de coût c définie dans la formule (5.1) est simplement la prime de réassurance du traité CAT XL pour le péril tempête, calculée dans le chapitre précédent. Ainsi, le coût marginal de réassurance peut se réécrire comme :

$$\forall i \in [\![1, n_i]\!],\ CM(i) = Prime(N) - Prime(N\backslash i)$$

avec $CM(i)$ le coût marginal du site n°i, $Prime$ correspondant à la prime de réassurance du traité CAT XL modélisée pour le péril tempête, N représente l'ensemble des sites du portefeuille DAB Domestique et $N\backslash i$ représente l'ensemble des sites du portefeuille DAB Domestique privé du site industriel i. Finalement, le coût marginal du site n°i peut se définir comme la prime de réassurance à payer en plus lorsque le souscripteur fait rentrer le site n°i dans le portefeuille RI de l'assureur.

Le coût marginal est calculé pour l'ensemble des sites du portefeuille RI seulement. Grâce à la méthode de calcul de la prime de réassurance avec incertitude secondaire présentée dans la section 4.3, la valeur $Prime(N)$ est connue et est égale à 27.8 M€. Ainsi, seule la valeur $Prime(N\backslash i)$ est à calculer. La démarche suivie est présentée ci-dessous :

1. Le site RI n°i est retiré de l'ELT associée au produit RI. Cette étape est possible car le niveau de détail de l'ELT de ce produit est à la maille site.
2. L'ELT RI au niveau site sans le site n°i est agrégée à la maille portefeuille grâce à la méthode présentée dans la partie 3.3.
3. Ensuite, les ELT associées aux sept produits différents de la garantie DAB Domestique sont agrégées. On obtient donc une nouvelle ELT-DABDOM.
4. La prochaine étape consiste à construire une nouvelle fois l'YLT, mais cette fois-ci à partir de l'ELT du portefeuille DAB Domestique ne contenant pas les pertes liées au site RI n°i :
   (a) Afin de comparer $Prime(N)$ et $Prime(N\backslash i)$, les évènements survenant chaque année de simulation ainsi que l'intensité de ces évènements sont les mêmes

que ceux utilisés pour la construction de l'YLT complète tenant compte de l'incertitude secondaire (table 4.14). Ainsi, pour construire la nouvelle YLT, les années de simulation, les « EVENTID » ainsi que l'intensité des pertes notée u_j restent inchangés par rapport au tableau 4.14.

(b) Les paramètres de lois Bêta associés à chaque évènement de la nouvelle ELT-DABDOM doivent être recalculés puisque les lignes associées au site RI n°i ont été supprimées de l'ELT RI initiale. Les taux de destruction associés à chaque évènement de l'YLT sont ensuite resimulés à l'aide de la fonction quantile appliquée à u_j des nouvelles lois Bêta.

5. Une fois l'YLT privée du site i obtenue, les cessions au traité CAT XL sont calculées pour l'ensemble des évènements. $Prime(N\backslash i)$ correspond à la moyenne de ces cessions sommées par année de simulation, à laquelle est ajouté 20% de l'écart-type de ces cessions annuelles.

Finalement, la méthode utilisée pour calculer $Prime(N\backslash i)$ est parfaitement similaire à la méthode de calcul de $Prime(N)$ sauf que les pertes du catalogue d'évènements tempête associées au site industriel n°i ont été supprimées de l'ELT-DABDOM. Les premières lignes de la nouvelle YLT pour laquelle a été retiré le site RI n°1 sont affichées dans le tableau 5.1.

EVENTID	α	β	Fonction de répartition	Taux de destruction simulé	EXPOSURE	Perte simulée
3198285	0.42	116k	0.78	4e-6	80Mrd	460k
3194241	0.48	322k	0.17	9e-8	268Mrd	29k
3178791	0.53	27k	0.17	1e-6	29Mrd	32k
3200165	0.35	367k	0.35	8e-8	166Mrd	21k
3200008	0.17	276k	0.93	4e-5	9Mrd	379k

TABLE 5.1 – Premières lignes de l'YLT au niveau portefeuille pour la garantie DAB Domestique privé du site RI n°1

En vue de la nouvelle YLT du tableau 5.1, le site industriel n°1 n'est pas impacté par les évènements 3178791 et 3200008 car les paramètres de lois Bêta restent inchangés par rapport à l'YLT 4.14, construite avec le portefeuille DAB Domestique entier. Comme l'intensité du taux de destruction reste la même, la perte simulée reste également inchangée.

5.3 Présentation des résultats

Les coûts marginaux de réassurance CAT XL (péril tempête) sont calculés pour l'ensemble des sites RI. Leur distribution peut être observée sur les figures 5.1 et 5.2.

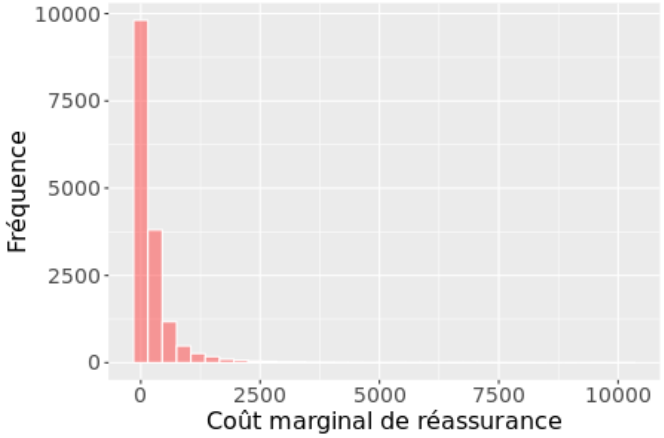


FIGURE 5.1 – Histogramme des coûts marginaux de réassurance des sites RI

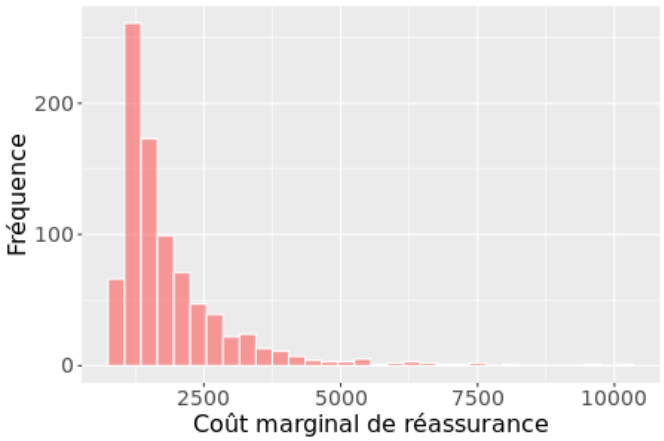


FIGURE 5.2 – Histogramme des coûts marginaux de réassurance des sites RI supérieurs à 1 000 €

Ainsi, on peut remarquer que certains sites RI disposent de coûts marginaux de réassurance élevés, allant jusque 10 k€, mais que la majorité des sites ont des coûts marginaux inférieurs à 200 €.

Afin d'identifier les zones contribuant fortement au coût de réassurance du traité CAT XL du portefeuille RI et pour le péril tempête, les coûts marginaux de réassurance des sites industriels ont été sommés par département et rapportés à l'engagement assuré RI du département sur la figure 5.3.

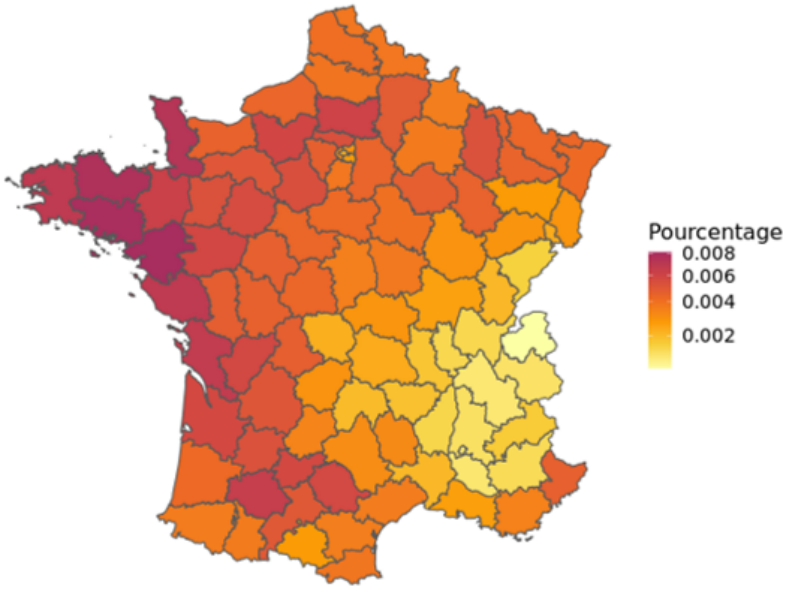


FIGURE 5.3 – Coûts marginaux de réassurance rapportés à l'engagement assuré par département pour le portefeuille RI

D'après le zonier vent contruit chez Generali à l'aide de RMS, l'aléa tempête est plus élevé à l'ouest de la France. Ainsi, les coûts marginaux de réassurance du traité CAT XL représentent bien cet aléa. En effet, les départements 22, 56 et 44 font partie des zones coûtant le plus cher en réassurance pour les évènements tempête RMS à Generali, pour le portefeuille RI. Cependant, la sinistralité historique due au péril tempête de Generali France n'est pas particulièrement élevée au nord-ouest de la France. Par conséquent, malgré que les évènements stochastiques construits par RMS se basent sur les évènements historiques, ce logiciel identifie des zones fortement exposées à l'aléa tempête qui n'ont pas connu beaucoup de sinistres historiques dans le passé. Cela peut être dû au fait que la région Bretagne dispose de méthodes de construction résistantes au vent.

5.4 Limites de l'approche

L'objectif principal de cette étude est de mettre à disposition de la souscription RI un indicateur sur le coût de réassurance engendré par l'acquisition d'une nouvelle affaire dans le portefeuille. Le coût marginal est donc pertinent puisqu'il évalue le prix de réassurance à payer en plus par Generali France pour le traité CAT XL lors de l'acquisition d'un nouveau site industriel. Cependant, le coût marginal est un indicateur moins adapté sur les sites du portefeuille RI existant, celui-ci présente des inconvénients :

— Le coût marginal de réassurance a été calculé sur les sites industriels du portefeuille existant comme si chaque site était rentré en dernier dans le portefeuille, et ne

tient pas compte de l'ordre réel d'entrée des sites au sein de Generali. Ainsi, une solution aurait été de calculer les coûts marginaux en fonction de l'ordre réel d'entrée dans le portefeuille RI ou d'utiliser une méthode d'allocation comme la valeur de Shapley. Celle-ci donne un poids égal à chaque combinaison possible d'ordres d'entrée du site considéré dans le portefeuille assuré. La valeur de Shapley du site n°i est notée x_i et se calcule par la formule suivante :

$$x_i = \frac{1}{n!} \sum_{\substack{S \subset N \\ i \in S}} (n-s)!(s-1)![c(S) - c(S \backslash i)]$$

n représente le nombre total de joueurs et s est la taille de la coalition S. Cette dernière demande ainsi un temps de calcul plus long que la méthode des coûts marginaux, en particulier lorsque le nombre de joueurs n est élevé.

— Le calcul des coûts marginaux sur le portefeuille existant ne constitue pas une méthode d'allocation de la partie RI de la prime du traité CAT XL aux sites industriels. En effet, en raison des dépendances des sites touchés par les mêmes évènements tempête, la somme des coûts marginaux des sites industriels n'est pas exactement égale à la contribution du produit RI à la prime du CAT XL, pour le péril tempête. Toutefois, cette différence est très faible : l'écart relatif entre ces deux valeurs est de 3%. Ainsi la méthode de calcul des coûts marginaux des sites est très proche d'une méthode de répartition de la prime RI du CAT XL pour le péril tempête.

Malgré les inconvénients de l'utilisation du coût marginal afin de représenter la contribution individuelle des sites du portefeuille RI existant, celle-ci a été retenue dans le but de mettre en place opérationnellement l'utilisation de l'indicateur de réassurance au niveau site pour la souscription et ainsi, avoir des informations sur le coût de réassurance supplémentaire lors de l'acquisition d'un nouveau site industriel.

Pour conclure, la méthode de calcul des coûts marginaux de réassurance des sites du portefeuille RI existant mise en place dans ce chapitre permet la constitution d'une base d'apprentissage fiable et robuste afin d'être en mesure de prédire le coût marginal de réassurance d'un nouveau site industriel à l'aide de ses caractéristiques.

Chapitre 6

Prédiction du coût marginal de réassurance d'un site industriel

L'objectif de ce chapitre est de calibrer plusieurs algorithmes de prédiction afin d'estimer les coûts marginaux de réassurance des sites RI, à l'aide de leurs caractéristiques. Il s'agira également de choisir un algorithme donnant des résultats suffisamment corrects afin qu'un souscripteur RI puisse avoir une information fiable quant au coût de réassurance supplémentaire engendré par l'acquisition d'un nouveau site.

6.1 Présentation de la base de données

Afin de prédire le coût marginal de réassurance RI lors de l'étape de souscription, une base de données a été construite à l'aide des caractéristiques renseignées par le souscripteur RI afin que l'approche puisse être mise en place simplement.

Nom de variable	Utilisation	Type de variable	Taux de remplissage (en %)
Coût marginal de réassurance	VAE	Numérique	100
Type d'occupation	Explicative	Nominale	64
Libellé de réseau	Explicative	Nominale	100
Activité du site	Explicative	Nominale	98
CRESTA	Explicative	Nominale	100
Zonier vent MRC	Explicative	Nominale	99
Surface (en m^2)	Explicative	Ordinale	100
Valeur au m^2	Explicative	Ordinale	100
Somme assurée bâtiment	Explicative	Ordinale	100
Somme assurée contenu	Explicative	Ordinale	100
Somme assurée PE	Explicative	Ordinale	100
Somme assurée BM	Explicative	Ordinale	100
Franchise du site	Explicative	Ordinale	100
Engagement assuré du site pour la garantie tempête	Explicative	Ordinale	100
Age du contrat (en années)	Explicative	Ordinale	100

TABLE 6.1 – *Dataset* pour la prédiction des coûts marginaux de réassurance des sites RI

Il est nécessaire d'effectuer quelques commentaires sur la table présentée :

— La garantie tempête est une garantie obligatoire comprise dans la couverture DAB des sites RI afin d'être protégé contre les sinistres liés aux tempêtes, puisque ceux-ci ne sont pas cessibles à la CCR au titre du régime CatNat.

— L'engagement assuré correspond à la somme assurée nette de LCI et de coassurance, pour la garantie tempête.

— Le type d'occupation correspond au libellé « Locataire » ou « Propriétaire ». Un locataire aura toujours une somme assurée bâtiment nulle et un propriétaire pourra avoir une somme assurée contenu nulle (notamment si le bâtiment est loué et non meublé).

— Il n'existe actuellement aucun zonier vent propre au produit RI au sein de Generali. Ainsi, un zonier vent construit spécifiquement pour le produit MRC sera utilisé comme variable explicative. Il donne une indication sur l'intensité de l'aléa tempête en France métropolitaine au niveau département.

— Les variables telles que les sommes assurées, l'engagement, la surface, la valeur au mètre carré, etc. étaient dans un premier temps de type numérique continue. Elles ont été retraitées en variables ordinales afin de simplifier l'interprétation des modèles. Une variable ordinale est une variable catégorielle pour laquelle un ordre a été choisi, contrairement à une variable nominale pour laquelle aucun ordre d'intensité n'est spécifié.

— La matrice de corrélation entre les variables explicatives et la VAE est présentée en annexe sur la figure 1. Certaines variables, comme le zonier vent MRC et le

CRESTA, sont très corrélées en raison de leur construction. Ainsi, une sélection de variables pourra être mise en place avant l'apprentissage des modèles de prédiction.

Pour l'ensemble des algorithmes de prédiction testés par la suite, les sites du portefeuille RI seront séparés en une base d'apprentissage (70% des sites) et une base de test (30% des sites).

