

Mémoire présenté devant l'ENSAE Paris
pour l'obtention du diplôme de la filière Actuariat
et l'admission à l'Institut des Actuaire
le 07/03/2025

Par : **Willy MARAIS**

Titre : **Risque grêle : construction d'un zonier et proposition
d'une politique de souscription responsable**

Confidentialité : ☒ NON ☐ OUI (Durée : ☐ 1 an ☐ 2 ans)

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

Membres présents du jury de la filière

Entreprise : Exiom Partners

Nom : Antoine HERANVAL

Signature :



*Membres présents du jury de l'Institut
des Actuaire*

Directeur du mémoire en entreprise :

Nom : Maylis KOUEK

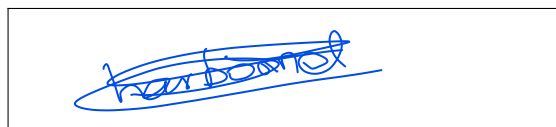
Signature :



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

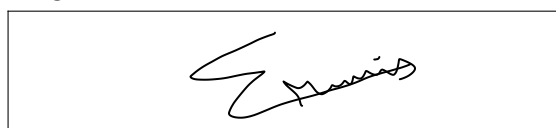
Secrétariat :

Signature du responsable entreprise



Bibliothèque :

Signature du candidat



Résumé

Phénomène météorologique auquel la France est de plus en plus exposée, la grêle est un risque dont l'anticipation et la gestion sont devenues indispensables pour les assureurs. En effet, l'année 2022 a enregistré plus d'un million de sinistres liés à la grêle dont les dégâts sont estimés à plus de 4,8 milliards d'euros. Ce mémoire a pour objectif de construire un zonier grêle pour la région Auvergne-Rhône-Alpes, afin de tarifier une garantie dans l'assurance MRH et de proposer une politique de souscription responsable et engagée.

La première partie examine le phénomène naturel de la grêle, ses impacts croissants liés au dérèglement climatique et les couvertures assurantielles existantes. La deuxième partie présente trois approches de conception d'un zonier : un procédé itératif, l'utilisation de modèles linéaires généralisés et le recours à la théorie de la crédibilité. Ces outils permettent d'attribuer des niveaux de risque précis aux communes, en intégrant la fréquence et l'intensité des événements de grêle ainsi que les caractéristiques des habitations. Enfin, la dernière partie propose une politique de souscription responsable qui tient compte des évolutions climatiques et réglementaires auxquelles les organismes assurantiels pourraient être confrontés. Le but étant de suggérer une démarche qui garantit une tarification équitable et accessible.

Ce travail présente une approche méthodologique innovante sur le risque grêle, qui permet d'appréhender l'exposition d'un assureur à ce risque dans le cadre d'une garantie dans l'assurance MRH. Il souligne également l'importance de l'engagement des acteurs assurantiels dans la construction d'une résilience collective et durable face aux aléas climatiques.

Mots-clés : grêle, zonier, théorie de la crédibilité, modèle de Bühlmann-Straub, garantie Tempête/Grêle/Neige, politique de souscription, tarification, rapport Langreny, GLM, dérèglement climatique

Abstract

Hail is a meteorological phenomenon to which France is increasingly exposed. Its anticipation and management have become essential for insurers. Indeed, 2022 recorded over one million hail-related claims, with damages estimated at more than €4.8 billion. This study aims to price hail coverage in home insurance for the Auvergne-Rhône-Alpes region through a geographical-zone design and to propose a responsible and committed underwriting policy.

The first part examines the natural phenomenon of hail, its growing impacts due to climate change, and existing insurance coverages. The second part presents three zone-based approaches : an iterative process, the use of generalized linear models, and the application of credibility theory. These tools make it possible to assign precise risk levels to towns, incorporating the frequency and severity of hail events as well as the characteristics of housing. Finally, the last part proposes a responsible underwriting policy that takes into account the climatic and regulatory changes that insurance organizations may face. The goal is to suggest an approach that ensures fair and accessible pricing.

This work introduces an innovative methodological approach to hail risk, enabling insurers to assess their exposure to this risk within the framework of home insurance coverage. It also highlights the importance of insurers' commitment to building collective and sustainable resilience to climate hazards.

Keywords : hail, zone-based approach, credibility theory, Bühlmann-Straub model, Storm/Hail/Snow coverage, underwriting policy, pricing, Langreny report, GLM, climate change

Note de synthèse

Partie 1 : La grêle : du phénomène naturel aux solutions assurantielles

Le changement climatique constitue un défi majeur pour le secteur assurantiel, en raison de l'intensification et de la multiplication des événements climatiques extrêmes, tels que les épisodes de grêle. En France, cette évolution s'est traduite en 2022 par une année exceptionnelle : plus d'un million de sinistres liés à la grêle ont été enregistrés, pour un coût total estimé à 4,5 milliards d'euros, dont 3,3 milliards pris en charge par la garantie TGN.

Malgré l'ampleur de ces chiffres, le risque grêle reste sous-exploré dans le domaine assurantiel. Il est souvent éclipsé par d'autres aléas climatiques plus fréquents ou plus coûteux, et les données disponibles sur ce phénomène demeurent limitées. Cependant, les projections climatiques mettent en évidence une augmentation de l'intensité des épisodes de grêle, nécessitant une adaptation urgente des pratiques assurantielles pour préserver la viabilité économique des contrats, tout en répondant aux attentes des assurés.

Ce mémoire s'inscrit dans cette perspective et se concentre sur la modélisation du risque grêle dans le cadre de la garantie MRH, en particulier dans une zone fortement exposée : la région Auvergne-Rhône-Alpes. Les travaux visent à créer un zonier permettant d'évaluer le risque grêle à l'échelle communale, ainsi qu'élaborer une politique de souscription équitable qui tienne compte des disparités régionales et des caractéristiques des biens assurés.

Partie 2 : Construction d'un zonier grêle

Pour répondre à ces objectifs, la première étape consiste en la construction d'un zonier dédié au risque grêle. Ce travail s'appuie sur le retraitement et l'analyse de trois bases de données mises à disposition, couvrant la période de 2019 à 2023. Ces bases incluent des informations sur le portefeuille d'assurés, les sinistres déclarés et les événements climatiques recensés, offrant ainsi une vue d'ensemble complète et détaillée pour modéliser le risque.

À partir de ces bases de données, trois approches ont été étudiées pour construire le zonier. Ce dernier correspond à l'élaboration d'un score de risque, représenté graphiquement en tenant compte des contraintes géographiques. Dans le cadre de ce mémoire, le score est basé sur les primes pures et permet une classification des zones selon leur niveau d'exposition.