6.2 Modèles et concepts théoriques utilisés

6.2.1 Modèle linéaire gaussien

La régression linéaire est un modèle de prédiction basé sur une relation linéaire entre les variables explicatives et la VAE. Par exemple, un modèle linéaire à p variables explicatives et appliqué à n observations est défini par la formule suivante : $Y = X\beta + W$, avec $Y \in \mathbb{R}^n$ le vecteur contenant la VAE, $X \in \mathbb{R}^{n\times p}$ la matrice déterministe des variables explicatives, $\beta \in \mathbb{R}^p$ le vecteur des coefficients linéaires et $W = (W^{(i)})_{i\in[\![1,n]\!]} \in \mathbb{R}^n$ où les $W^{(i)}$ sont des variables aléatoires i.i.d d'espérance nulle et de variance σ^2. Dans le cas du modèle linéaire gaussien, les $W^{(i)} \sim \mathcal{N}(0,\, \sigma^2)$. Ainsi, $\forall i \in [\![1, n]\!]$:

$$Y^{(i)} = \beta_0 + \beta_1 x_1^{(i)} + ... + \beta_p x_p^{(i)} + W^{(i)}$$

Afin d'expliquer les valeurs de Y, il est nécessaire d'estimer le vecteur $\beta = (\beta_j)_{j\in[\![0,p]\!]}$, en minimisant l'écart quadratique ci-dessous. Cette approche est appelée méthode des moindres carrés :

$$\sum_{i=1}^{n} (Y^{(i)} - (\beta_0 + \beta_1 x_1^{(i)} + ... + \beta_p x_p^{(i)}))^2$$

Dans le cas de la régression linéaire multiple, c'est-à-dire avec $p > 1$, la solution de ce problème est exacte et s'écrit sous forme matricielle : $\hat{\beta} = (X^\intercal X)^{-1} X^\intercal Y$, avec $X^\intercal$ la matrice transposée de X.

Afin de connaître les variables explicatives importantes d'un modèle linéaire gaussien, le test de Student peut être utilisé. Celui-ci consiste à vérifier l'hypothèse $H_0 : \beta_j = c = 0$ contre $H_1 : \beta_j \neq c = 0$. Ainsi, si l'hypothèse nulle H_0 est vérifiée, la variable explicative j n'est pas significative pour le modèle. Pour chaque variable explicative du modèle, la statistique de test suit une loi de Student sous H_0, d'après [Vermet, 2020]. Par conséquent, on peut retirer la variable explicative j du modèle si la p-value du test est supérieure à α (avec un niveau de confiance de $1 - \alpha$).

6.2.2 GLM

Les modèles linéaires généralisés sont utilisés lorsque les conditions du modèle linéaire ne sont pas adaptées à la VAE. En effet, la régression linéaire suppose que les prédictions sont continues et appartienennt à $\mathbb{R}$. Cependant, en actuariat, il est commun de modéliser

des coûts, appartenant à $\mathbb{R}_+$. Les GLM utilisés sont alors Gamma (A.1.6) ou log-normal (A.1.5). Dans le cas de la modélisation d'une variable binaire, prenant ses valeurs dans $\{0, 1\}$, le GLM couramment utilisé est Bernoulli (A.1.1). En reprenant les notations de la partie 6.2.1 et en s'appuyant sur les résultats du cours [Ailliot, 2021], pour n'importe quel GLM, les variables aléatoires $(Y_i)_{i\in[\![1,n]\!]}$ sont indépendantes. Seuls les GLM Gamma, log-normal et Bernoulli (appelé également régression logistique) seront utilisés dans ce chapitre :

— dans le cas du GLM Gamma, il est supposé que

$$\forall i \in [\![1, n]\!],\ Y_i \sim \gamma(\alpha,\ \beta(x_1^{(i)}, ..., x_p^{(i)})),$$

— dans le cas du GLM log-normal, il est supposé que

$$\forall i \in [\![1, n]\!],\ Z_i = \ln(Y_i) \sim \mathcal{N}(\mu(x_1^{(i)}, ..., x_p^{(i)}),\ \sigma^2),$$

et ainsi Y_i suit une loi log-normale,

— quant à la régression logistique, on suppose que

$$\forall i \in [\![1, n]\!],\ Y_i \sim \mathcal{B}(1,\ \pi(x_1^{(i)}, ..., x_p^{(i)})).$$

La relation entre la combinaison des variables explicatives et la VAE est définie par une fonction de lien g telle que :

$$\forall i \in [\![1, n]\!],\ g(\mathrm{E}[Y_i]) = \beta_0 + \beta_1 x_1^{(i)} + ... + \beta_p x_p^{(i)}$$

Afin d'estimer le vecteur de coefficients β, la méthode du maximum de vraisemblance (A.3.2) est utilisée [Ailliot, 2021]. Les fonctions de lien utilisées dans cette étude sont :

— $x \mapsto \ln(x)$ pour les GLM Gamma et log-normal,

— $x \mapsto \ln\left(\frac{x}{1-x}\right)$ pour le GLM Bernoulli.

En outre, le test de Student (6.2.1) s'applique également aux modèles GLM afin d'identifier les variables significatives.

6.2.3 Sélection de variables

Les méthodes de sélection de variables utilisées dans cette étude sont : *backward*, *forward* et *stepwise*. Elles seront appliquées aux GLM.

D'après [Ailliot, 2021], la méthode *backward* consiste à partir du modèle contenant toutes les variables explicatives disponibles, aussi appelé modèle complet, et à éliminer les variables une à une si la métrique d'évaluation du modèle utilisée est améliorée par la suppression de la variable. A chaque étape, la variable supprimée est celle dont la suppression optimise la métrique d'évaluation.

La méthode *forward* consiste à partir du modèle ne contenant aucune variable explicative et à ajouter les variables une à une si la métrique d'évaluation du modèle utilisée

est améliorée par l'ajout de la variable. A chaque étape, la variable ajoutée est celle permettant d'obtenir la valeur optimale de la métrique d'évaluation.

L'inconvénient de la méthode *backward* (respectivement *forward*) est qu'une variable supprimée (respectivement ajoutée) ne peut être ajoutée (respectivement retirée) par la suite. Ainsi, il existe la méthode *stepwise*. Celle-ci combine les deux approches citées ci-dessus [Ailliot, 2021].

Par la suite, les méthodes de sélection de variables se baseront sur un critère d'évaluation de modèle appelé l'AIC : $AIC = 2k - 2\ln(L)$ où k est le nombre de paramètres à estimer pour le modèle et L est le maximum de la fonction de vraisemblance du modèle.

6.2.4 Théorie des valeurs extrêmes

Dans cette section seront présentés quelques résultats sur la loi des dépassements de seuil utilisés par la suite, d'après l'article [Coles, 2001]. Si X est la variable aléatoire étudiée, on s'intéresse à la loi des dépassements de seuil u définie par la fonction de survie $y \mapsto F_u(y) = \mathrm{P}(X > y + u | X > u)$. Cette loi peut être estimée par une loi de Pareto Généralisée, ou GPD, sous certaines conditions. Différentes méthodes existent afin de déterminer le seuil optimal. Deux d'entres elles seront présentées dans l'étude : le graphique de la moyenne des excès ainsi que celui de l'estimation des paramètres d'une loi GPD en fonction de différents seuils.

La fonction moyenne des excès est définie par $\mathrm{E}[X - u | X > u]$. Dans le cas où la distribution GPD peut être utilisée pour modéliser la loi des dépassements de seuil u_0, alors elle est également valide $\forall u > u_0$. Dans le cas où $u > u_0$, $\mathrm{E}[X - u | X > u]$ est une fonction linéaire de u. Le graphique permettant de déterminer le seuil optimal présente la fonction moyenne des excès en fonction de différents seuils. Il s'agit alors de déterminer la valeur de l'axe des abscisses à partir de laquelle la courbe affichée est linéaire.

La deuxième figure permettant de déterminer le seuil optimal d'une distribution X présente les paramètres estimés de la loi GPD pour la fonction $y \mapsto F_u(y)$ pour différentes valeurs de u. Elle permet notamment de compléter la première méthode présentée ci-dessus. Les paramètres de forme et d'échelle de la loi GPD estimés devraient être constants à partir du seuil u_0 optimal.

6.2.5 Algorithmes de Machine Learning

Trois algorithmes de régression seront présentés dans cette section : l'arbre de décision, le *Random Forest* et le *XGBoost*. Les résultats présentés se basent sur le cours [Vermet, 2021]. On note Y la VAE quantitative à valeurs réelles, X la matrice des variables explicatives (il y en a p) et n le nombre d'observations dans l'échantillon d'apprentissage : $Y = f(X) + \epsilon$ avec ϵ représentant le bruit. Les couples de variables aléatoires

$(X_i, Y_i)_{i\in[\![1,n]\!]}$ sont i.i.d de loi inconnue.

L'objectif de l'arbre de régression, utilisant la méthode CART, est de « partitionner l'espace des valeurs des variables explicatives en rectangles sur lesquels la VAE est constante » [Vermet, 2021]. Les régions sont appelées feuilles de l'arbre et sont notées $R_1, ..., R_J$. Par conséquent l'objectif est de minimiser :

$$\sum_{j=1}^{J} \sum_{i\in R_j} (y_i - \hat{y_j})^2$$

où $\hat{y_j} = \frac{1}{n_j} \sum_{i\in R_j} y_i$ et n_j le nombre d'observations dans la feuille R_j. Ce problème est transformé en pratique en approche récursive en définissant deux régions $R_-(j,s) = \{X_j < s\}$ et $R_+(j,s) = \{X_j \geq s\}$ avec X_j la j$^{\text{ème}}$ variable explicative, s correspondant à un seuil. La variable explicative ainsi que le seuil sont choisis à chaque itération en minimisant la variance des deux sous-arbres obtenus :

$$\frac{1}{n} \sum_{i\in R_-(j,s)} (y_i - y_{\hat{R}_-})^2 + \frac{1}{n} \sum_{i\in R_+(j,s)} (y_i - y_{\hat{R}_+})^2$$

Le processus récursif s'arrête grâce à un critère d'arrêt défini initialement, comme par exemple une profondeur maximale ou un nombre d'observations minimal dans les nœuds. Cela constitue la première étape de l'algorithme CART. La deuxième étape consiste à élaguer l'arbre de décision obtenu, appelée *pruning*. Cette étape vise à réduire la complexité de l'arbre final en regroupant certaines feuilles si l'erreur de prévision sur la base de test diminue. Ce processus permet notamment de limiter le surapprentissage.

L'algorithme *Random Forest* consiste à agréger plusieurs arbres de décision construits indépendamment les uns des autres. Le nombre de prédicteurs B ainsi que m, représentant le nombre de variables explicatives tirées aléatoirement parmi l'ensemble des p variables à chaque coupure des différents arbres, sont fixés initialement. Le choix de restreindre le nombre de variables explicatives à utiliser à chaque coupure permet de diminuer la corrélation entre les B arbres de décision et ainsi, améliorer le prédicteur final. De plus, pour chaque arbre, l'échantillon d'apprentissage n'est pas complet mais $l \leq n$ observations sont tirés aléatoirement sans ou avec remise (notamment dans le cas $l = n$). Pour une observation quelconque, le modèle *Random Forest* prédit la moyenne des B prédictions des différents arbres CART utilisés.

Le dernier algorithme présenté, *XGBoost* ou *Extreme Gradient Boosting*, repose sur le même principe que l'algorithme *Random Forest* sauf que les B arbres de décision utilisés ne sont pas indépendants. En effet, ils sont construits les uns à la suite des autres en corrigeant les erreurs des précédents ([Chen et Guestrin, 2016]). Cette approche est dite adaptative, également appelée *boosting*. L'ensemble des algorithmes de *boosting* donnent un poids aux observations de la base d'apprentissage. Ce poids augmente au fil des arbres

utilisés si les observations sont mal prédites par les arbres précédents, dans le but d'améliorer les arbres au fur et à mesure de la phase d'apprentissage. Les premiers algorithmes utilisant la méthode *boosting* sont Adaboost pour une VAE binaire puis l'algorithme de Drucker pour la régression et enfin, le *gradient boosting machine*. La fonction de perte Q à minimiser dans les modèles de régression peut être quadratique, exponentielle, etc. Le modèle *gradient boosting machine*, tout comme l'algorithme *XGBoost*, repose sur l'application d'une méthode d'optimisation du gradient afin de minimiser la fonction de perte Q. En effet, le deuxième modèle est entraîné sur le gradient de Q liées aux prédictions du premier modèle, et ainsi de suite. La principale différence entre le *gradient boosting machine* et le *XGboost* est que ce dernier utilise des modèles plus régularisés afin d'éviter le surapprentissage.

6.2.6 Interprétabilité des modèles de Machine Learning

Premièrement, afin d'identifier les variables importantes des modèles basés sur plusieurs arbres de décision, il existe une fonction du langage de programmation Python permettant de classer les variables par ordre de significativité basée sur la diminution moyenne d'impureté dans les noeuds des arbres de décision utilisés. L'impureté est mesurée à chaque nœud et indique l'efficacité de chaque caractéristique (variable ou modalité dans le cas de variables catégorielles) à réduire l'incertitude lors de la séparation des données d'apprentissage. La diminution moyenne d'impureté est calculée sur l'ensemble des arbres utilisés dans le cas des algorithmes *Random Forest* et *XGBoost*.

En outre, afin d'interpréter n'importe quel modèle de prédiction, il existe une autre méthode permettant d'évaluer l'importance des variables, appelée valeur SHAP. La mise en œuvre de la méthode SHAP repose sur l'estimation des valeurs de Shapley (5.4) de chaque variable et pour chaque individu de la base de données, selon [Dehoux et Liao, 2019]. Pour une observation donnée, la valeur de Shapley d'une variable représente sa contribution à la différence entre la valeur prédite par le modèle et la moyenne de toutes les prédictions de celui-ci. Ainsi, l'effet d'une variable est mesuré en testant toutes les combinaisons de variables possibles du *dataset*. Puis, le calcul de la moyenne des valeurs SHAP d'une variable sur toutes les prédictions donne la valeur SHAP globale de la variable. Cette dernière méthode sera priviligiée afin de connaître les variables importantes pour les algorithmes *Random Forest* et *XGBoost*.

6.2.7 Métriques d'évaluation utilisées

Dans le cas des algorithmes de régression, deux métriques seront utilisées : la RMSE et la corrélation de Spearman.
Afin de comparer les performances des modèles de régression, la RMSE est généralement

utilisée. Elle est définie par la formule suivante :

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{n}}$$

avec n le nombre d'observations dans la base, y_i la valeur réelle de l'observation n°i et $\hat{y}_i$ la prédiction du modèle pour l'observation n°i.
De plus, une des utilisations opérationnelles possibles du coût marginal de réassurance étant l'identification des sites industriels contribuant fortement à la prime de réassurance payée par Generali pour le traité CAT XL, il est important que le modèle soit en mesure de reproduire l'ordre d'intensité des coûts marginaux des sites industriels. Ainsi, la corrélation de Spearman sera utilisée. Elle correspond à une corrélation de rang, appartenant à l'intervalle $[-1, 1]$, et permet d'étudier la corrélation entre deux variables aléatoires X et Y sans que la relation entre les deux soit de type affine. Finalement, cette dernière consiste à calculer la corrélation de Pearson (A.2.1) entre les rangs des valeurs de X et Y, notés $rang_X$ et $rang_Y$. La corrélation de Spearman est définie par la formule suivante :

$$r_S = \frac{\text{Cov}\left(rang_X, rang_Y\right)}{\sigma_{rang_X}\sigma_{rang_Y}}$$

Dans le cas des modèles de classification, les métriques utilisées sont différentes de celles utilisées pour les modèles de régression. Concernant la régression logistique, la variable réponse appartient à $\{0, 1\}$. Néanmoins, le modèle prédit la probabilité que l'observation ait comme valeur de VAE 1. Ainsi, il est nécessaire de définir un seuil sur cette probabilité. Pour cela, différentes métriques peuvent être maximisées en fonction du seuil : la précision, le rappel, le f1-score et l'*accuracy*.
La précision peut se définir comme le taux de prédictions correctes parmi les prédictions à 1 :

$$précision = \frac{TP}{TP + FP}$$

Quant au rappel, il représente le taux de sites appartenant à la classe 1 détectés par le modèle :

$$rappel = \frac{TP}{TP + FN}$$

Les sigles TP, FP et FN sont définis ci-dessous :

- TP représente le taux de valeurs correctement prédites à 1,
- FP représente le taux de valeurs prédites à 1 mais appartenant en réalité à la classe 0,
- FN correspond au taux de sites prédits à 0 mais appartenant en réalité à la classe 1,
- TN correspond au taux de valeurs correctement estimées à 0 par le modèle.

Le f1-score se base à la fois sur la précision et sur le rappel :

$$f1 - score = \frac{2}{\frac{1}{précision} + \frac{1}{rappel}}$$

De plus, l'*accuracy* évalue le pourcentage d'individus bien classés par le modèle et correspond donc à :

$$accuracy = \frac{TP + TN}{TP + FP + TN + FN}$$

L'erreur du modèle est alors définie par $1 - accuracy$.

6.3 Prédiction des coûts marginaux de réassurance

6.3.1 Séparation des coûts de réassurance graves et attritionnels

Afin de prédire les coûts marginaux de réassurance des sites RI (appartenant à $\mathbb{R}_+$) du CAT XL pour le péril tempête par des algorithmes de régression, il est important dans un premier temps d'étudier leur distribution. En effet, suivant la queue de distribution de ces coûts, les GLM Gamma et log-normal étant des lois à queue fine ([Thérond, 2012]), ils ne pourront pas toujours s'appliquer. Afin de vérifier la présence d'une queue de distribution épaisse sur les coûts marginaux de réassurance, on a comparé, à l'aide de la méthode des moments, les quantiles empiriques avec les quantiles théoriques d'ajustement de lois à nos coûts marginaux, sur la figure 6.1.