Première approche : construction d'un zonier par itération

La première méthode de construction consiste à élaborer séparément :

- **Un zonier intensité**, basé sur la taille maximale des grêlons observée par commune, en tenant compte de l'écart entre les coûts moyens observés et ceux prédits par régression.
- **Un zonier fréquence**, calculé à partir du nombre d'événements de grêle enregistrés annuellement, en distinguant ceux ayant généré des sinistres.

Ces deux zoniers sont ensuite combinés à l'aide de la grille de croisement utilisée par la MRN.

Bien que cette approche soit relativement simple à mettre en œuvre et compréhensible pour des acteurs non experts, elle présente plusieurs limites significatives telles que l'intégration réduite des spécificités du portefeuille, le manque de robustesse scientifique, et le peu de fiabilité des résultats. Pour ces raisons, cette méthode n'a pas été retenue pour la suite de l'étude.

Deuxième approche : construction d'un zonier par GLM

La deuxième approche se base sur l'utilisation des GLM pour modéliser la fréquence et le coût moyen des sinistres. Ces modèles utilisent des variables explicatives telles que la fréquence et l'intensité des épisodes de grêle, ainsi que des caractéristiques des habitations, en particulier le nombre de pièces. Une fonction de lien logarithmique est choisie pour exprimer les relations multiplicatives entre les variables, facilitant l'interprétation des résultats et l'ajustement des primes.

Trois bases de données sont structurées selon différents niveaux de détail :

- **Base 1** : Une ligne par commune (4 500 lignes).
- **Base 2** : Une ligne par commune et par année (22 000 lignes).
- **Base 3** : Une ligne par commune, année et nombre de pièces (170 000 lignes).

Modélisation de la fréquence des sinistres

La fréquence des sinistres est modélisée à l'aide d'une distribution binomiale négative, en raison de la sur-dispersion des données. Différents modèles sont évalués à l'aide des p-valeurs des coefficients, du R^2 , de la log-vraisemblance, du critère d'Akaike (AIC) et du F-score.

Le modèle retenu est celui utilisant la base 1, avec le nombre d'assurés et le nombre d'événements par classe de grêlons (nombre d'événements A1, nombre d'événements A2, etc.) comme variables explicatives, et une distribution binomiale négative. Les indicateurs obtenus pour ce modèle sont les suivants :

Base de données	Variables explicatives	Distribution	P-valeur des coefficients	R^2	Log-vraisemblance	AIC
Base 1	- Nombre d'assurés - Nombre d'événements par intensité	Négative binomiale	Toutes < 0.1%	62%	-11328.60	22671

De plus, les résultats obtenus pour le F-score se retrouve dans le tableau suivant :

Base de données	Distribution	F-score global	F-score pas de sinistre	F-score un sinistre	F-score plus d'un sinistre
Base 1	Négative binomiale	0.517	0.59	0.622	0.277

Modélisation du coût moyen

Un travail similaire est à présent réalisé pour modéliser le coût moyen des sinistres. Celui-ci est ajusté à l'aide d'une distribution Gamma, adaptée pour représenter la nature asymétrique des données, marquées par une queue longue caractéristique des sinistres élevés. A nouveau, le modèle retenu utilise la base 1, combinée à cette distribution, avec le nombre d'événements par classe comme variables explicatives principales.

Base de données	Variables explicatives	Distribution	P-valeur par coefficient	R^2	F-score global
Base 1	Nombre d'événements et intensité	Gamma	A1 : 0.723 A2 : 0.129 A3 : 0.232 A4 : < 0.1% A5 : < 0.1%	4.4%	0.466

Cependant, les indicateurs de ce modèle sont insatisfaisants. En effet, les p-valeurs élevées pour certains coefficients traduisent une faible significativité statistique, et le coefficient de détermination R^2 , relativement faible, reflète une capacité limitée du modèle à expliquer la variabilité des données. Ces résultats étaient assez attendus compte tenu de la complexité à prédire le coût moyen, avec un modèle relativement simple.

Ajustement du scoring : classification et lissage

Une fois les primes pures calculées pour chaque commune, une étape de classification est réalisée afin d'attribuer un niveau de risque (faible, modéré et élevé) à chacune d'elles.

Ce zonier est ensuite amélioré par un lissage géographique, qui réduit la dispersion des zones de risque.

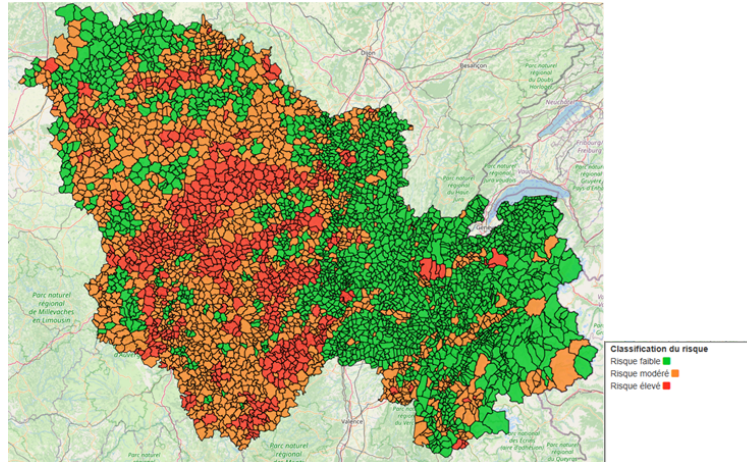


FIGURE 1 – Zonier par GLM de l'exposition de l'assureur MRH au risque grêle pour la région Auvergne-Rhône-Alpes après lissage

Le modèle GLM présente plusieurs limites, notamment un nombre restreint de variables explicatives, ce qui empêche de capter toute la complexité des phénomènes. De plus, le faible R^2 reflète une explication limitée de la variabilité des données par le modèle. Cela suggère donc l'utilisation d'une méthode plus adaptée comme la théorie de la crédibilité.

Troisième approche de zonier : théorie de la crédibilité

Le zonier GLM est challengé par la méthode de la théorie de la crédibilité, avec notamment le modèle de Bühlmann-Straub. Ce modèle repose sur une méthodologie qui combine les données globales du portefeuille avec celles spécifiques à des sous-groupes homogènes d'assurés, en fonction de caractéristiques similaires. Ce modèle permet de calculer la prime pure en agrégeant la fréquence et le coût moyen des sinistres, pour ensuite établir un scoring des risques. Le modèle de Bühlmann-Straub attribue un poids optimal entre les moyennes globales (au niveau du portefeuille) et les moyennes spécifiques (au niveau des sous-groupes), déterminé par le paramètre de crédibilité α , qui dépend de la variance entre groupes et de la variance intra-groupe. Cette approche permet de produire des estimations plus fiables, indépendamment de la quantité de données fournies par sous-groupes.

Modélisation de la fréquence des sinistres

Pour utiliser le modèle de Bühlmann-Straub, des groupes homogènes en fréquence à l'intérieur, et différents les uns des autres, sont créés en deux étapes :

- Regroupement des communes par risque lié aux événements de grêle
- Croisement avec le nombre de pièces des habitations

Cela permet de créer 12 groupes, sur lesquels le modèle de Bühlmann-Straub est appliqué, permettant d'obtenir les résultats suivants :

Types d'événements \ Nb pièces	1 à 3 pièces	4 à 7 pièces	8 pièces ou plus
Pas d'événements	0,092%	0,101%	0,243%
Basse fréquence	0,291%	0,480%	0,670%
Moyenne fréquence	0,659%	1,303%	1,427%
Haute fréquence	3,509%	7,175%	4,654%

FIGURE 2 – Fréquence prédite par le modèle de Bühlmann-Straub en fonction du nombre de pièces et du nombre d'événements de grêle subis

Modélisation du coût moyen d'un sinistre

Le coût moyen est modélisé à partir de deux critères :

- Départements
- Nombre de pièces

Cela permet de créer 9 groupes, sur lesquels le modèle de Bühlmann-Straub est appliqué, permettant d'obtenir les résultats suivants :

Départements\Nb pièces	1 à 4 pièces	5 à 7 pièces	8 pièces ou plus
1, 18, 38, 69	4 195,03€	4 464,88€	5 146,84€
42, 43, 71, 73	4 680,05€	5 509,19€	6 300,12€
3, 58, 63, 74	8 485,01€	9 119,58€	7 316,64€

FIGURE 3 – Cout moyen prédit par Bühlmann-Straub en fonction du département et du nombre de pièces

Construction et classification du zonier

En combinant la fréquence et le coût moyen, une prime pure est calculée pour chaque assuré, puis agrégée au niveau communal. Les communes sont classées en trois catégories de risque exactement de la même manière que pour le zonier fait avec les GLM. Puis le lissage géographique est appliqué de la même façon, permettant d'obtenir le zonier suivant :

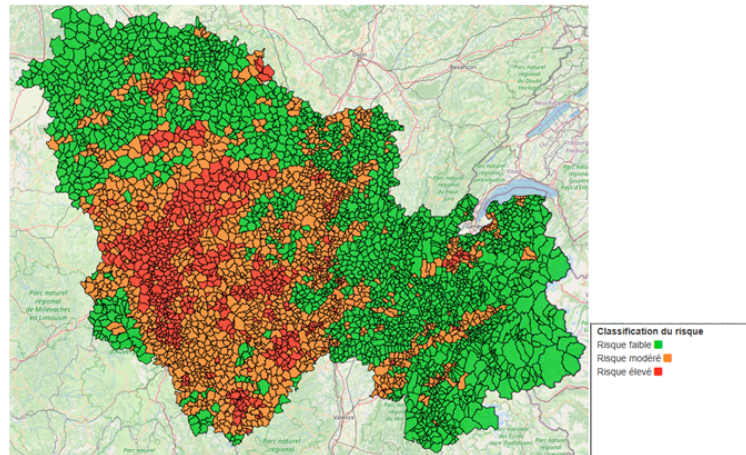


FIGURE 4 – Zonier après lissage basé sur la théorie de la crédibilité de l'exposition de l'assureur MRH au risque grêle pour la région Auvergne-Rhône-Alpes

La théorie de la crédibilité présente certaines limites malgré ses avantages. Tout d'abord, cette méthode peut paraître abstraite et nécessite une bonne maîtrise des concepts sous-jacents pour une mise en œuvre correcte. De plus, la qualité des résultats dépend en grande partie de la pertinence des groupes de risques définis en amont, rendant cette étape cruciale.

Comparaison des méthodes de construction de zonier

Les trois méthodes explorées présentent des avantages et des limites. Cependant, le zonier lissé issu de la théorie de la crédibilité est conservé pour plusieurs raisons, dont :

- L'adéquation du recours à la théorie de la crédibilité dans le cadre d'études ayant une faible quantité de données
- L'homogénéité de la cartographie obtenue
- La cohérence des zones identifiées comme étant à risque avec celles ayant subies des épisodes de grêles intenses comme le montrent les cartes ci-dessous :

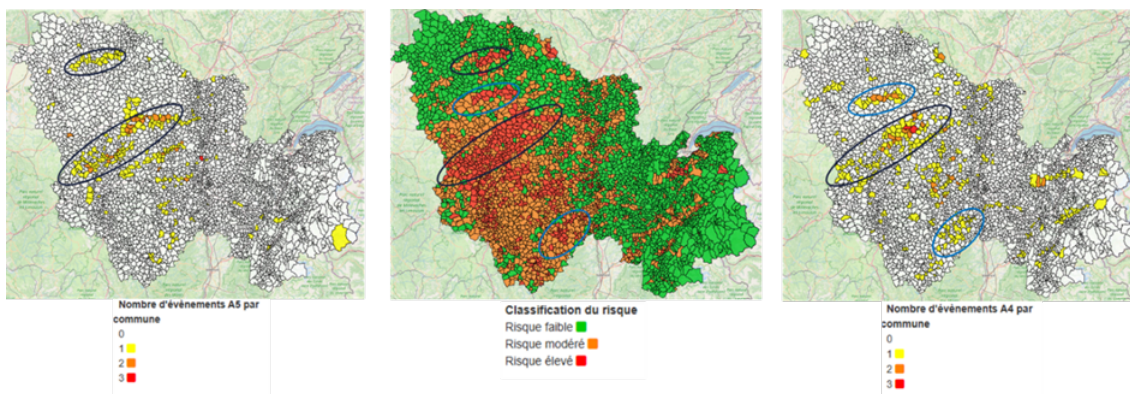


FIGURE 5 – Cartes du nombre d'évènements A5 à gauche, A4 à droite et le zonier par théorie de la crédibilité de l'exposition de l'assureur MRH au risque grêle pour la région Auvergne-Rhône-Alpes au centre