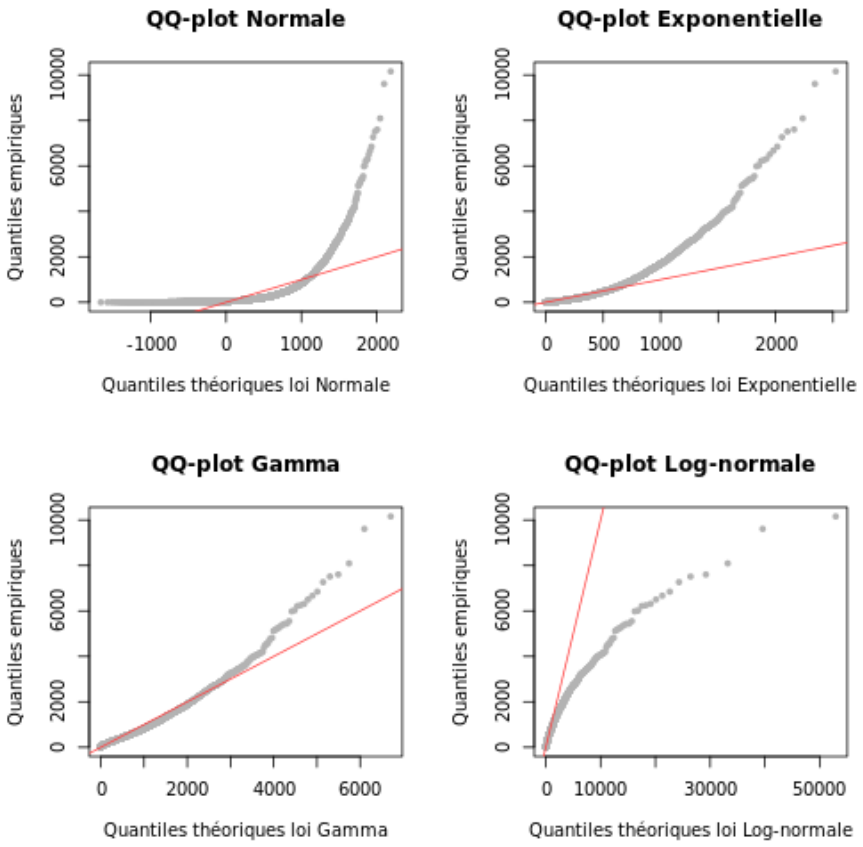


FIGURE 6.1 – QQ-plot des coûts marginaux de réassurance

Le graphique quantile-quantile avec la distribution théorique gaussienne est concave. Cela traduit l'existence d'une queue de distribution à droite. De plus, celui avec la loi exponentielle l'est également. Il semble donc que la distribution des coûts marginaux de réassurance contienne une queue lourde à droite. D'après les deux derniers graphiques quantile-quantile, les lois Gamma et log-normale ne semblent pas non plus être adaptées à nos données. Ainsi, les GLM associés ne peuvent être utilisés directement sur les coûts marginaux de réassurance dans le but de les modéliser.

En vue des graphiques précédents, afin que les GLM Gamma et log-normal soient plus adaptés à la distribution modélisée, il convient de séparer la charge des coûts marginaux de réassurance grave et attritionnelle. Pour cela, les coûts marginaux des sites RI doivent être écrêtés à l'aide d'un seuil grave. Le seuil sur les coûts marginaux de réassurance est donc choisi de sorte que la charge écrêtée, c'est-à-dire la charge des coûts marginaux inférieurs au seuil grave, se rapproche d'une loi Gamma et/ou log-normale. La charge écrêtée

correspond à la charge attritionnelle alors que la charge de sur-écrêtement correspond à la charge au-dessus du seuil grave. Le seuil choisi est 900. Dans un premier temps, les méthodes classiques de détermination de seuil provenant de la théorie des valeurs extrêmes sont appliquées (figures 6.2 et 6.3).

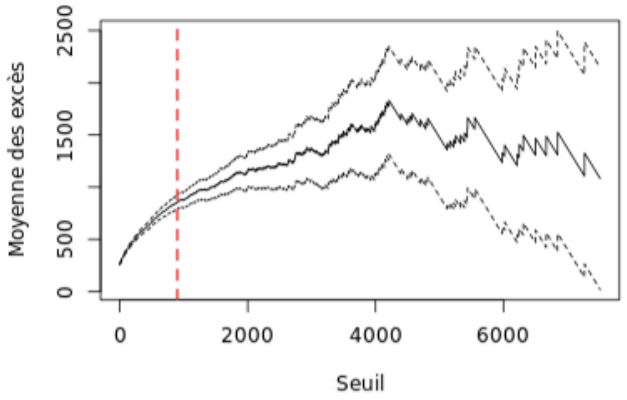


FIGURE 6.2 – Fonction moyenne des excès en fonction du seuil sur les coûts marginaux

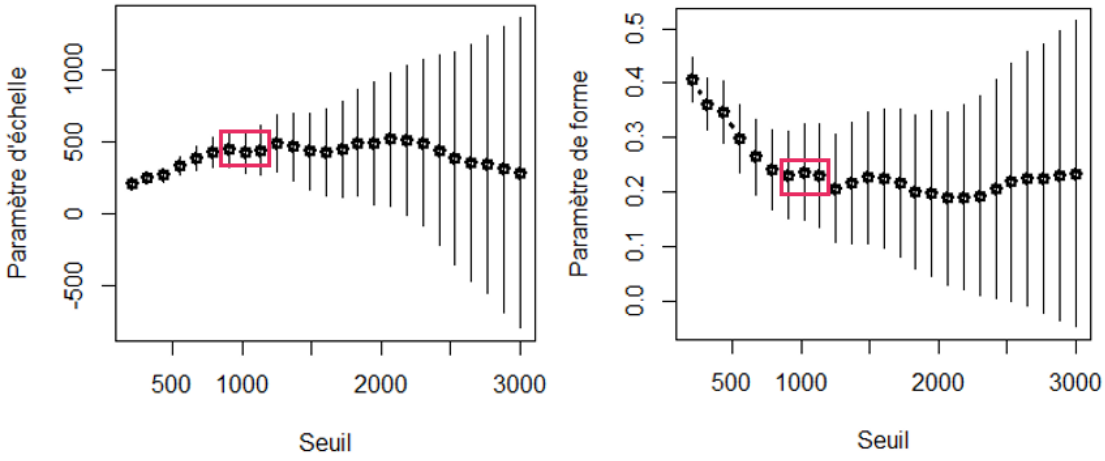


FIGURE 6.3 – Paramètres d'une loi GPD en fonction du seuil sur les coûts marginaux

Les droites et encadrés rouges des figures 6.2 et 6.3 correspondent au seuil de 900. Sur le premier graphique, la courbe est linéaire à partir de ce seuil. A partir du seuil égal à 2 500, il n'y a plus assez d'observations pour que la courbe soit suffisamment fiable. De plus, à l'aide des graphique de la figure 6.3, on observe que les paramètres d'échelle et de forme pour une loi GPD sont constants autour de la valeur 900. Néanmoins, ces méthodes d'estimation de seuil se basent sur les paramètres de distribution du sur-écrêtement. D'autres tests ont donc été réalisés sur la distribution des coûts marginaux inférieurs au seuil de

900, notamment sur la figure 6.4.

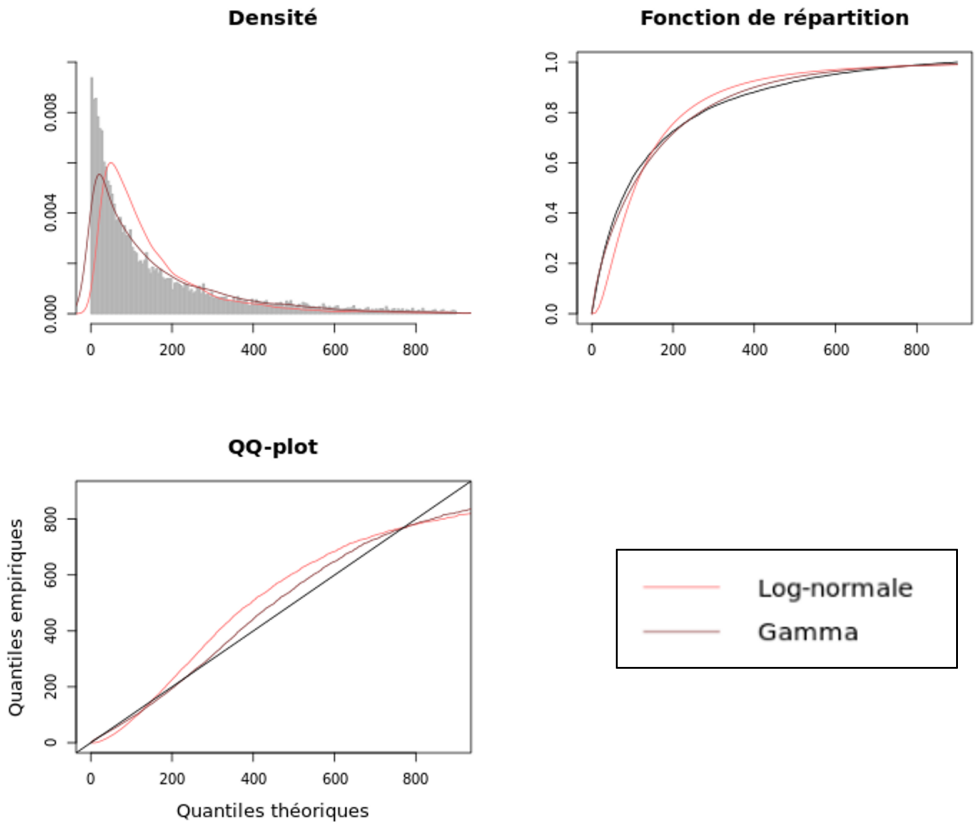


FIGURE 6.4 – Ajustement de lois log-normale et Gamma à la distribution des coûts marginaux de réassurance inférieurs à 900

A l'aide des graphiques ci-dessus, on peut voir que la définition du seuil sur les coûts marginaux de réassurance permet un ajustement des lois Gamma et log-normale plus adéquat. Afin de séparer la modélisation de la charge grave et attritionnelle, on peut décomposer les coûts marginaux de réassurance à l'aide de la formule suivante :

$$CM = CM_{écrêté} + P_{grave} \cdot Sur\text{-}écrêtement \tag{6.1}$$

avec :

— CM le coût marginal de réassurance,

— $CM_{écrêté} = \begin{cases} CM & \text{si } CM \leq 900 \\ 900 & \text{sinon} \end{cases}$,

— $P_{grave} = \begin{cases} 0 & \text{si } CM \leq 900 \\ 1 & \text{sinon} \end{cases}$,

— $Sur\text{-}écrêtement = \begin{cases} 0 & \text{si } CM \leq 900 \\ CM - 900 & \text{sinon} \end{cases}$.

La démarche suivie pour la modélisation est la suivante :

1. modélisation de $CM_{écrêté}$ par un GLM Gamma puis par un GLM log-normal,
2. modélisation de P_{grave} par une régression logistique,
3. modélisation de *Sur-écrêtement* par un arbre de décision de faible profondeur.

Premièrement, l'estimation des coûts marginaux de réassurance écrêtés à 900, également appelés charge attritionnelle, est effectuée à l'aide d'un GLM Gamma puis d'un GLM log-normal. Afin de sélectionner les variables à utiliser en entrée des modèles GLM, la méthode de sélection de variables *stepwise* est utilisée. Les variables sélectionnées pour les deux GLM sont les suivantes :

— l'ensemble des variables explicatives sauf la somme assurée BM pour le GLM log-normal,
— seule la variable libellé de réseau n'est pas significative pour le modèle GLM Gamma.

De plus, afin de confirmer la significativité des variables sélectionnées, le test de Student est effectué pour chacune d'entre elles : l'ensemble des variables explicatives, pour les deux modèles de prédiction, ont des p-value inférieures à 5% et sont donc importantes à un niveau de confiance de 95%. L'apprentissage du GLM Gamma (respectivement log-normal) sur le coût marginal de réassurance écrêté donne une RMSE sur la base de test égale à 167 (respectivement 137). Les graphiques 6.5 et 6.6 présentent les graphes des prédictions en fonction des valeurs de coût marginal écrêté observées. La RMSE associée au modèle GLM Gamma est plus élevée que pour le GLM log-normal car ce dernier prédit des valeurs de coûts marginaux de réassurance moins élevées par rapport aux observations.

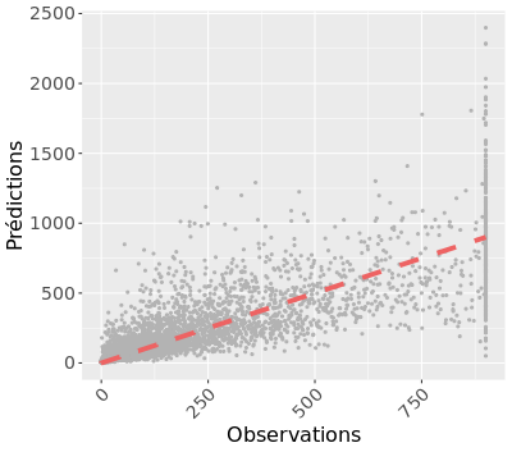


FIGURE 6.5 – Prédictions par un GLM Gamma des coûts marginaux écrêtés en fonction des observations

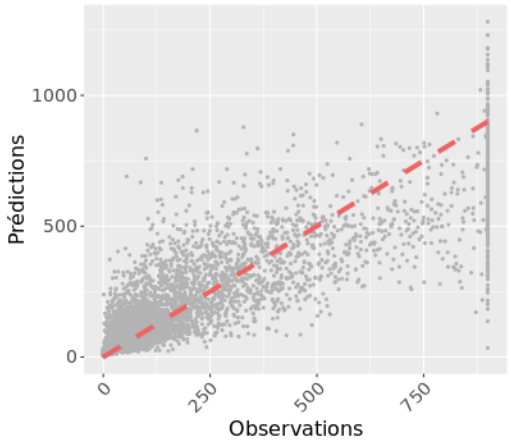


FIGURE 6.6 – Prédictions par un GLM log-normal des coûts marginaux écrêtés en fonction des observations

Deuxièmement, on souhaite prédire, à l'aide d'un modèle de régression logistique, la probabilité que le coût marginal du site industriel considéré soit supérieur à 900. L'objectif final étant d'obtenir P_{grave}, soit 0 ou 1, il est nécessaire de définir un seuil sur cette probabilité à partir duquel la prédiction est égale à 1. Ce seuil est déterminé en maximisant le f1-score sur la base de test, soit pour un seuil à 0.26 et un f1-score égal à 63%. De plus, le graphique 6.7 présente l'erreur sur les bases d'apprentissage et de test, permettant de confirmer le choix du seuil à 0.26.

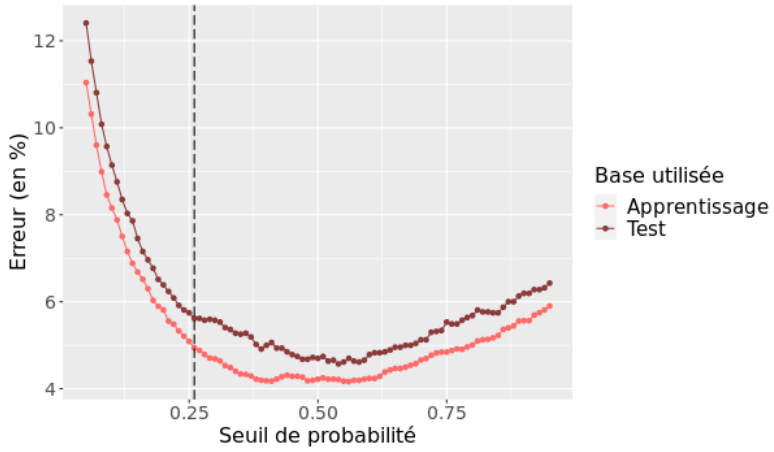


FIGURE 6.7 – Régression logistique pour les coûts marginaux : taux d'erreur en fonction du seuil

Pour un seuil de probabilité égal à 0.26, les résultats de la régression logistique sont

résumés dans le tableau 6.2.

	Erreur	Rappel	Précision	F1-score
Base d'apprentissage	5%	79%	57%	66%
Base de test	6%	71%	56%	63%

TABLE 6.2 – Métriques d'évaluation de la régression logistique pour les coûts marginaux

Enfin, la charge de sur-écrêtement est modélisée par un arbre de décision de faible profondeur. Seules quelques variables seront sélectionnées dans la modélisation : l'engagement assuré pour la garantie tempête, la somme assurée bâtiment, la surface, la valeur au mètre carré ainsi que le zonier vent du produit MRC. Le schéma de l'arbre de décision utilisé est affiché sur la figure 6.8.

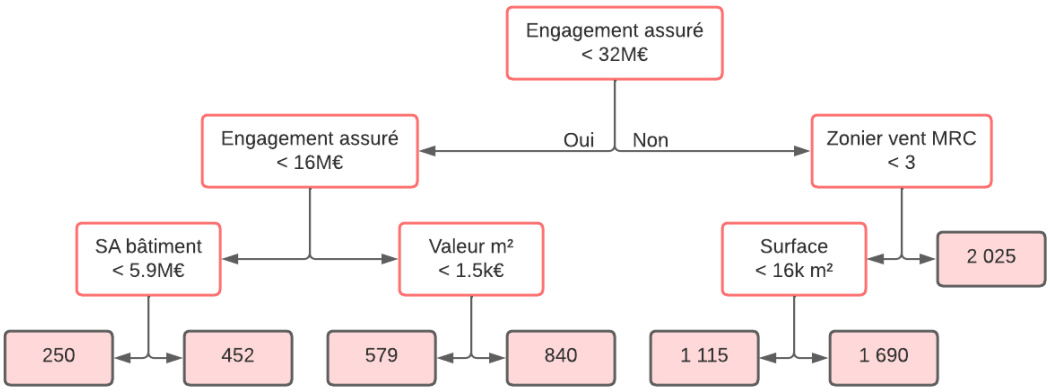


FIGURE 6.8 – Arbre de décision modélisant le sur-écrêtement des coûts marginaux

Pour conclure, en regroupant les trois étapes de modélisation par la formule 6.1, la RMSE ainsi que la corrélation de Spearman sont calculées sur les prédictions du coût marginal de réassurance total des sites du portefeuille RI de la base de test.

Algorithme sur la charge attritionnelle	RMSE test	Corrélation de Spearman
GLM Gamma	306	0.89
GLM log-normal	295	0.87

TABLE 6.3 – RMSE et corrélation de Spearman du coût marginal de réassurance sur la base de test avec séparation des graves et attritionnels

Ainsi, la RMSE obtenue lorsque la charge attritionnelle est modélisée par un GLM

Gamma est plus élevée qu'avec un GLM log-normal. Ce dernier semble donc plus performant pour prédire les coûts marginaux. Les graphiques 6.9 et 6.10 représentent les prédictions du coût marginal de réassurance total en fonction des observations.

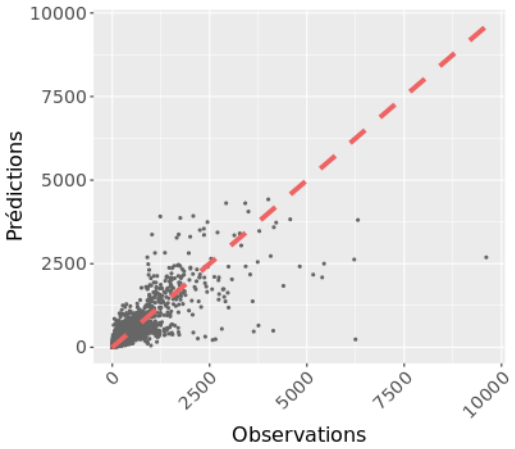


FIGURE 6.9 – Prédictions des coûts marginaux totaux en fonction des observations pour le modèle avec GLM Gamma

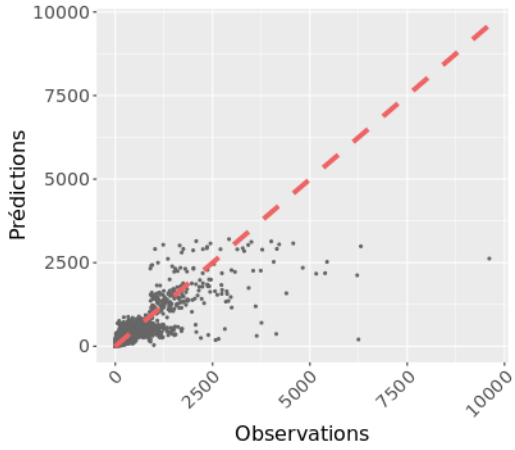


FIGURE 6.10 – Prédictions des coûts marginaux totaux en fonction des observations pour le modèle avec GLM log-normal

Sur la figure 6.10, à partir de 2 000 €, le modèle sous-estime la VAE alors que sur le graphique 6.9, le modèle prédit correctement les coûts marginaux de réassurance élevés jusqu'au seuil de 3 500 € environ. Ainsi, malgré que le modèle de prédiction composé du GLM log-normal ait une RMSE plus faible que celui s'appuyant sur le GLM Gamma pour le coût écrêté, ce dernier est plus performant à prédire les coûts marginaux de réassurance élevés et sera donc utilisé en priorité afin d'identifier les sites industriels contribuant fortement au traité CAT XL pour le péril tempête.

6.3.2 Algorithmes de Machine Learning

En vue des résultats performants des algorithmes de Machine Learning, un *Random Forest* et un *XGBoost* seront testés afin de prédire les coûts marginaux de réassurance. Aucune séparation de la charge n'est nécessaire puisqu'aucune hypothèse n'est faite sur la distribution de la VAE pour ces deux modèles.

Afin de déterminer les variables importantes pour ces algorithmes, les valeurs de SHAP sont calculées. Les variables explicatives importantes sont similaires pour le *Random Forest* et le *XGBoost* : l'engagement assuré, l'ensemble des sommes assurées, la surface, la valeur au mètre carré, le type d'occupation, la franchise du site ainsi que le zonier vent MRC. Les graphiques 6.11 et 6.12 sont des exemples de valeurs SHAP prises par deux variables explicatives pour l'algorithme *Random Forest*.

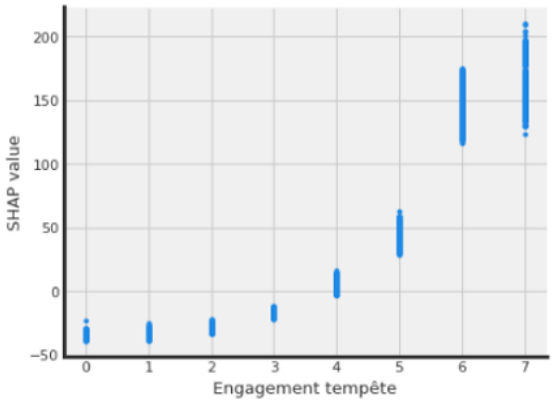


FIGURE 6.11 – *Random Forest* : valeur SHAP de l'engagement assuré

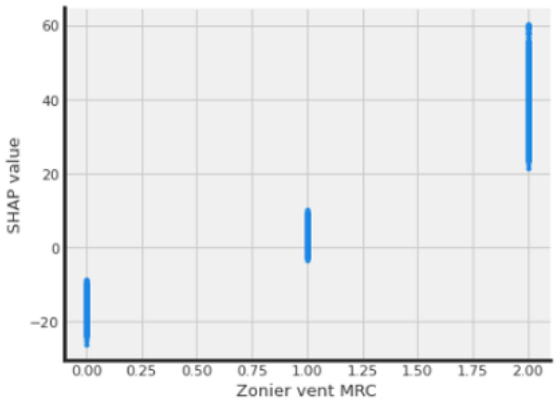


FIGURE 6.12 – *Random Forest* : valeur SHAP du zonier vent MRC

L'engagement assuré et le zonier vent MRC sont des variables ordinales. Ainsi, leurs modalités sont triées par ordre croissant d'intensité. Plus la modalité d'engagement assuré est élevée et plus celui-ci est important. De même, plus la modalité du zonier est élevée et plus le département est exposé au risque tempête, évalué pour le portefeuille MRC. Par conséquent, d'après les graphiques ci-dessus, le coût marginal de réassurance des sites RI augmente avec le montant d'engagement assuré pour la garantie tempête et l'exposition du site à l'aléa tempête. Cela est cohérent puisqu'un engagement assuré important implique des sinistres élevés et donc une forte contribution aux cessions annuelles du traité CAT XL. De plus, malgré que le zonier ait été construit sur le produit MRC, celui-ci semble adapté aux sites RI puisque le coût marginal de réassurance pour le péril tempête augmente lorsque que la zone dans laquelle se trouve le site industriel est exposée à l'aléa tempête MRC.

La RMSE ainsi que la corrélation de Spearman associées à chaque modèle de prédiction ont été calculées sur la base de test et également sur la base d'apprentissage afin d'éviter le phénomène de surapprentissage.

Algorithme sur la charge totale	RMSE test	RMSE apprentissage	Corrélation de Spearman
Random Forest	414	399	0.77
XGBoost	317	260	0.90

TABLE 6.4 – RMSE et corrélation de Spearman du coût marginal de réassurance sur la base de test pour le *Random Forest* et le *XGBoost*

Ainsi, le modèle *XGBoost* est plus performant que l'algorithme *Random Forest* au regard de l'ensemble des métriques d'évaluation. Sa RMSE est proche de celles des deux modèles de prédiction précédents. Les graphiques des prédictions en fonction des observations du

coût marginal de réassurance du traité CAT XL, pour le péril tempête, sont présentés sur les figures 6.13 et 6.14.

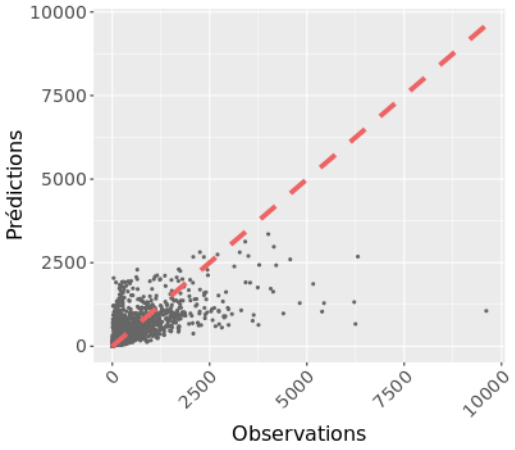


FIGURE 6.13 – Prédictions des coûts marginaux en fonction des observations pour le *Random Forest*

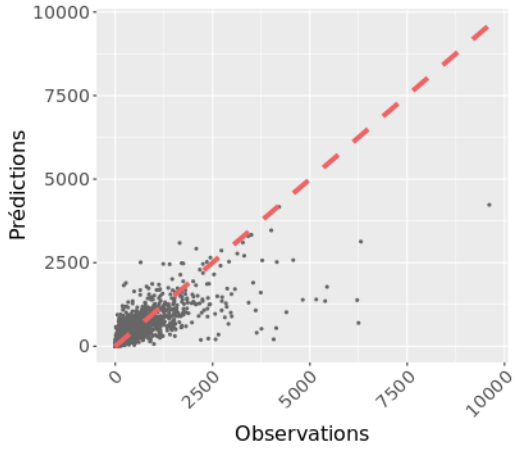


FIGURE 6.14 – Prédictions des coûts marginaux en fonction des observations pour le *XGBoost*

Ces graphiques confirment les propos précédents. Encore une fois, l'algorithme *Random Forest* est moins performant sur les valeurs élevées de coûts marginaux par rapport au *XGBoost*. Cependant, sur la figure 6.14, on peut voir que l'algorithme a des difficultés à prédire les valeurs de VAE supérieures à 3 000.

6.4 Modèle fréquence/coût moyen de cession de réassurance

6.4.1 Présentation du modèle

Généralement, la tarification des contrats d'assurance se base sur un modèle « *Fréquence* × *Coût moyen* » pour lequel les deux grandeurs citées sont indépendantes. Grâce à l'YLT au niveau portefeuille obtenue pour l'ensemble des produits couverts par la garantie DAB Domestique dans le tableau 4.14, une fréquence annuelle de contribution au dépassement de la priorité du traité CAT XL peut être calculée pour l'ensemble des sites RI. En effet, celle-ci se définit par la formule suivante pour le site industriel n°i :

$$\frac{Nombre\ de\ contributions\ du\ site\ i}{50000}$$

où *Nombre de contributions du site i* correspond au nombre de fois où le site RI n°i contribue à un évènement de l'YLT dont la perte est supérieure à la priorité du CAT XL.

Afin de calculer la fréquence de l'ensemble des sites industriels, dans un premier temps, les lignes de l'YLT ayant une perte (colonne « LOSS ») supérieure à la priorité du CAT XL sont sélectionnées. Ensuite, pour chaque site industriel, on regarde par combien

d'évènements sélectionnés celui-ci est touché. L'ELT au niveau site du portefeuille RI permet de connaître les évènements impactant chaque site industriel. Cela nous donne le nominateur de la fraction ci-dessus. Pour finir, il suffit de diviser ce nombre par les 50 000 années de simulation afin d'obtenir la fréquence annuelle de dépassement. L'intérêt de l'utilisation de cette fréquence est que celle-ci est déterministe pour les sites du portefeuille RI existant (dont les pertes sont comptabilisées dans l'YLT) et ainsi, il n'est pas nécessaire de la modéliser dans un premier temps. On dispose donc d'une grandeur déterministe pouvant être en mesure d'expliquer une partie du coût marginal de réassurance. La corrélation de Spearman entre le coût marginal de réassurance et la fréquence définie ci-dessus est égale à 0.64. L'histogramme des fréquences obtenues pour l'ensemble des sites industriels du portefeuille existant est affiché sur le graphique 6.15.

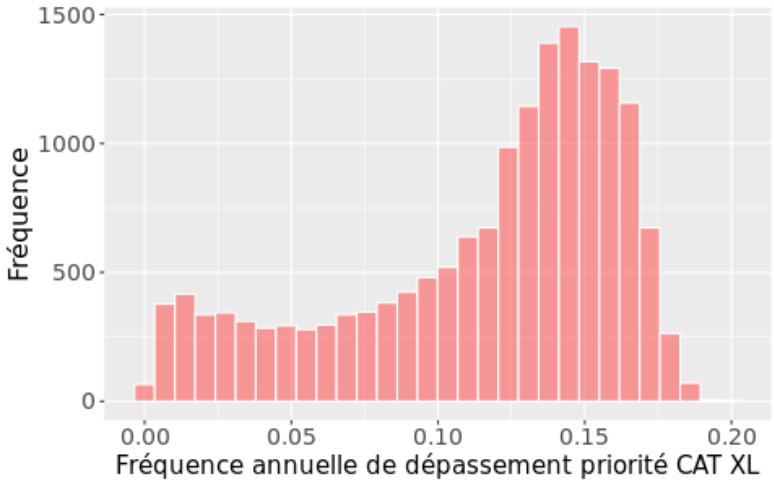


FIGURE 6.15 – Histogramme de la fréquence annuelle de contribution au dépassement de la priorité du CAT XL des sites RI pour les évènements tempête

Il est à noter qu'aucun site du portefeuille RI n'a une fréquence nulle. Cela signifie que chaque site assuré est impacté au moins une fois par un évènement de l'YLT du portefeuille DAB Domestique dont la perte est supérieure à la priorité du traité CAT XL.

Le coût marginal de réassurance du traité CAT XL, pour le péril tempête, peut donc être décomposé de cette façon :

$$CM = Fréquence \cdot Coût\ moyen\ de\ réassurance \tag{6.2}$$

Le coût marginal de réassurance étant la contribution du site au prix de réassurance du CAT XL de l'année 2023, celui-ci peut être vu comme la prime de réassurance d'un site RI sur l'année 2023 et donc comme la prime de réassurance annuelle du site. Ainsi, le coût moyen de cession de réassurance peut se définir comme étant la contribution marginale moyenne du site aux cessions du CAT XL pour un évènement sur les 50 000 années de

simulation, c'est-à-dire le montant de contribution moyen pour un site RI sur les années d'« historique » de l'YLT. Son expression pour le site industriel n°i est la suivante :

$$\frac{CM(i) \cdot 50000}{Nombre\ de\ contributions\ du\ site\ i} = \frac{CM(i)}{Fréquence(i)}$$

Au nominateur, il est nécessaire de multiplier le coût marginal de réassurance par le nombre d'années de simulation puisque celui-ci correspond à une charge annuelle et non pas à une charge sur l'ensemble des années de simulation. En effet, dans les méthodes de tarification des primes d'assurance, le coût moyen se définit de la façon suivante :

$$Coût\ moyen = \frac{\sum_{i=1}^{n} S_i}{\sum_{i=1}^{n} N_i},$$

avec n le nombre d'années dans l'historique des sinistres, S_i le montant des sinistres pour l'année i et N_i le nombre de sinistres survenus durant l'année i. Dans cette section, il suffira ainsi d'appliquer un modèle de prédiction au coût moyen de cession de réassurance défini ci-dessus.