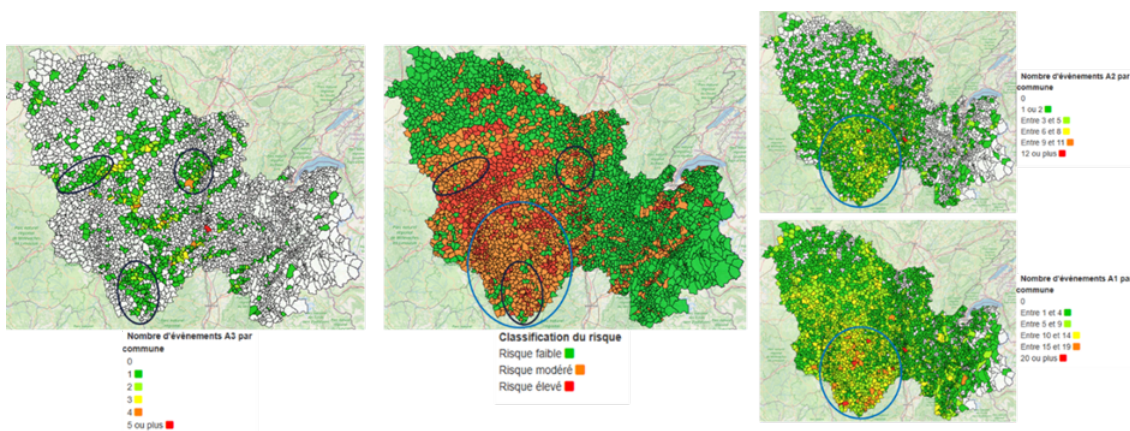


FIGURE 6 – Cartes du nombre d'évènements A3 à gauche, A2 en haut à droite, A1 en bas à droite et le zonier par théorie de la crédibilité de l'exposition de l'assureur MRH au risque grêle pour la région Auvergne-Rhône-Alpes au centre

Partie 3 : Définition d'une politique de souscription à partir du zonier

La construction du zonier permet de définir une politique de souscription durable et juste. Cette mise en application permet concrètement :

- De construire un tarif basé sur des primes pures ajustées, reflétant précisément le risque par zone.
- De mettre en évidence les disparités ou injustices éventuelles dans la répartition des primes et des risques, offrant une vision claire des ajustements nécessaires.
- Anticiper les évolutions climatiques et réglementaires à travers des scénarios prospectifs.

Dans un premier temps, une grille tarifaire est calibrée en se basant sur le zonier, de manière à garantir l'équilibre actuariel. La grille tarifaire se base sur deux variables : la zone de risque et le nombre de pièces du logement. Un coefficient α ajuste les primes en fonction de la taille des logements, garantissant que les primes couvrent exactement les coûts des sinistres. La grille tarifaire obtenue est la suivante :

Zone de risque Nombre de pièces	Bas	Moyen	Haut
1	4,28€	27,77€	141,40€
2	5,43€	33,37€	163,83€
3	6,58€	38,97€	186,26€
4	7,73€	44,57€	208,69€
5	8,88€	50,17€	231,12€
6	10,03€	55,77€	253,55€
7	11,18€	61,37€	275,98€
8	12,33€	66,97€	298,41€
9	13,48€	72,57€	320,84€
10	14,63€	78,17€	343,27€
11	15,78€	83,77€	365,70€
12	16,93€	89,37€	388,13€
13	18,08€	94,97€	410,56€
14	19,23€	100,57€	432,99€
15	20,38€	106,17€	455,42€

Les primes varient considérablement entre les zones, avec des différences marquées dans les zones à risque élevé. Par exemple, la garantie MRH d'une maison de 10 pièces coûte trente fois plus dans une zone rouge que dans une zone verte. Cette tarification engendre une disparité entre assurés des différentes zones, ce qui pourrait freiner les assurés à la souscription et contribuer au désengagement des assureurs. De plus, avec l'intensification des risques climatiques, la sinistralité devrait se dégrader dans les zones exposées et par conséquent accentuer les inégalités observées et les potentiels désengagements.

Compte tenu de ces constats, il est pertinent de définir une politique de souscription durable et juste. Concrètement, cela revient à tenir compte de l'augmentation de la sinistralité liée au dérèglement climatique attendu, et à proposer des solutions pour inciter l'assureur à s'engager dans les zones à risque.

Adaptation de la grille tarifaire face au dérèglement climatique

Le dérèglement climatique pourrait modifier la fréquence et l'intensité des événements de grêle, entraînant une augmentation des sinistres. Une analyse du groupe Covéa prévoit une augmentation de 40% des orages de grêle d'ici 2050, entraînant une hausse de 20% des sinistres.

Ainsi, pour maintenir l'équilibre, les primes estimées devront augmenter progressivement de 20% d'ici 2050, ajustées sur une croissance annuelle linéaire de 0,92%. Cependant cette augmentation annuelle est proposée que dans 5 ans, afin de prendre en compte le caractère exceptionnel de l'année 2022. Un ajustement du zonier sera également nécessaire pour tenir compte du déplacement des zones les plus exposées.

Adaptation de la grille tarifaire face à des changements réglementaires

Des évolutions réglementaires sont proposées dans le rapport Langrenay (2024) afin de renforcer l'assurabilité des risques climatiques. Les travaux de ce mémoire s'attardent sur une mesure en particulier visant à réduire la discrimination géographique dans les zones à haut risque tout en garantissant une couverture équitable. Celle-ci consiste à moduler la surprime Cat Nat en fonction de l'exposition géographique. L'impact de cette surprime modulée sur la prime pure grêle, comparée à une surprime uniforme de 20% (surprime destinée à financer le régime d'indemnisation des catastrophes naturelles) appliquée à toutes les zones est présenté ci-dessous :

Zones de risque / Nombre de pièces	Bas		Moyen		Haut	
	Prime pure grêle avec surprime Cat Nat uniforme de 20%	Prime pure grêle avec surprime Cat Nat modulée	Prime pure grêle avec surprime Cat Nat uniforme de 20%	Prime pure grêle avec surprime Cat Nat modulée	Prime pure grêle avec surprime Cat Nat uniforme de 20%	Prime pure grêle avec surprime Cat Nat modulée
1	5,14 €	8,56 €	33,32 €	30,55 €	169,68 €	144,23 €
2	6,52 €	10,86 €	40,04 €	36,71 €	196,60 €	167,11 €
3	7,90 €	13,16 €	46,76 €	42,87 €	223,51 €	189,99 €
4	9,28 €	15,46 €	53,48 €	49,03 €	250,43 €	212,86 €
5	10,66 €	17,76 €	60,20 €	55,19 €	277,34 €	235,74 €
6	12,04 €	20,06 €	66,92 €	61,35 €	304,26 €	258,62 €
7	13,42 €	22,36 €	73,64 €	67,51 €	331,18 €	281,50 €
8	14,80 €	24,66 €	80,36 €	73,67 €	358,09 €	304,38 €
9	16,18 €	26,96 €	87,08 €	79,83 €	385,01 €	327,26 €
10	17,56 €	29,26 €	93,80 €	85,99 €	411,92 €	350,14 €
11	18,94 €	31,56 €	100,52 €	92,15 €	438,84 €	373,01 €
12	20,32 €	33,86 €	107,24 €	98,31 €	465,76 €	395,89 €
13	21,70 €	36,16 €	113,96 €	104,47 €	492,67 €	418,77 €
14	23,08 €	38,46 €	120,68 €	110,63 €	519,59 €	441,65 €
15	24,46 €	40,76 €	127,40 €	116,79 €	546,50 €	464,53 €

FIGURE 7 – Grille tarifaire avec surprime Cat Nat uniforme de 20% ou modulée

Le tableau révèle une augmentation de 40% entre la prime Cat Nat uniforme et celle modulée pour la zone verte, une diminution de 9% pour la zone orange, et une diminution de 18% dans la zone rouge. Ces ajustements rendent la couverture plus viable en allégeant la charge financière des assurés les plus exposés. Toutefois, un écart significatif demeure entre les zones vertes et rouges, atteignant un rapport de 1 à 15 environ. Cela souligne la nécessité d'intégrer d'autres recommandations du rapport Langrenay, ainsi que des choix stratégiques des assureurs, comme des rééquilibrages entre différents types d'assurances.

Ainsi l'intégration des recommandations du rapport Langrenay encourage un système assurantiel plus équitable. La modulation des surprimes, combinée à des incitations fiscales et des mesures de prévention, garantit l'accessibilité de l'assurance dans les zones à haut risque. Cette étude démontre qu'il est possible de concilier viabilité financière, équité sociale et adaptation aux défis climatiques, renforçant ainsi le rôle clé de l'assurance dans la résilience collective face à l'intensification des aléas naturels.

Conclusion

La construction d'un zonier permet d'évaluer l'exposition de l'assureur au risque selon des paramètres géographiques et d'affiner la tarification des produits en conséquence. Par ailleurs, la définition d'une politique de souscription permet à l'assureur de piloter son activité et de l'aligner avec ses engagements. Elle pourrait également contribuer à promouvoir un système assurantiel équitable et pérenne. La démarche proposée dans ce mémoire favorise une approche collective en renforçant la collaboration entre assureurs, pouvoirs publics et citoyens, afin d'accroître la résilience des territoires face aux aléas climatiques.

Ce mémoire s'intéresse au risque de grêle, un sujet encore peu exploré dans la littérature actuarielle, malgré son importance croissante dans un contexte de dérèglement climatique. Il met en lumière l'importance d'une gestion collective des risques climatiques et propose une méthodologie basée sur la segmentation des risques et l'application de la théorie de la crédibilité pour construire un zonier adapté. Les travaux réalisés constituent une première base pour développer des stratégies assurantielles efficaces face aux défis climatiques.

Executive summary

Part 1 : Hail : from natural phenomena to insurance solutions

Climate change represents a major challenge for the insurance industry, as extreme weather events such as hail are becoming more frequent. In France, this trend resulted in an exceptional year in 2022 : more than a million hail-related claims were recorded, for a total estimated cost of 4.5 billion euros, of which 3.3 billion were covered by TGN (Storm hail snow) insurance.

Despite the scale of these figures, hail risk remains under-explored in the insurance field. It is often overshadowed by other more frequent or more costly climatic hazards, and the data available on this phenomenon remains limited. However, climate projections point to an increase in the intensity of hail events, requiring an urgent adaptation of insurance practices to preserve the economic viability of contracts, while meeting policyholders' expectations.

This work is in line with this perspective and focuses on modeling hail risk for home insurance, especially in a highly exposed area : the Auvergne-Rhône-Alpes region. The aim of the work is to create a zone-sectioned-map for assessing hail risk on a town scale, and to develop a fair underwriting policy which considers regional disparities and the characteristics of insured properties.

Part 2 : Construction of a hail zone

To meet these objectives, the first step consists of building map with multiple zones dedicated to hail risk exposure. This work is based on the reprocessing and analysis of three databases made available, covering the period from 2019 to 2023. These databases include information on the policyholder portfolio, reported claims and recorded climatic events, providing a complete and detailed overview for modeling the risk.

Based on these databases, three approaches were studied to identify the insurer's exposition per zone. The latter corresponds to the development of a risk score, represented graphically while considering geographical constraints. In the context of this work, the score is based on pure premiums and allows a classification of zones according to their level of exposure.

First approach : Iterative construction of a zoning system

The first zoning method involves separately creating :

- **An intensity-based zoning map**, built on the maximum observed hailstone size per town, considering the discrepancies between observed and predicted average costs from regression analysis.
- **A frequency-based zoning map**, calculated from the annual number of hail events, distinguishing those that resulted in claims.

These two zoning maps are then combined using the cross-reference grid applied by the MRN.

While this approach is relatively straightforward to implement and understandable for non-expert stakeholders, it has significant limitations, such as reduced integration of portfolio-specific characteristics, limited scientific robustness, and low reliability of results. For these reasons, this method was not selected for further study.

Second approach : Construction of a zoning system using GLMs

The second approach relies on GLMs to model the frequency and average cost of claims. These models incorporate explanatory variables such as the frequency and intensity of hail episodes and property characteristics, particularly the number of rooms. A logarithmic link function is selected to express the multiplicative relationships between variables, making it easier to interpret results and adjust premiums.

Three datasets are structured with varying levels of granularity :

- **Dataset 1** : One row per town (4,500 rows).
- **Dataset 2** : One row per town and year (22,000 rows).
- **Dataset 3** : One row per town, year, and number of rooms (170,000 rows).

Claims frequency modeling

Claims frequency is modeled using a negative binomial distribution, due to the over-dispersion of the data. Different models are evaluated using the p-values of the coefficients, the R^2 , the log-likelihood, the Akaike criterion (AIC) and the F-score.

The model selected is the one using dataset 1, with the number of policyholders and the number of events per hailstone class (number of A1 events, number of A2 events, etc.) as explanatory variables, and a negative binomial distribution. The indicators obtained for this model are as follows :

Database	Explanatory variables	Distribution	P-value of coefficients	R^2	Log-likelihood	AIC
Dataset 1	- Number of policyholders - Number of events by intensity	Negative binomial	All < 0.1%	62%	-11328.60	22671

In addition, the results obtained for the F-score can be seen in the following table :

Database	Distribution	Global F-score	F-score no claims	F-score one claim	F-score more than one claim
Dataset 1	Negative binomial	0.517	0.59	0.622	0.277

Average cost modeling

Similar work has now been carried out to model the average cost of claims. This is fitted using a Gamma distribution, adapted to represent the asymmetric nature of the data, marked by a long tail characteristic of high claims. Once again, the model adopted uses dataset 1, combined with this distribution, with the number of events per class as the main explanatory variables.

Database	Explanatory variables	Distribution	P-value of coefficients	R^2	Global F-score
Dataset 1	Number of events by intensity	Gamma	A1 : 0.723 A2 : 0.129 A3 : 0.232 A4 : < 0.1% A5 : < 0.1%	4.4%	0.466

However, the indicators of this model are unsatisfactory. Indeed, the high p-values for some coefficients reflect a low statistical significance, and the relatively low coefficient of determination R^2 reflects a limited capacity of the model to explain the variability of the data. These results were expected given the complexity of predicting the average cost with a relatively simple model.