6.4.2 Utilisation d'une méthode d'interpolation spatiale sur la fréquence

La fréquence annuelle est déterministe pour les sites du portefeuille existant RI et a été calculée. Toutefois, afin que le souscripteur puisse disposer de cette fréquence lors de la souscription d'un nouveau site industriel, il est nécessaire que chaque point de coordonnées de la France métropolitaine ait une valeur de fréquence associée. Ainsi, une méthode d'interpolation spatiale est utilisée : la pondération inverse à la distance, plus couramment appelée *Inverse Distance Weighting*. Les méthodes d'interpolation spatiale consistent à assigner une valeur d'une grandeur recherchée à tout point de coordonnées d'une zone fixée à l'aide des points pour lesquels cette grandeur est connue.

Le processus de pondération inverse à la distance est défini par la formule suivante :

$$u(x) = \frac{\sum_{k=0}^{N} w_k(x)^p \cdot u_k}{\sum_{k=0}^{N} w_k(x)^p} \tag{6.3}$$

avec :

- u la grandeur à interpoler, ici la fréquence annuelle,
- x le point de coordonnées pour lequel on cherche à interpoler la grandeur,
- N le nombre de points connus à utiliser pour la pondération, c'est-à-dire le nombre de plus proches voisins à utiliser,

— $w_k(x) = \frac{1}{d(x, x_k)}$ où $d(x, x_k)$ correspond à la distance entre le point x et le point connu x_k,
— $p > 0$.

Avant d'appliquer cette méthode, comme on cherche une valeur de fréquence annuelle pour l'ensemble des points de coordonnées de la France métropolitaine, il est nécessaire de définir une grille de résolution. La résolution choisie pour la grille utilisée dans la méthode de pondération inverse à la distance est égale 200 mètres. L'interpolation spatiale permet d'obtenir la valeur de fréquence annuelle de chaque carré de la grille et plus précisément du sommet haut gauche du carré. Ainsi, plus la résolution choisie est faible et plus il y a de valeurs à déterminer. Les résultats seront donc plus précis. Un exemple de grille de résolution de 20 kilomètres sur la France métropolitaine est présenté sur la figure 6.16.

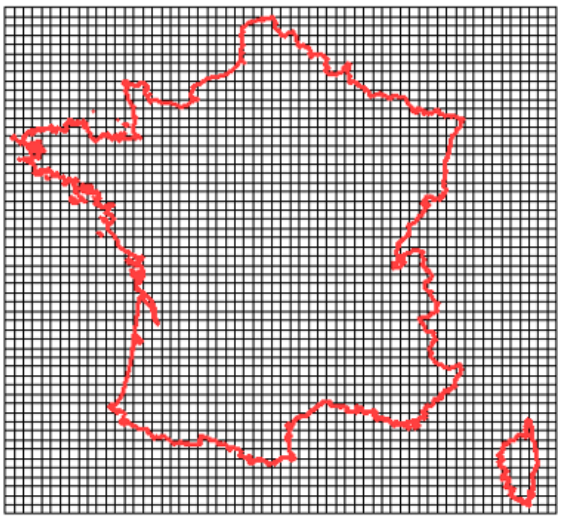


FIGURE 6.16 – Grille de résolution de 20 kilomètres sur la France métropolitaine

De plus, afin de déterminer les paramètres p et N optimaux de la formule 6.3, une base d'apprentissage et une base de test sont créées pour évaluer les résultats de la méthode d'interpolation spatiale, à l'aide de la RMSE sur les fréquences annuelles prédites. Les paramètres permettant de minimiser la RMSE sur la base de test sont $p = 1$ et $N = 100$. Il est à noter qu'une fois les paramètres optimaux obtenus, l'algorithme d'*Inverse Distance Weighting* est appliqué sur l'ensemble des sites RI du portefeuille existant. La carte des fréquences annuelles de contribution des sites RI au traité CAT XL pour le péril tempête obtenue est présentée sur la carte 6.17.

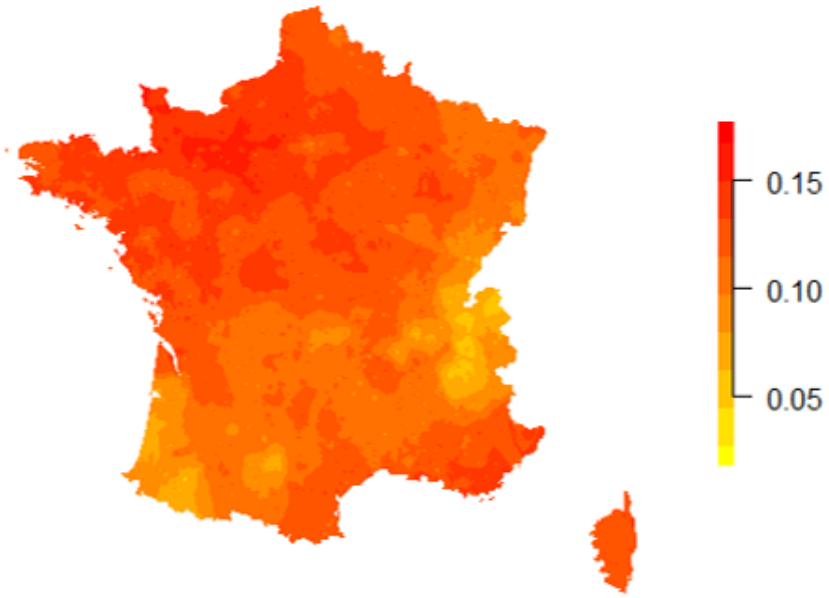


FIGURE 6.17 – Fréquences annuelles de contribution au dépassement de la priorité du CAT XL des sites RI interpolées

6.4.3 Prédiction des coûts moyens de cession de réassurance

Cette partie se concentre sur la modélisation des coûts moyens de cession de réassurance associés aux sites RI, décrits dans la formule 6.2. Comme la distinction graves/attritionnels a permis d'identifier les sites industriels ayant des coûts marginaux de réassurance élevés, nous allons comparer la distribution des coûts moyens de cession de réassurance à des lois statistiques afin de remarquer si la distinction est nécessaire pour ces derniers.

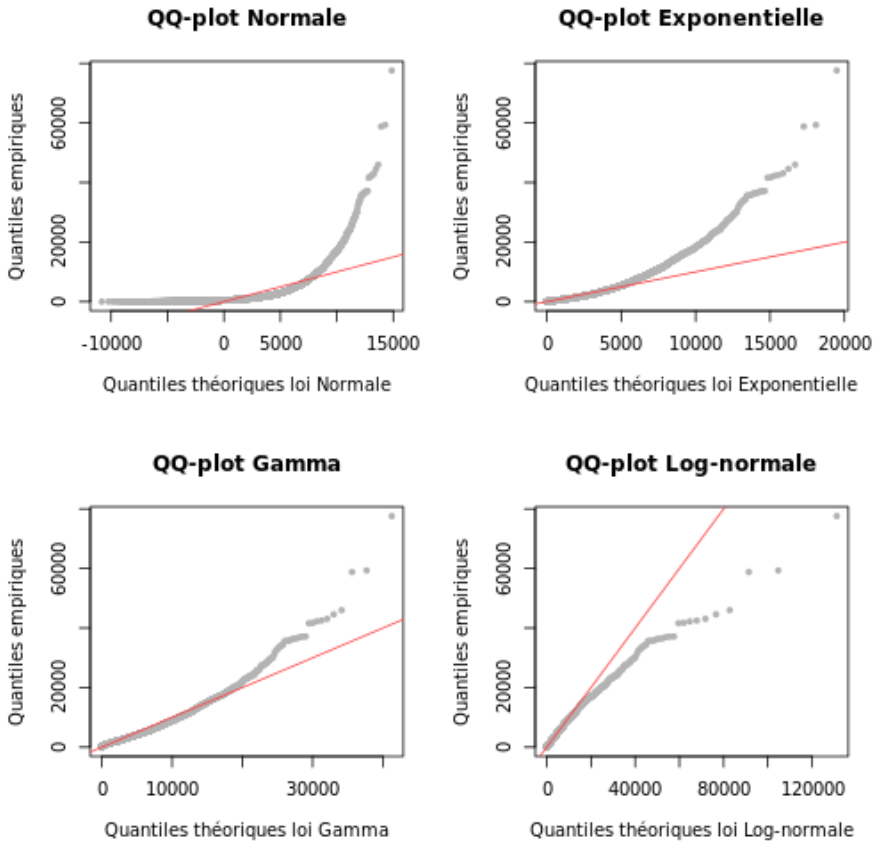


FIGURE 6.18 – QQ-plot des coûts moyens de cession de réassurance

De même que pour les coûts marginaux de réassurance, la distribution des coûts moyens semblent détenir une queue épaisse en vue des graphiques quantile-quantile normale et exponentielle de la figure 6.18. De plus, les lois Gamma et log-normale ne s'ajustent pas correctement à leur distribution. Par conséquent, une distinction de la charge des coûts moyens de cession de réassurance grave et attritionnelle sera réalisée. Le seuil grave choisi sur les coûts moyens est égal à 4 800. Celui-ci est choisi de la même façon que le seuil sur le coût marginal dans la section 6.3.1, à l'aide de la théorie des valeurs

extrêmes. Les différents graphiques sont présentés en annexe dans la section A.4. Une démarche de modélisation similaire à celle suivie pour les coûts marginaux sera adoptée. Le coût marginal de réassurance peut alors se décomposer de cette façon :

$$CM = Fréquence \cdot Coût\ moyen \tag{6.4}$$
$$= Fréquence \cdot (Coût\ moyen_{écrêté} + P_{grave} \cdot Sur\text{-}écrêtement) \tag{6.5}$$

avec :

- CM le coût marginal de réassurance,
- $Fréquence$ la fréquence annuelle de contribution des sites RI au dépassement de la priorité du traité CAT XL pour le péril tempête,
- $Coût\ moyen_{écrêté} = \begin{cases} Coût\ moyen & \text{si } Coût\ moyen \leq 4800 \\ 4800 & \text{sinon} \end{cases}$,
- $P_{grave} = \begin{cases} 0 & \text{si } Coût\ moyen \leq 4800 \\ 1 & \text{sinon} \end{cases}$,
- $Sur\text{-}écrêtement = \begin{cases} 0 & \text{si } Coût\ moyen \leq 4800 \\ Coût\ moyen - 4800 & \text{sinon} \end{cases}$.

La démarche suivie pour la modélisation est la suivante :

1. modélisation de $Coût\ moyen_{écrêté}$ par des GLM Gamma et log-normal,
2. modélisation de P_{grave} par une régression logistique,
3. modélisation de $Sur\text{-}écrêtement$ par un arbre de décision de faible profondeur.

Premièrement, un GLM Gamma ainsi qu'un GLM log-normal sont calibrés sur la base d'apprentissage des coûts moyens de cession de réassurance écrêtés à 4 800. Le GLM Gamma donne une RMSE sur la base de test égale à 1 062 et le GLM log-normal permet d'obtenir une RMSE égale à 897. Les RMSE sont plus élevées que celles obtenues lors de la modélisation des coûts marginaux attritionnels car les coûts moyens atteignent des valeurs beaucoup plus élevées que ces derniers. En effet, le maximum du coût marginal de réassurance est d'environ 10 000 et le coût moyen maximal est environ égal à 75 000. Les graphes des prédictions des coûts moyens de cession de réassurance des sites RI appartenant à la base de test en fonction des valeurs réelles sont présentés sur les figures 6.19 et 6.20. Une nouvelle fois, on observe que la RMSE associée au modèle GLM log-normal est moins élevée que celle obtenue pour le GLM Gamma car les coûts moyens prédits par ce dernier atteignent des valeurs plus élevées que les prédictions du GLM log-normal mais également bien supérieures aux valeurs réelles pour certains sites.

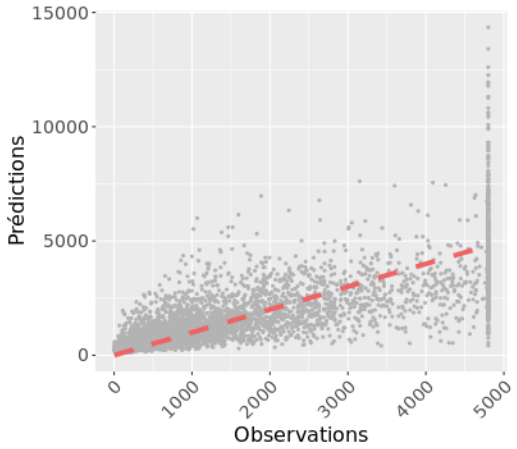


FIGURE 6.19 – Prédictions par un GLM Gamma des coûts moyens écrêtés en fonction des observations

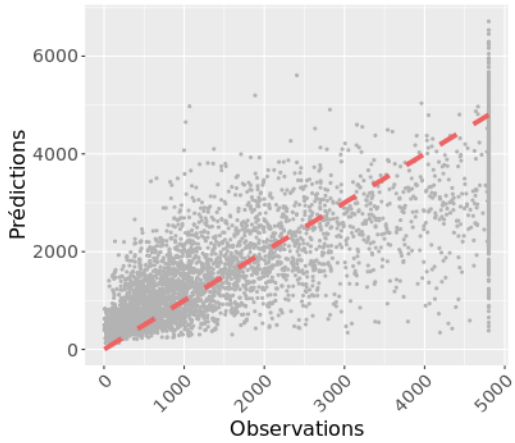


FIGURE 6.20 – Prédictions par un GLM log-normal des coûts moyens écrêtés en fonction des observations

Deuxièmement, une régression logistique est utilisée afin de modéliser P_{grave}. Cette valeur appartient à l'ensemble $\{0, 1\}$. Or, le GLM Bernoulli prédit la probabilité qu'un site RI ait un coût moyen de cession de réassurance pour le traité CAT XL et le péril tempête supérieur à 4 800. Il est donc nécessaire de définir un seuil de probabilité sur les prédictions du modèle. En maximisant le f1-score, on obtient un seuil égal à 0.34, pour un f1-score de 65%. En outre, d'après le graphique 6.21, les erreurs du modèle sur les bases d'apprentissage et de test sont faibles pour ce seuil.

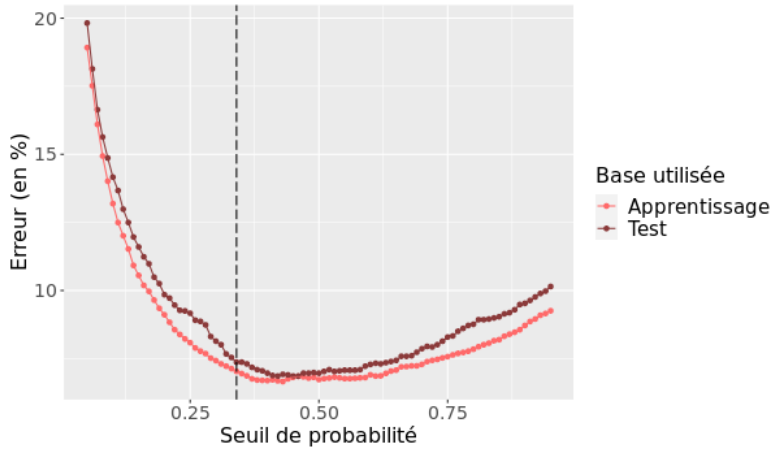


FIGURE 6.21 – Régression logistique pour les coûts moyens : taux d'erreur en fonction du seuil

L'ensemble des métriques calculées sur les bases d'apprentissage et de test définies ini-

tialement afin d'évaluer les résultats de la régression logistique sont affichées :

	Erreur	Rappel	Précision	F1-score
Base d'apprentissage	7%	68%	63%	65%
Base de test	7%	65%	65%	65%

TABLE 6.5 – Métriques d'évaluation de la régression logistique pour les coûts moyens

Pour finir, la charge de sur-écrêtement associée au coût moyen de cession de réassurance est modélisée par un arbre de décision de profondeur égale à deux. Les variables importantes du modèle sont décrites dans le schéma 6.22.

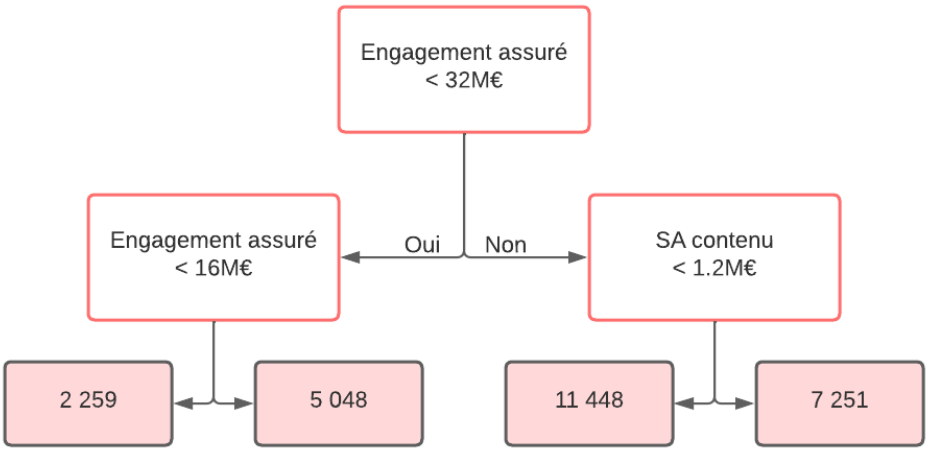


FIGURE 6.22 – Arbre de décision modélisant le sur-écrêtement des coûts moyens

Les résultats de prédiction des coûts marginaux de réassurance de la base de test, lorsque la fréquence et les trois modèles permettant d'estimer les coûts moyens de cession de réassurance sont assemblés d'après la formule 6.5, sont présentés dans le tableau 6.6.