Scoring adjustment : classification and smoothing

Once the pure premiums are calculated for each municipality, a classification step is carried out to assign a risk level (low, moderate and high) to each of them, followed by geographical smoothing.

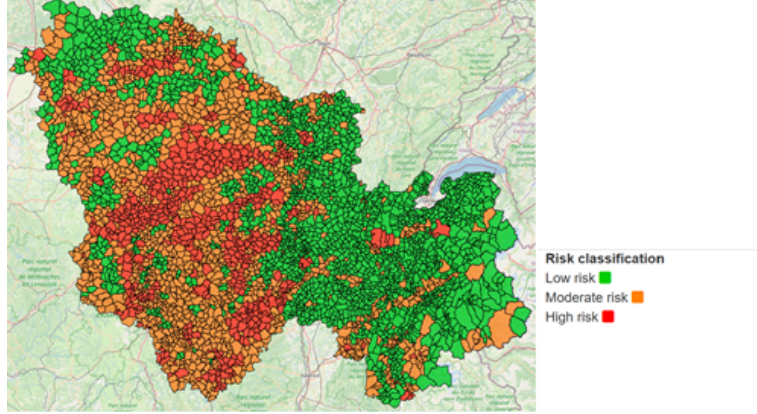


FIGURE 8 – Zoning by GLM of the exposure of the insurer to hail risk for the Auvergne-Rhône-Alpes region after smoothing

The GLM model has several limitations, including a limited number of explanatory variables, which prevents capturing the full complexity of the phenomena. Additionally, the low R^2 indicates that the model explains only a small portion of the variability in the data. This suggests the need for a more suitable approach, such as credibility theory.

Third approach to zoning : credibility theory

The GLM-based zoning model is challenged by the credibility theory approach, particularly using the Bühlmann-Straub model. This model combines portfolio-wide data with data specific to homogeneous subgroups of policyholders, based on similar characteristics. It calculates the pure premium by aggregating claim frequency and average cost, and then establishes a risk scoring system. The Bühlmann-Straub model assigns an optimal weight between the global averages (portfolio level) and specific averages (subgroup level), determined by the credibility parameter α , which depends on the variance between groups and the variance within groups. This approach produces more reliable estimates, regardless of the amount of data provided by the subgroups.

Modeling Claim Frequency

To apply the Bühlmann-Straub model, homogeneous groups in terms of claim frequency, distinct from each other, are created in two steps :

- Grouping towns based on hail event risk.
- Cross-referencing with the number of rooms in properties.

This process results in 12 groups, to which the Bühlmann-Straub model is applied, giving the following results :

Event types / Nb rooms	1 to 3 rooms	4 to 7 rooms	8 rooms or more
No events	0,092%	0,101%	0,243%
Low frequency	0,291%	0,480%	0,670%
Medium frequency	0,659%	1,303%	1,427%
High frequency	3,509%	7,175%	4,654%

FIGURE 9 – Frequency predicted by the Bühlmann-Straub model as a function of the number of rooms and the number of hail events experienced

Modeling average claims costs

The average cost is modelled based on two criteria :

- Departments
- Number of rooms

This creates 9 groups, to which the Bühlmann-Straub model is applied, giving the following results :

Departments / Nb rooms	1 to 4 rooms	5 to 7 rooms	8 rooms or more
1, 18, 38, 69	4195,03€	4464,88€	5146,84€
42, 43, 71, 73	4680,05€	5509,19€	6300,12€
3, 58, 63, 74	8485,01€	9119,58€	7316,64€

FIGURE 10 – Average cost predicted by the Bühlmann-Straub model as a function of department and number of rooms

Zonal construction and classification

By combining frequency and average cost, a pure premium is calculated for each policyholder, then aggregated to the town level. Towns are classified into three risk categories in the same way as for the GLM map. Geographical smoothing is then applied in the same way, to obtain the following zoning :

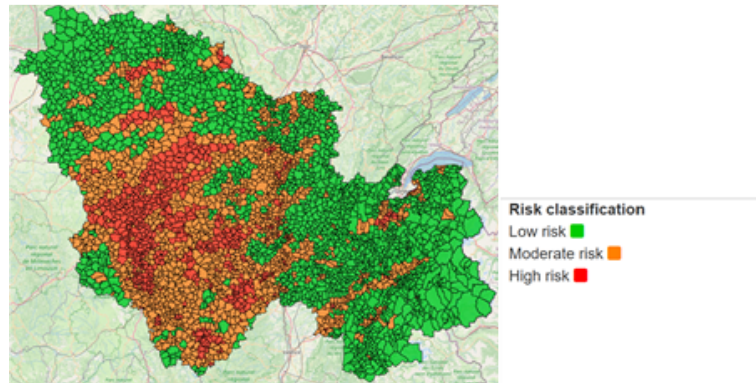


FIGURE 11 – Zone-sectioned-map after smoothing, based on the theory of credibility of the insurer's exposure to hail risk for the Auvergne-Rhône-Alpes region

Despite its benefits, credibility theory has certain limitations. Firstly, this method can appear abstract and requires a strong understanding of the underlying concepts for correct implementation. Moreover, the quality of the results largely depends on the relevance of the risk groups defined beforehand, making this step critical.

Comparison of zoning methods

The three methods explored offer both advantages and limitations. However, the smoothed zoning model derived from credibility theory is retained for several reasons, including :

- The adequacy of using credibility theory in studies with limited data.
- The homogeneity of the resulting map.
- The alignment of zones identified as high-risk with those that have experienced intense hail episodes, as demonstrated by the maps below :

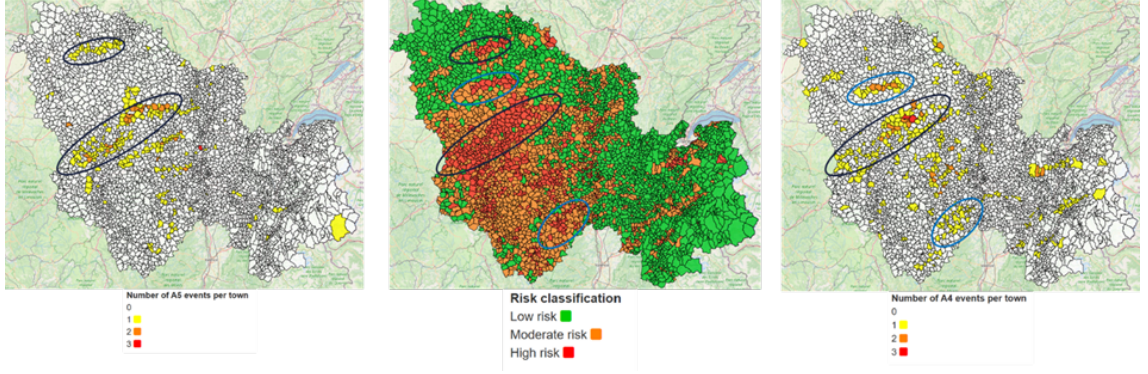


FIGURE 12 – Maps of the number of events A5 on the left, A4 on the right and the zonier by theory of the credibility of the insurer's exposure to hail risk for the Auvergne-Rhône-Alpes region in the center

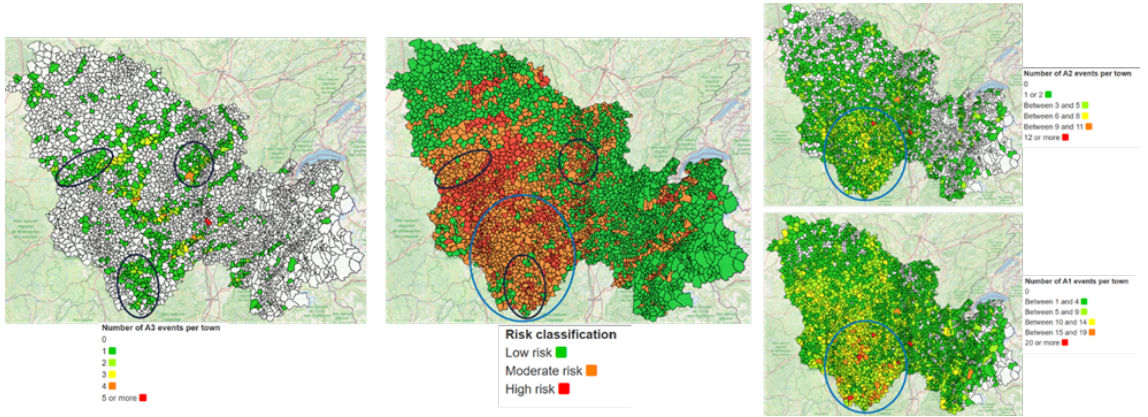


FIGURE 13 – Maps of the number of events A3 left, A2 top right, A1 bottom right and the zone-sectioned map by credibility theory of the insurer's exposure to hail risk for the Auvergne-Rhône-Alpes region in the center

Part 3 : Defining an underwriting policy based on the zoning model

The construction of the zoning model enables the definition of a sustainable and fair underwriting policy. This practical application specifically allows :

- The development of a pricing structure based on adjusted pure premiums, accurately reflecting the risk by zone.
- The identification of potential disparities or inequities in the distribution of premiums and risks, providing a clear view of necessary adjustments.
- Anticipation of climate and regulatory changes through prospective scenarios.

Initially, a pricing table is calibrated based on the zoning model to ensure actuarial balance. The pricing table relies on two variables : the risk zone and the number of rooms in the house. A coefficient α adjusts premiums according to the size of the property, ensuring that premiums exactly cover the costs. The resulting pricing table is as follows :

Number of rooms \ Risks zone	Low	Medium	High
1	4,28€	27,77€	141,40€
2	5,43€	33,37€	163,83€
3	6,58€	38,97€	186,26€
4	7,73€	44,57€	208,69€
5	8,88€	50,17€	231,12€
6	10,03€	55,77€	253,55€
7	11,18€	61,37€	275,98€
8	12,33€	66,97€	298,41€
9	13,48€	72,57€	320,84€
10	14,63€	78,17€	343,27€
11	15,78€	83,77€	365,70€
12	16,93€	89,37€	388,13€
13	18,08€	94,97€	410,56€
14	19,23€	100,57€	432,99€
15	20,38€	106,17€	455,42€

Premiums vary significantly between zones, with marked differences in high-risk areas. For instance, the home insurance premium (MRH) for a 10-room house is thirty times higher in a red zone than in a green zone. This pricing disparity creates inequalities among policyholders in different zones, potentially discouraging subscriptions and contributing to insurers' disengagement. Moreover, with the intensification of climate risks, claims are expected to worsen in exposed areas, further amplifying these inequalities and potential disengagements.

Given these findings, it is essential to define a sustainable and fair underwriting policy. This involves accounting for the expected increase in claims due to climate change and proposing solutions to encourage insurers to remain engaged in high-risk zones.

Adapting the price table to climate change

Climate change is likely to alter the frequency and intensity of hail events, leading to an increase in claims. An analysis by the Covéa group forecasts a 40% rise in hailstorms by 2050, resulting in a 20% increase in claims.

To maintain balance, estimated premiums will need to gradually increase by 20% by 2050, adjusted through an annual linear growth of 0.92%. However, this annual adjustment is proposed to begin in five years to account for the exceptional nature of 2022. Additionally, the zoning model will need to be updated to reflect shifts in the most exposed areas.

Adapting the pricing table to regulatory changes

Regulatory changes proposed in the Langrenoy report (2024) aim to enhance the insurability of climate risks. This work focuses on a specific measure designed to reduce geographical discrimination in high-risk zones while ensuring equitable coverage. The measure involves adjusting the natural catastrophes (Nat Cat) surcharge based on geographical exposure. The impact of this adjusted surcharge on the hail pure premium, compared to a uniform 20% surcharge (intended to fund the natural disaster compensation scheme) applied across all zones, is presented below :

Risk zones / Number of rooms	Low		Medium		High	
	Pure hail premium with uniform Nat Cat surcharge of 20%	Pure hail premium with modulated Nat Cat surcharge	Pure hail premium with uniform Nat Cat surcharge of 20%	Pure hail premium with modulated Nat Cat surcharge	Pure hail premium with uniform Nat Cat surcharge of 20%	Pure hail premium with modulated Nat Cat surcharge
1	5,14 €	8,56 €	33,32 €	30,55 €	169,68 €	144,23 €
2	6,52 €	10,86 €	40,04 €	36,71 €	196,60 €	167,11 €
3	7,90 €	13,16 €	46,76 €	42,87 €	223,51 €	189,99 €
4	9,28 €	15,46 €	53,48 €	49,03 €	250,43 €	212,86 €
5	10,66 €	17,76 €	60,20 €	55,19 €	277,34 €	235,74 €
6	12,04 €	20,06 €	66,92 €	61,35 €	304,26 €	258,62 €
7	13,42 €	22,36 €	73,64 €	67,51 €	331,18 €	281,50 €
8	14,80 €	24,66 €	80,36 €	73,67 €	358,09 €	304,38 €
9	16,18 €	26,96 €	87,08 €	79,83 €	385,01 €	327,26 €
10	17,56 €	29,26 €	93,80 €	85,99 €	411,92 €	350,14 €
11	18,94 €	31,56 €	100,52 €	92,15 €	438,84 €	373,01 €
12	20,32 €	33,86 €	107,24 €	98,31 €	465,76 €	395,89 €
13	21,70 €	36,16 €	113,96 €	104,47 €	492,67 €	418,77 €
14	23,08 €	38,46 €	120,68 €	110,63 €	519,59 €	441,65 €
15	24,46 €	40,76 €	127,40 €	116,79 €	546,50 €	464,53 €

FIGURE 14 – Price list with Nat Cat 20% uniform or modulated surcharge

The table shows a 40% increase between the uniform Nat Cat premium and the modulated one for the green zone, a 9% decrease for the orange zone, and an 18% decrease in the red zone. These adjustments make coverage more viable by lightening the financial burden of the most exposed policyholders. However, a significant gap remains between green and red zones, reaching a ratio of around 1 to 15. This underscores the need to integrate other recommendations from the Langreny report, as well as insurers' strategic choices, such as rebalancing between different types of insurance.