Algorithme sur la charge moyenne attritionnelle	RMSE test	Corrélation de Spearman
GLM Gamma	306	0.90
GLM log-normal	298	0.90

TABLE 6.6 – RMSE et corrélation de Spearman du coût marginal de réassurance sur la base de test pour le modèle fréquence/coût moyen

Ainsi, la séparation du coût marginal de réassurance en fréquence et coût moyen permet seulement d'améliorer légèrement la corrélation de Spearman entre les observations

et prédictions. En effet, les RMSE obtenues ici sont similaires aux modèles de séparation des graves et attritionnels pour les coûts marginaux (tableau 6.3). De plus, il est à noter que pour un nouveau site industriel, la fréquence sera interpolée et ne sera donc plus une valeur déterministe. Par conséquent, il apparaîtra une erreur supplémentaire sur la prédiction des coûts marginaux de réassurance. Pour finir, les graphiques des prédictions en fonction des observations de la base de test pour les modèles fréquence/coût moyen de cession de réassurance sont reportées sur les figures 6.23 et 6.24.

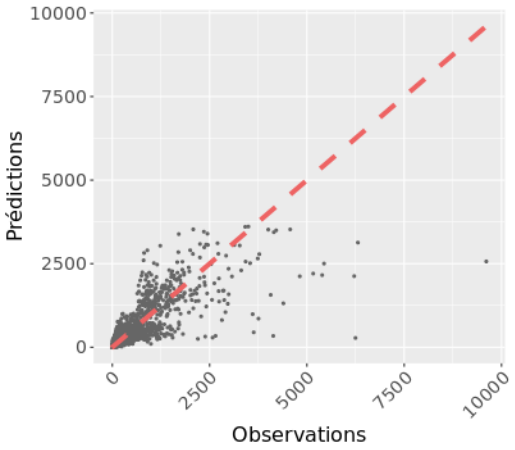


FIGURE 6.23 – Prédictions des coûts marginaux totaux en fonction des observations pour le modèle fréquence/coût avec GLM Gamma

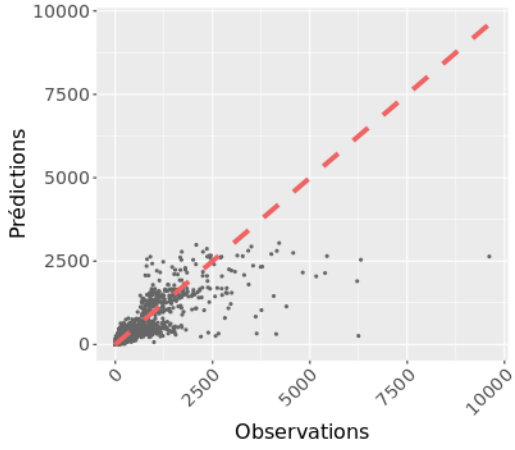


FIGURE 6.24 – Prédictions des coûts marginaux totaux en fonction des observations pour le modèle fréquence/coût avec GLM log-normal

On observe que les points sont plus centrées autour des droites que pour les graphiques 6.9 et 6.10 ce qui indique que les prédictions sont plus proches des observations. Néanmoins, les algorithmes ne sont pas performants sur les coûts marginaux de réassurance élevés, notamment à partir de 3 500 pour la figure 6.23 et 2 500 pour la figure 6.24.

Pour conclure, pour l'ensemble des modèles de prédiction présentés précédemment, les valeurs extrêmes des coûts marginaux de réassurance ne sont pas correctement captées par les algorithmes malgré la définition d'un seuil grave pour quatre d'entre eux.

6.5 Définition de catégories sur le coût marginal de réassurance

Nous disposons d'une base de données d'un peu plus de 10 000 sites assurés RI pour entraîner les modèles de prédiction. De plus, les modèles testés dans les parties précédentes sous-estiment les valeurs de coûts marginaux de réassurance élevés. Afin que l'indicateur mis à disposition de la souscription RI soit fiable, trois classes sont définies sur les coûts marginaux de réassurance pour le traité CAT XL et le péril tempête. Cette

information de classe sur le coût marginal de réassurance pourra s'intégrer à un outil de souscription sous forme d'alerte suivant la classe prédite du nouveau site industriel. L'indicateur ne sera donc pas une information exacte sur le coût de réassurance mais une indication sur le niveau de coût de réassurance supplémentaire engendré par l'acquisition du nouveau site. Ainsi, dans le but d'évaluer les modèles, on ne regarde plus directement la valeur du coût marginal prédite par le modèle mais le pourcentage de bien classés dans les prédictions de chaque modèle. Il est à noter que les classes ne sont pas directement prédites. En effet, les coûts marginaux de réassurance sont prédits dans un premier temps à l'aide des algorithmes présentés ci-dessous, puis classés en fonction de la valeur prédite. Les classes définies ainsi que le pourcentage de sites RI appartenant à chaque classe sont décrits dans la table 6.7.

Classe	Intervalle sur le coût marginal	Pourcentage de sites RI	Pourcentage de charge totale de coût marginal
Faible	$[0, 200]$	68	17
Moyen	$]200, 900]$	26	41
Fort	> 900	6	42

TABLE 6.7 – Définition des classes sur le coût marginal de réassurance pour le traité CAT XL et le péril tempête

Les prédictions sur la base de test issues des six modèles précédents sont classés dans les trois catégories sur le coût marginal de réassurance : faible, moyen et fort. Les pourcentages de bien classés sur les valeurs prédites sont affichés dans le tableau 6.8.

Algorithme	Pourcentage de bien classés		
	Faible	Moyen	Fort
Coût marginal - GLM Gamma	89	78	74
Coût marginal - GLM log-normal	86	85	70
Random Forest	88	67	53
XGBoost	89	77	59
Fréquence/coût - GLM Gamma	89	79	74
Fréquence/coût - GLM log-normal	87	84	74

TABLE 6.8 – Pourcentage de sites bien classés dans les prédictions du coût marginal de réassurance sur la base de test

Les pourcentages affichés confirment les observations faites sur les graphiques précédents. En effet, l'algorithme *Random Forest* ne prédit correctement que 53% des sites industriels dans la classe la plus élevée. Dans l'ensemble, le pourcentage de sites RI bien classés dans la catégorie faible est similaire pour les six modèles. Pour la classe moyenne, les algorithmes utilisant un GLM log-normal affichent des pourcentages supérieurs aux quatre autres. La classe nécessitant l'identification du plus grand nombre de sites industriels est la classe d'intensité forte. En effet, celle-ci regroupe l'ensemble des sites dont le coût

marginal de réassurance pour le traité CAT XL et le péril tempête est supérieur à 2 000 €. Ce tableau montre donc que les algorithmes de Machine Learning ne sont pas performants pour prédire les coûts marginaux des sites RI de cette classe. En revanche, la séparation de la charge grave et attritionnelle permet de bien classer un plus grand nombre de sites dont le coût marginal de réassurance est élevé. En outre, l'ajout de la fréquence permet d'améliorer de 4% le pourcentage de bien classés dans le groupe fort pour le modèle GLM log-normal. Quant au GLM Gamma, ajouter la fréquence permet seulement d'améliorer le pourcentage de la classe moyenne des coûts marginaux de réassurance de 1%. Ainsi, celle-ci ne permet pas d'améliorer significativement les prédictions. En supplément, un biais dans les prédictions sera ajouté lorsque la fréquences des nouveaux sites sera interpolée et non plus déterministe.

6.6 Limites de l'approche de prédiction

Premièrement, les coûts marginaux de réassurance du traité CAT XL, pour le péril tempête, des sites RI dépendent des autres sites du portefeuille DAB Domestique entier touchés par les mêmes évènements tempête. En effet, si une zone en France métropolitaine ne cédait pas encore de sinistres liés aux tempêtes au traité de réassurance CAT XL, en raison d'un engagement assuré global trop faible dans cette zone, et qu'un nouveau site ayant un engagement assuré important est souscrit, celui-ci pourrait engendrer l'activation du traité et ainsi changerait les coûts marginaux des différents sites RI déjà présents dans la zone. Ces propos sont illustrés par l'exemple suivant : supposons une région en France contenant trois sites RI impactés par un évènement tempête dont le taux de destruction est égal à 5% en moyenne. Les sinistres peuvent être cédés à un traité de réassurance de type excédent de sinistre par évènement 5M XS 1M. Les trois sites assurés, leur engagement ainsi que le sinistre dû à l'évènement associé sont décrits dans le tableau 6.9.

Numéro de site RI	Engagement assuré	Sinistre dû à l'évènement tempête
1	3M	0.05×3M = 150k
2	4M	200k
3	1.5M	75k
Total	8.5M	425k

TABLE 6.9 – Exemple d'une région touchée par un évènement tempête et contenant trois sites assurés RI

En vue du total des montants de sinistres causés par l'évènement tempête, aucune cession au traité CAT XL n'est réalisée. Si on suppose que l'YLT n'est composée que des trois lignes ci-dessus, les coûts marginaux de réassurance des trois sites sont nuls car la prime de réassurance est nulle. Cependant, si un site RI avec un engagement assuré de 14 M€ avait été souscrit juste avant la catastrophe naturelle, les coûts marginaux de réassurance seraient différents de 0. En effet, d'après le tableau 6.10, l'YLT servant d'exemple serait modifiée.

Numéro de site RI	Engagement assuré	Sinistre dû à l'évènement tempête
1	3M	0.05×3M = 150k
2	4M	200k
3	1.5M	75k
4	14M	700k
Total	22.5M	1.125M

TABLE 6.10 – Exemple d'une région touchée par un évènement tempête et contenant quatre sites assurés RI

Ainsi, la cession au CAT XL est désormais égale à 125 k€. Comme l'YLT n'est composée que d'une année, la prime de réassurance est également égale à 125 k€. Les coûts marginaux de réassurance des sites RI s'obtiennent en enlevant un à un les sites de l'YLT et en recalculant la prime de réassurance du traité (5.1) :

$$CM(i) = Prime(N) - Prime(N \backslash i),$$

avec $CM(i)$ le coût marginal de réassurance du site n°i, $Prime(N)$ la prime de réassurance obtenue avec les quatre sites RI (125 k€) et $Prime(N \backslash i)$ la prime de réassurance obtenue en retirant le site n°i du portefeuille. Les résultats des coûts marginaux sont présentés dans le tableau 6.11.

Numéro de site RI	$Prime(N)$	$Prime(N \backslash i)$	Coût marginal de réassurance
1	125k	0	125k
2	125k	0	125k
3	125k	50k	75k
4	125k	0	125k

TABLE 6.11 – Exemple de calcul des coûts marginaux de réassurance de quatre sites RI

Ainsi, dans l'exemple, l'acquisition d'un nouveau site industriel dont l'engagement assuré est élevé a fortement fait varier les coûts marginaux de réassurance des sites déjà assurés auparavant. Cet exemple est réalisé sur l'entrée en portefeuille d'un contrat RI mais l'évolution des coûts marginaux peut être due à n'importe quel autre produit du portefeuille DAB Domestique. Par conséquent, si l'engagement assuré de ce portefeuille change significativement au cours d'une certaine période de temps, il sera nécessaire d'exécuter une nouvelle fois les modèles mis en place dans l'étude :

1. modélisation des pertes liées au péril tempête par le logiciel RMS,
2. construction de l'YLT via une méthode de simulation,
3. calibrage du modèle de prédiction des coûts marginaux de réassurance sur le nouveau portefeuille RI.

Il sera nécessaire de recalibrer le modèle de prédiction seulement dans le cas où l'engagement assuré du portefeuille DAB Domestique change significativement à la maille

département puisque celui-ci utilise la variable explicative de localisation CRESTA. Ainsi, ce modèle pourra être revu à des intervalles de temps plus élevés que les deux autres. Une solution pourrait être de relancer les modèles tous les ans. Sinon, il faudrait effectuer des tests de sensibilité afin de déterminer un seuil sur la variation d'engagement assuré du portefeuille DAB Domestique à partir duquel relancer les modèles. Malheureusement, ces tests ne pourront être menés que sur le produit RI puisque l'ELT associée est la seule du portefeuille DAB Domestique décrite à la maille site.

Deuxièmement, les coûts marginaux de réassurance des sites industriels sont proches d'une méthode d'allocation de la contribution du produit RI à la prime du traité CAT XL, pour le péril tempête. Ainsi, si cette dernière change de façon significative d'une année à l'autre, il sera nécessaire de refaire tourner l'ensemble des trois modèles cités ci-dessous puisque les ordres de grandeur des coûts marginaux pourraient varier. Il faudra notamment modifier la définition des catégories sur le coût marginal.

Finalement, les ELT ainsi que les YLT devront être recalculées si la version de RMS est améliorée puisque la modélisation des pertes dues à des évènements tempête en Europe sera différente de celle actuellement utilisée. Des évènements plus extrêmes pourront notamment être construits par RMS en raison du dérèglement climatique. Ainsi, les coûts marginaux du portefeuille existant pourraient être plus élevés dans les zones fortement exposées à l'aléa tempête. De plus, de nouvelles zones pourraient être identifiées comme exposées à des évènements tempête. Par conséquent, les modèles de prédiction devront également être recalibrés.

Pour conclure, ce chapitre permet la mise en place de plusieurs algorithmes de prédiction afin de prédire au mieux le coût marginal d'un nouveau site industriel à l'aide de ses caractéristiques renseignées par le souscripteur. Néanmoins, l'ensemble des algorithmes utilisés sous-estiment les coûts marginaux de réassurance élévés. Il a donc été nécessaire de définir trois catégories sur les coûts marginaux : faible, moyen et fort. En vue des résultats obtenus pour les différents algorithmes (tableau 6.8), le modèle fréquence/coût moyen pour lequel la charge attritionnelle du coût moyen est modélisée par un GLM log-normal, est préconisé. Cependant, il faut prendre en compte le fait que la fréquence annuelle d'un nouveau site industriel contiendra un biais due à l'interpolation spatiale. Ainsi, il pourra être préférable d'utiliser le modèle utilisant la distinction grave/attritionnelle sur la charge du coût marginal pour lequel la charge écrêtée est modélisée par un GLM Gamma. Dans un premier temps, l'utilisation opérationnelle de ce modèle pourrait être la mise en place d'une alerte si un nouveau site RI a un coût marginal de réassurance, pour le CAT XL et le péril tempête, prédit dans la classe élevée. Il est à noter que les classes sur le coût marginal ne sont pas directement prédites par des modèles de classification puisque l'objectif à long terme est l'utilisation de la valeur de coût marginal et ainsi, le processus de prédiction est déjà mis en place. En outre, il faut également prendre en compte le fait que les coûts marginaux calculés dans l'étude sur le portefeuille actuel ne sont pas stables dans le temps en raison de la modification du profil

du portefeuille DAB Domestique de la cédante mais également en raison du changement de la prime du traité CAT XL d'une année à l'autre ainsi que de l'amélioration des modèles de RMS. Ces trois phénomènes impliquent de relancer la construction des ELT et de l'YLT. Ils peuvent également entraîner le changement d'ordre de grandeur et/ou la modification de la répartition du coût de réassurance du portefeuille RI existant. Dans ce cas, il sera également nécessaire de relancer les modèles de prédiction.

Chapitre 7

Conclusion

L'objectif principal de ce mémoire était l'obtention, en un temps de calcul raisonnable, d'un indicateur sur le prix de réassurance supplémentaire engendré par l'acquisition d'un nouveau site industriel, dans le but que celui-ci soit utilisé par la souscription RI. Pour cela, il était nécessaire de calculer dans un premier temps cet indicateur sur les sites industriels du portefeuille actuel de Generali France. L'indicateur choisi est le coût marginal de réassurance et correspond au prix de réassurance supplémentaire à payer par Generali pour l'entrée d'un nouveau site industriel en dernier dans le portefeuille RI. L'étude s'est concentrée sur la structure de réassurance des polices protégées par la garantie DAB Domestique composée d'un traité excédent de sinistre par risque, appelé DAB par risque et s'exerçant au niveau du sinistre, puis d'un traité excédent de sinistre par évènement, appelé CAT XL et s'exerçant sur la somme des sinistres concernés par un même évènement. L'approche consistait à calculer la contribution marginale des sites industriels à la prime de réassurance du traité CAT XL, pour le péril tempête.

Le calcul des coûts marginaux de réassurance des sites industriels du portefeuille existant de Generali constitue la première étape du mémoire. Le coût marginal se basant sur la prime du CAT XL pour les évènements tempête, il était nécessaire de la calculer. Le CAT XL est un traité de réassurance employé par Generali France afin de couvrir différents évènements climatiques et s'appliquant à l'ensemble des sinistres du portefeuille DAB Domestique. L'étude portant sur le péril tempête, les pertes liées à celui-ci devaient être estimées sur l'ensemble des sites associés. Toutefois, les sinistres historiques dus aux catastrophes naturelles sont peu nombreux et ne représentent pas suffisamment les risques actuels du réchauffement climatique et donc les périls de plus en plus extrêmes. Ainsi, un logiciel de modélisation CAT a été utilisé afin de modéliser les pertes du portefeuille DAB Domestique. Celui-ci se nomme RMS et fournit des tables en sortie appelées *Event Loss Table*. Néanmoins, ces dernières ne contiennent aucune information sur l'année de survenance de l'évènement. Or, la prime de réassurance du CAT XL se calcule comme une moyenne de cessions annuelles. Il a donc fallu mettre en place une méthode de simulation de type Monte Carlo afin de disposer de cette information. Cela a permis l'obtention d'une table appelée *Year Loss Table*. Deux méthodes de simulation différentes

ont été présentées : une pour laquelle la perte liée à un évènement correspondait à une perte moyenne et une pour laquelle les pertes ont été simulées par des lois Bêta. Cette dernière a été retenue puisqu'elle permettait d'obtenir des pertes beaucoup plus élevées que la première, en adéquation avec le contexte climatique actuel et certains résultats fournis par RMS. Après avoir calculé la prime de réassurance pour l'ensemble des sites du portefeuille DAB Domestique, le calcul du coût marginal d'un site industriel consiste à reproduire cette même approche mais en supprimant initialement ce site de l'ELT du produit RI récupérée sur RMS.

Le calcul des coûts marginaux sur l'ensemble des sites industriels du portefeuille actuel a permis de détenir des résultats robustes afin de mettre en place, dans un deuxième temps, un proxy de calcul du coût marginal pour le traité CAT XL et le péril tempête d'un nouveau site industriel. En effet, il n'est pas possible de reproduire le processus détaillé ci-dessus en raison du fait que la souscription RI nécessite d'avoir un indicateur de réassurance en un temps de calcul raisonnable. Or, l'utilisation de RMS ainsi que la récupération des tables en sortie du logiciel demandent un temps de calcul trop important. Pour pallier à ce problème, des algorithmes de prédiction ont été testés sur les coûts marginaux du portefeuille existant à l'aide de variables explicatives correspondant aux données renseignées par le souscripteur lors de l'acquisition d'un nouveau site industriel dans le portefeuille. Plusieurs algorithmes de régression ont été utilisés : algorithmes de Machine Learning, modélisation distincte de la charge des coûts marginaux attritionnelle et grave, modèle fréquence/coût moyen de cession de réassurance. Ces deux derniers types de modèles permettaient d'obtenir de meilleurs résultats de prédiction. Cependant, les coûts marginaux extrêmes étaient sous-estimés pour l'ensemble des modèles. Ainsi, afin d'obtenir un indicateur robuste de réassurance au niveau site, des classes ont été construites sur les coûts marginaux : faible, moyen et fort. Par conséquent, au lieu d'évaluer directement le résultat de régression, le pourcentage de sites industriels bien classés a été observé pour chaque modèle. En effet, l'information de classe pourrait être affichée sous forme d'alerte sur un des outils utilisés par la souscription RI. Le modèle le plus performant correspond au modèle fréquence/coût moyen de cession de réassurance car il permet de capter 74% des sites de la classe élevée, ce qui est satisfaisant pour un premier proxy de calcul.

Toutefois, les travaux présentent quelques limites qui pourront être améliorées à l'avenir. Premièrement, malgré que la méthode de simulation des pertes dues aux évènements tempête à l'aide de lois Bêta permet d'obtenir de meilleurs résultats sur la construction de l'YLT, celle-ci ne tient pas compte de la troncature des lois due aux conditions d'assurance en raison de la complexité de calcul. La principale limite de l'approche concerne la modification des coûts marginaux actuellement calculés sur le portefeuille existant pouvant être causée par la variation d'engagement assuré du portefeuille DAB Domestique, le changement de la prime du traité CAT XL ainsi que par l'utilisation d'une nouvelle version de RMS. En cas de survenance d'un de ces phénomènes, il sera nécessaire de reproduire l'approche de modélisation des pertes par RMS, de récupération des nouvelles

ELT ainsi que de relancer la méthode de simulation pour la construction de l'YLT sur le portefeuille DAB Domestique de Generali. Le modèle de prédiction utilisé par la souscription RI nécessitera d'être relancé seulement dans le cas où un changement significatif d'ordre de grandeur des coûts marginaux et/ou une modification de la répartition de la prime du traité CAT XL sera observée à la maille département, utilisé comme variable explicative. De plus, les modèles de prédiction sont basés sur les caractéristiques des sites industriels et sont donc moins exposés à ces phénomènes. Ces limites constituent des pistes d'amélioration possibles de la méthode mise en place dans ce mémoire.

Pour conclure, cette étude s'est concentrée uniquement sur le traité CAT XL et le péril tempête. Afin d'obtenir un indicateur global de la réassurance s'appliquant aux sites du produit RI, il serait intéressant que le même type d'approche soit effectué sur d'autres traités et/ou d'autres périls. Ainsi, la méthode développée dans ce mémoire pourrait servir de base aux approches applicables aux autres périls/traités de réassurance. En vue de la hauteur des engagements des sites MRI, il pourrait également être pertinent d'étudier leur contribution marginale au coût de réassurance du traité CAT XL pour le péril tempête. Cela nécessiterait de disposer de leur ELT à la maille site en sortie de RMS. En outre, le proxy de calcul du coût marginal de réassurance d'un nouveau site industriel mis en place, pour le traité CAT XL et le péril tempête, pourrait être utilisé opérationnellement si un grand compte, c'est-à-dire une police détenant de nombreux sites et une somme assurée très importante, est souscrit en cours d'année.

Annexes

A.1 Lois statistiques usuelles

A.1.1 Loi de Bernoulli

La loi de Bernoulli de paramètre $p \in [0,1]$ se note $\mathcal{B}(p)$ et est à valeurs dans $\{0,1\}$. Soit $X \sim \mathcal{B}(p)$, alors $\mathrm{P}(X=1) = p$ et ainsi, $\mathrm{P}(X=0) = 1-p$.

A.1.2 Loi de Poisson

Une variable aléatoire X à valeurs dans $\mathbb{N}$ suit une loi de Poisson de paramètre $\lambda > 0$ si, et seulement si :

$$\forall k \in \mathbb{N},\ \mathrm{P}(X=k) = \mathrm{e}^{-\lambda}\,\frac{\lambda^k}{k!}$$

A.1.3 Loi uniforme continue

Une variable aléatoire U à valeurs dans $[a;b]$, avec $a,\ b \in \mathbb{R}$, $a < b$, suit une loi uniforme continue si, et seulement si, sa fonction de densité de probabilité est de la forme :

$$\forall x \in \mathbb{R},\ f(x) = \begin{cases} \dfrac{1}{b-a} & \text{si } a < x < b \\ 0 & \text{sinon} \end{cases}$$

A.1.4 Loi normale

Une variable aléatoire X à valeurs dans $\mathbb{R}$ suit une loi normale (ou gaussienne) d'espérance $\mu \in \mathbb{R}$ et de variance $\sigma^2 > 0$ si, et seulement si, sa fonction de densité de probabilité est de la forme :

$$\forall x \in \mathbb{R},\ f(x;\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}}\,\mathrm{e}^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

On note ainsi $X \sim \mathcal{N}(\mu,\ \sigma^2)$.

A.1.5 Loi log-normale

Une variable aléatoire Y à valeurs dans $\mathbb{R}_+^*$ suit une loi log-normale de paramètres $\mu \in \mathbb{R}$ et $\sigma > 0$ si, et seulement si, $X = \ln(Y)$ suit une loi normale d'espérance μ et de

variance σ^2. Sa densité de probabilité s'écrit de cette façon :

$$\forall x > 0,\ f_Y(x;\mu,\sigma) = \frac{1}{x\sigma\sqrt{2\pi}}\,\mathrm{e}^{-\dfrac{\ln(x-\mu)^2}{2\sigma^2}} = \frac{1}{x}\,f_X(\ln(x)\,;\mu,\sigma)$$

A.1.6 Loi Gamma

Une variable aléatoire X à valeurs dans $\mathbb{R}_+^*$ suit une loi Gamma de paramètres α et β strictement positifs si, et seulement si, sa fonction de densité de probabilité est de la forme :

$$\forall x > 0,\ f(x;\alpha,\beta) = x^{\alpha-1}\frac{\beta^\alpha\,\mathrm{e}^{-\beta x}}{\Gamma(\alpha)}$$

où Γ désigne la fonction Gamma d'Euler.

$$\forall z \in \mathbb{C} \text{ avec } \mathrm{Re}(z) > 0,\ \Gamma(z) = \int_0^{+\infty} t^{z-1}\,\mathrm{e}^{-t}\,dt$$

A.1.7 Loi Bêta

Une variable aléatoire X à valeurs dans $[0;1]$ suit une loi Bêta de paramètres α et β strictement positifs si, et seulement si sa fonction de densité de probabilité est de la forme :

$$\forall x \in [0,1],\ f(x;\alpha,\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}x^{\alpha-1}(1-x)^{\beta-1}$$

où Γ désigne la fonction Gamma d'Euler.

A.1.8 Loi du χ^2

Supposons que $X_1, ..., X_k$, k variables aléatoires indépendantes de loi $\sim \mathcal{N}(0,\,1)$ alors :

$$X = \sum_{i=1}^{k} X_i^2$$

suit une loi du χ^2 à k degrés de liberté et est notée $\chi^2(k)$.

A.1.9 Loi de Student

Supposons $Z \sim \mathcal{N}(0,\,1)$ et $U \sim \chi^2(k)$ indépendante de Z, alors la variable aléatoire :

$$X = \frac{Z}{\sqrt{U/k}}$$

suit une loi de Student à $k > 0$ degrés de liberté.

A.2 Corrélations

A.2.1 Corrélation de Pearson

Le coefficient de corrélation de Pearson mesure la relation linéaire entre deux variables aléatoires X et Y. Il est noté $\rho_{X,Y}$ et se calcule de la façon suivante :

$$\rho_{X,Y} = \frac{\mathrm{Cov}\,(X,Y)}{\sigma_X \sigma_Y} \text{ où } \mathrm{Cov}\,(X,Y) = \mathrm{E}\,[XY] - \mathrm{E}\,[X]\,\mathrm{E}\,[Y]$$

$\rho_{X,Y} \in [-1,1]$. Une valeur proche de 1 (respectivement -1) indique une forte corrélation positive (respectivement négative) entre X et Y alors qu'une corrélation proche de 0 indique que les variables ne sont quasiment pas corrélées.

A.2.2 V de Cramér

La corrélation de Pearson, présentée ci-dessus s'applique à des variables numériques. Quant au V de Cramér $\in [0,1]$, il mesure la corrélation entre deux variables catégorielles. Il se base sur la statistique de test du χ^2 et est à valeurs dans [0,1]. Plus le résultat est élévé et plus la dépendance entre les deux variables étudiées est forte.

Soient deux variables aléatoires U et V à respectivement r et k modalités. L'échantillon d'observations est de taille n. Le nombre de fois où les valeurs (U_i, V_j) sont identifiées est noté $n_{i,j}$. La statistique de test du χ^2 est donnée par la formule :

$$\chi^2 = \sum_{i,j} \frac{(n_{i,j} - \frac{n_i n_j}{n})^2}{\frac{n_i n_j}{n}}$$

n_i (respectivement n_j) correspond au nombre de fois où la modalité U_i respectivement (V_j) est observée : $n_i = \sum_j n_{i,j}$ (respectivement $n_j = \sum_i n_{i,j}$). Le V de Cramér s'obtient en prenant la racine carré de : la statistique de test du χ^2 divisée par la taille de l'échantillon d'observations n puis par le minimum de $\{k-1, s-1\}$.

A.2.3 Matrice de corrélation du *dataset* utilisé pour la prédiction des coûts marginaux de réassurance

Les variables explicatives du *dataset* 6.1 sont toutes catégorielles et la VAE est de type numérique. Afin de calculer la corrélation entre ces variables et le coût marginal de réassurance, ce dernier est retraité en variable catégorielle pour l'application. La méthode utilisée pour le retraitement consiste à séparer les valeurs du coût marginal en huit modalités contenant chacune un même nombre d'observations. Le calcul de corrélation entre variables catégorielles s'effectue à l'aide d'un V de Cramér.

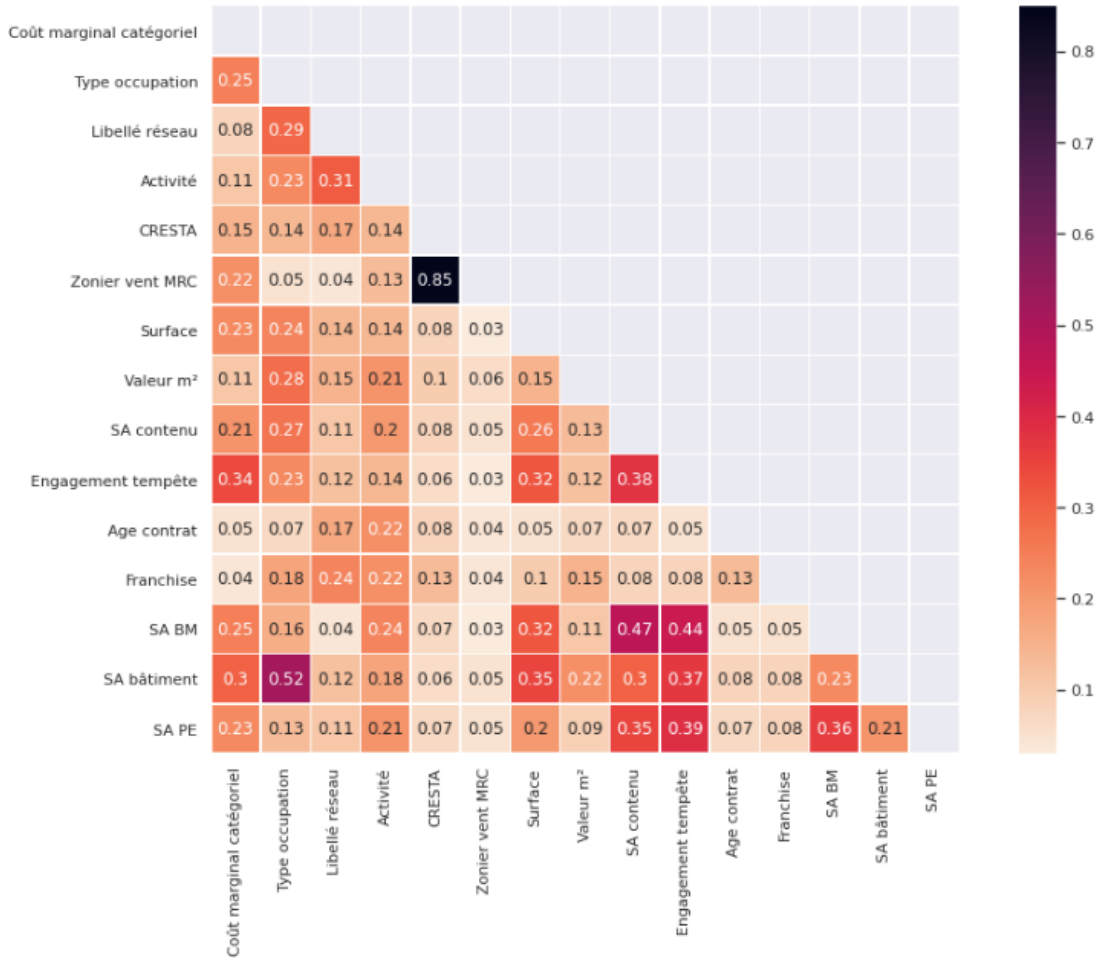


FIGURE 1 – Corrélation entre les variables explicatives et la VAE retraitée en variable catégorielle

A.3 Ajustement de lois statistiques

A.3.1 Méthode des moments

La méthode des moments consiste à calculer les paramètres d'une loi statistique à l'aide de l'estimation des moments sur un échantillon d'observations. En effet, les paramètres s'expriment généralement en fonction de la moyenne et de l'écart-type. Soit $X_1, ..., X_n$ un échantillon i.i.d. La moyenne empirique est $m_x = \frac{1}{n}\sum_{i=1}^{n} X_i$ et l'écart-type empirique sans biais correspond à $\sigma_x = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - m_x)^2}$.

A.3.2 Maximum de vraisemblance

La méthode du maximum de vraisemblance est une autre approche permettant d'estimer les paramètres d'une loi statistique ajustée à un échantillon d'observations $X_1, ..., X_n$ i.i.d. Soit F_θ la loi de l'échantillon et θ le paramètre de loi à estimer. On note :

$$f(x;\theta) = \begin{cases} P_\theta(X = x) & \text{si } X \text{ est de loi discrète} \\ f_\theta(x) & \text{sinon, avec } f_\theta \text{ la fonction de densité} \end{cases}$$

La fonction de vraisemblance se note :

$$L(x_1, ..., x_n; \theta) = \prod_{i=1}^{n} f(x_i; \theta)$$

Ainsi, le paramètre $\hat{\theta}$ est estimé en maximisant la fonction de vraisemblance L par rapport au paramètre θ.

A.4 Détermination du seuil grave sur les coûts moyens de cession de réassurance

Les graphiques 2 et 3 nous permettent de choisir le seuil d'écrêtement des coûts moyens de cession de réassurance, à l'aide des méthodes classiques de détermination de la théorie des valeurs extrêmes.

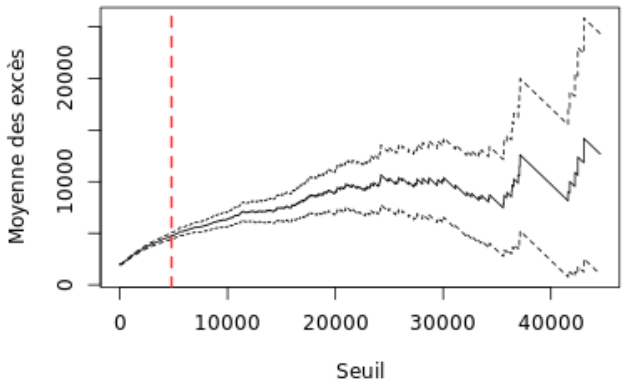


FIGURE 2 – Fonction moyenne des excès en fonction du seuil sur les coûts moyens

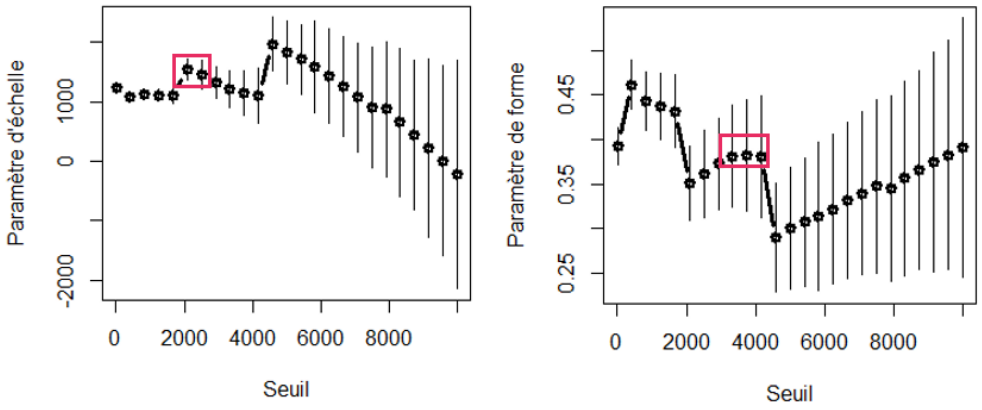


FIGURE 3 – Paramètres d'une loi GPD en fonction du seuil sur les coûts moyens

De plus, la qualité d'ajustement de lois Gamma et log-normale à la distribution des coûts moyens de cession de réassurance inférieurs au seuil de 4 800 confirme ce choix de seuil, d'après la figure 4.

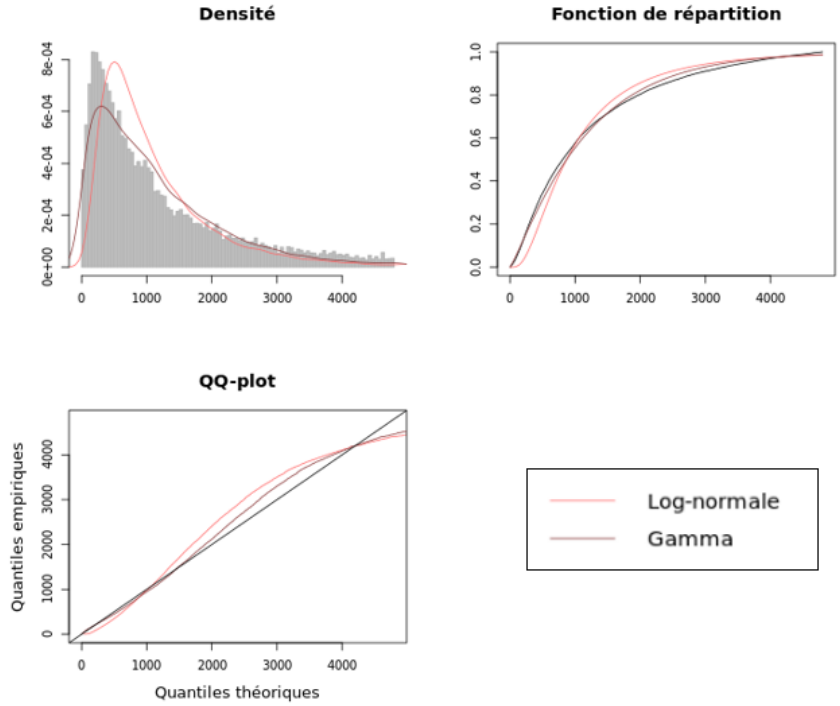


FIGURE 4 – Ajustement de lois log-normale et Gamma à la distribution des coûts moyens de cession de réassurance inférieurs à 4 800

Table des figures

1.1 Structure de réassurance 2023 de Generali France pour le produit RI . . . 2

2.1 Les différents types de traités de réassurance obligatoire 5
2.2 Exemple de traités excédent de sinistre par risque et excédent de sinistre par évènement 6
2.3 Structure de réassurance des risques climatiques du portefeuille DAB Domestique 10
2.4 Tranches du traité excédent de sinistre par risque, le DAB par risque . . . 11
2.5 Tranches du traité excédent de sinistre par évènement, le CAT XL 11

3.1 Schéma de modélisation des catastrophes naturelles 15
3.2 Principaux évènements tempête historiques en France 16
3.3 Courbes de vulnérabilité pour la tempête 17
3.4 Courbes OEP et AEP du portefeuille DAB Domestique calculées par RMS pour la tempête 18
3.5 Densité de lois Bêta 19
3.6 Engagement assuré tempête par département du produit RI 21

4.1 Histogramme des pertes moyennes des sites RI modélisées par RMS supérieures à 200k€ 32
4.2 Erreur relative absolue en fonction du nombre de simulations pour la méthode de simulation des pertes sans incertitude secondaire 39
4.3 Erreur relative absolue en fonction du nombre de simulations pour la méthode de simulation des pertes avec incertitude secondaire 40

5.1 Histogramme des coûts marginaux de réassurance des sites RI 47
5.2 Histogramme des coûts marginaux de réassurance des sites RI supérieurs à 1 000 € 47
5.3 Coûts marginaux de réassurance rapportés à l'engagement assuré par département pour le portefeuille RI 48

6.1 QQ-plot des coûts marginaux de réassurance 59
6.2 Fonction moyenne des excès en fonction du seuil sur les coûts marginaux . 60
6.3 Paramètres d'une loi GPD en fonction du seuil sur les coûts marginaux . . 60

6.4 Ajustement de lois log-normale et Gamma à la distribution des coûts marginaux de réassurance inférieurs à 900 . . . 61
6.5 Prédictions par un GLM Gamma des coûts marginaux écrêtés en fonction des observations . . . 63
6.6 Prédictions par un GLM log-normal des coûts marginaux écrêtés en fonction des observations . . . 63
6.7 Régression logistique pour les coûts marginaux : taux d'erreur en fonction du seuil . . . 63
6.8 Arbre de décision modélisant le sur-écrêtement des coûts marginaux . . . 64
6.9 Prédictions des coûts marginaux totaux en fonction des observations pour le modèle avec GLM Gamma . . . 65
6.10 Prédictions des coûts marginaux totaux en fonction des observations pour le modèle avec GLM log-normal . . . 65
6.11 *Random Forest* : valeur SHAP de l'engagement assuré . . . 66
6.12 *Random Forest* : valeur SHAP du zonier vent MRC . . . 66
6.13 Prédictions des coûts marginaux en fonction des observations pour le *Random Forest* . . . 67
6.14 Prédictions des coûts marginaux en fonction des observations pour le *XGBoost* . . . 67
6.15 Histogramme de la fréquence annuelle de contribution au dépassement de la priorité du CAT XL des sites RI pour les évènements tempête . . . 68
6.16 Grille de résolution de 20 kilomètres sur la France métropolitaine . . . 70
6.17 Fréquences annuelles de contribution au dépassement de la priorité du CAT XL des sites RI interpolées . . . 71
6.18 QQ-plot des coûts moyens de cession de réassurance . . . 72
6.19 Prédictions par un GLM Gamma des coûts moyens écrêtés en fonction des observations . . . 74
6.20 Prédictions par un GLM log-normal des coûts moyens écrêtés en fonction des observations . . . 74
6.21 Régression logistique pour les coûts moyens : taux d'erreur en fonction du seuil . . . 74
6.22 Arbre de décision modélisant le sur-écrêtement des coûts moyens . . . 75
6.23 Prédictions des coûts marginaux totaux en fonction des observations pour le modèle fréquence/coût avec GLM Gamma . . . 76
6.24 Prédictions des coûts marginaux totaux en fonction des observations pour le modèle fréquence/coût avec GLM log-normal . . . 76

1 Corrélation entre les variables explicatives et la VAE retraitée en variable catégorielle . . . 88
2 Fonction moyenne des excès en fonction du seuil sur les coûts moyens . . . 89
3 Paramètres d'une loi GPD en fonction du seuil sur les coûts moyens . . . 90
4 Ajustement de lois log-normale et Gamma à la distribution des coûts moyens de cession de réassurance inférieurs à 4 800 . . . 90

Liste des tableaux

2.1 Contribution des périls et produits aux cessions historiques du CAT XL . 12
2.2 Montant moyen des sinistres historiques tempête 12

3.1 Taux de remplissage des variables pour la modélisation RMS sur les sites du portefeuille RI 20
3.2 Précision de géocodage des sites du portefeuille RI avant et après suppression des sites mal géocodés 21
3.3 Exemple d'ELT au niveau site 22
3.4 ELT 3.3 au niveau portefeuille 24

4.1 Exemple d'ELT au niveau portefeuille du produit MRH 26
4.2 Exemple d'ELT au niveau portefeuille du produit MRC 26
4.3 Regroupement au niveau portefeuille des ELT MRH 4.1 et MRC 4.2 . . . 27
4.4 Premières lignes de l'ELT au niveau portefeuille de la garantie DAB Domestique après agrégation des sept produits 27
4.5 Montant maximal des sinistres historiques tempête 28
4.6 Nombre d'évènements tempête des cinq premières années de simulation tirés dans une loi de Poisson 29
4.7 Exemple d'ELT au niveau portefeuille pour la garantie DAB Domestique . 29
4.8 Probabilités conditionnelles de survenance des évènements de l'ELT 4.7 . . 30
4.9 Tirage aléatoire dans les évènements de l'ELT 4.8 31
4.10 Attribution des évènements tempête aux années de simulation pour l'exemple d'ELT 4.8 31
4.11 Evènements de l'ELT-DABDOM survenant durant les deux premières années de simulation 31
4.12 Premières lignes de l'YLT au niveau portefeuille pour la garantie DAB Domestique sans incertitude secondaire 32
4.13 Cinq cessions annuelles de l'YLT 4.12 au traité de réassurance CAT XL . 33
4.14 Premières lignes de l'YLT au niveau portefeuille pour la garantie DAB Domestique avec incertitude secondaire 34
4.15 Paramètres de lois Bêta des sites RI 35
4.16 Cinq cessions annuelles de l'YLT 4.14 au traité de réassurance CAT XL . 36

4.17 Pourcentage d'erreur sur l'OEP 250Y inférieur à 1% et 2% pour un nombre de simulations différent 41
4.18 Ecart entre la prime de réassurance de marché et la prime issue d'un logiciel de modélisation CAT 42

5.1 Premières lignes de l'YLT au niveau portefeuille pour la garantie DAB Domestique privé du site RI n°1 46

6.1 *Dataset* pour la prédiction des coûts marginaux de réassurance des sites RI 51
6.2 Métriques d'évaluation de la régression logistique pour les coûts marginaux 64
6.3 RMSE et corrélation de Spearman du coût marginal de réassurance sur la base de test avec séparation des graves et attritionnels 64
6.4 RMSE et corrélation de Spearman du coût marginal de réassurance sur la base de test pour le *Random Forest* et le *XGBoost* 66
6.5 Métriques d'évaluation de la régression logistique pour les coûts moyens . 75
6.6 RMSE et corrélation de Spearman du coût marginal de réassurance sur la base de test pour le modèle fréquence/coût moyen 75
6.7 Définition des classes sur le coût marginal de réassurance pour le traité CAT XL et le péril tempête 77
6.8 Pourcentage de sites bien classés dans les prédictions du coût marginal de réassurance sur la base de test 77
6.9 Exemple d'une région touchée par un évènement tempête et contenant trois sites assurés RI 78
6.10 Exemple d'une région touchée par un évènement tempête et contenant quatre sites assurés RI 79
6.11 Exemple de calcul des coûts marginaux de réassurance de quatre sites RI . 79

Bibliographie

[RMS, 2019] (2019). https://www.rms.com/. Documentation RMS.

[Ailliot, 2021] Ailliot, P. (2021). Modèle linéaire, M1 EURIA. Université de Bretagne Occidentale, Année académique 2021–2022.

[Bernegger, 1997] Bernegger, S. (1997). The Swiss Re exposure curves and the MBBEFD distribution class. *Astin Bulletin*, 27(1):99–111.

[Berthet, 2009] Berthet, L. (2009). Polluted site assessment using Inverse Distance Weighted and Ordinary Kriging. page 6.

[Braham, 2022] Braham, G. A. (2022). Réassurance. Université de Bretagne Occidentale, Année académique 2022–2023.

[Buntun, 2016] Buntun, B. (2016). Property excess of loss reinsurance pricing in the Lloyd's market. *Mémoire d'actuariat*, pages 24–61.

[Chen et Guestrin, 2016] Chen, T. et Guestrin, C. (2016). XGBoost : A Scalable Tree Boosting System. *arXiv :1603.02754*.

[Coles, 2001] Coles, S. (2001). *An introduction to statistical modeling of extreme values*. Springer.

[de la Cruz, 2009] de la Cruz, J. (2009). Les risques événements naturels tempêtes - Couvertures de réassurances traditionnelles et non traditionnelles - Application sur un portefeuille d'assurance. *Mémoire d'actuariat*, pages 93–106.

[Dehoux et Liao, 2019] Dehoux, A. et Liao, C. (2019). Shap : Mieux comprendre l'interprétation de modèle. https://www.aquiladata.fr/insights/shap-mieux-comprendre-linterpretation-de-modeles/.

[Guggisberg, 2004] Guggisberg, D. (2004). Exposure rating. *Swiss Re*.

[Homer et Li, 2017] Homer, D. et Li, M. (2017). Notes on Using Property Catastrophe Model Results. *Casualty Actuarial Society*, 2:2–10.

[Lair, 2011] Lair, F. (2011). Gestion du risque catastrophes naturelles en assurance d'entreprises. *Mémoire d'actuariat*, pages 50–61.

[Sermet, 2018] Sermet, T. (2018). Allocation du chargement de réassurance et du coût du capital à la tarification des catastrophes naturelles. *Mémoire d'actuariat*, pages 41–55.

[Thérond, 2012] Thérond, P. (2012). Théorie des valeurs extrêmes. *L'actuariel*, (5).

[Vermet, 2020] Vermet, F. (2020). Régression linéaire. Université de Bretagne Occidentale, Année académique 2020–2021.

[Vermet, 2021] Vermet, F. (2021). Arbres de décision et méthodes ensemblistes. Université de Bretagne Occidentale, Année académique 2021–2022.

[Vuong, 2011] Vuong, C. (2011). La tempête européenne : estimation du besoin de capacité de réassurance et aide à la souscription. *Mémoire d'actuariat*, pages 28–34.

[Wang, 2012] Wang, Z. (2012). Tarification des couvertures de réassurance des catastrophes naturelles en excédent de sinistre par risque, comparaison des approches à l'exposition et à l'expérience. *Mémoire d'actuariat*, pages 94–102.