Integrating the recommendations of the Langreny report encourages a fairer insurance system. The modulation of excess premiums, combined with tax incentives and prevention measures, guarantees the accessibility of insurance in high-risk areas. This study demonstrates that it is possible to reconcile financial viability, social equity and adaptation to climate challenges, thus reinforcing the key role of insurance in collective resilience to the intensification of natural hazards.

Conclusion

The construction of a sectioned map makes it possible to assess the insurer's exposure to risk, according to geographical parameters and to fine-tune product pricing accordingly. Moreover, the definition of an underwriting policy enables the insurer to manage its business and align it with its commitments. It could also help promote a fair and sustainable insurance system. The approach proposed in this memoir favors a collective approach by strengthening collaboration between insurers, public authorities and citizens, to increase the resilience of territories in the face of climatic hazards.

This memoir focuses on the risk of hail, a subject still relatively unexplored in actuarial literature, despite its growing importance in a context of climate change. It highlights the importance of collective management of climatic risks, and proposes a methodology based on risk segmentation and the application of credibility theory to build a suitable zoning system. This work provides an initial basis for developing effective insurance strategies to meet the challenges of climate change.

Remerciements

Je tiens tout d’abord à remercier mon entreprise d’accueil, Exiom Partners, pour m’avoir accueilli et pour avoir mis à disposition les moyens nécessaires à la réussite de ce projet.

Je remercie sincèrement Maylis Kouek, ma tutrice d’entreprise, pour sa bienveillance, son expertise et le temps précieux qu’elle a consacré à m’encadrer. Ses conseils avisés, son enthousiasme et son soutien constant ont été des moteurs essentiels dans l’avancée de mes recherches et l’approfondissement de mes réflexions.

Je tiens également à adresser mes remerciements à Coralie Charbonnel, ma manager durant ce mémoire, pour m’avoir offert l’opportunité de mener ce projet. Son encadrement et ses retours éclairés ont enrichi ce mémoire et m’ont permis d’acquérir des compétences précieuses.

Un remerciement particulier va à Thomas Peyrat, Axel Kaboré, Richard Ahouansou et Alexandre Albertoli, collègues chez Exiom Partners, pour leur soutien et le partage de leurs connaissances. Leurs échanges constructifs, discussions enrichissantes et esprit d’équipe ont largement contribué à la qualité de ce mémoire.

Aussi, je souhaite remercier l’ENSAE Paris et, en particulier, Antoine Heranval, mon tuteur académique. Son expertise, ses remarques constructives et son accompagnement rigoureux ont joué un rôle déterminant dans la structuration et la profondeur de ce travail.

Enfin, je souhaite exprimer ma reconnaissance envers ma famille et mes proches. Leur soutien indéfectible, leur patience et leurs encouragements ont été une source inestimable de motivation et de réconfort, m’aidant à rester concentré et déterminé jusqu’à l’aboutissement de ce mémoire.

Table des matières

Résumé	1
Abstract	2
Note de synthèse	3
Executive summary	10
Remerciements	17
Introduction	20
1 La grêle : du phénomène naturel aux solutions assurantielles	22
1.1 <i>Présentation du phénomène grêle</i>	22
1.1.1 Définition et présentation du phénomène grêle	22
1.1.2 La grêle en France : tendances, saisonnalité, géographie, événements marquants	25
1.2 <i>La grêle en assurance</i>	28
1.2.1 Les garanties existantes	28
1.2.2 Bilan 2022 des événements de grêle : étude France Assureurs	30
1.2.3 Grêle en France : défis climatiques et réponses assurantielles	32
1.3 <i>Objectifs et enjeux du mémoire</i>	34
1.3.1 Objectifs	34
1.3.2 Enjeux	34
2 Construction d'un zonier grêle	36
2.1 <i>Présentation et traitement des données</i>	36
2.1.1 Sources des données : historiques des sinistres, portefeuilles, météorologiques, INSEE	36
2.1.2 Retraitements	39
2.1.3 Qualité et fiabilité des données	41
2.2 <i>Première approche : construction d'un zonier par itération</i>	42
2.2.1 Principe de la méthode	42
2.2.2 Résultats obtenus	44
2.3 <i>Deuxième approche : construction d'un zonier par GLM</i>	45
2.3.1 Introduction au GLM	45
2.3.2 Structure des modèles et choix des variables explicatives	47
2.3.3 Modélisation de la fréquence des sinistres par assuré et par commune	48
2.3.4 Modélisation du coût moyen d'un sinistre par commune	51
2.3.5 Résultats et interprétations	54
2.4 <i>Troisième approche : construction à partir de la théorie de la Crédibilité</i>	56
2.4.1 Présentation de la théorie de la crédibilité	56
2.4.2 Modèle de Bühlmann	57
2.4.3 Modèle de Bühlmann-Straub	60
2.4.4 Application de la théorie de la crédibilité pour la construction d'un zonier	63
2.4.5 Résultats et interprétations	68
2.5 <i>Comparaison des 3 méthodes</i>	69
2.5.1 Avantages et inconvénients de chaque méthode	69
2.5.2 Comparaison avec les cartographies d'événements Keraunos	70
2.5.3 Choix final du zonier	72
3 Définition d'une politique de souscription à partir du zonier	73

3.1	<i>Cycle de vie d'un produit d'assurance</i>	73
3.1.1	Création d'un produit	74
3.1.2	Modélisation	74
3.1.3	Politique de souscription	75
3.1.4	Commercialisation	76
3.1.5	Suivi du portefeuille	76
3.1.6	Provisionnement	77
3.2	<i>Définir une politique de souscription à partir du zonier</i>	79
3.2.1	Présentation de la tarification dans le cadre actuariel	79
3.2.2	Tarification de la garantie à partir du zonier	81
3.2.3	Définition d'une politique de souscription	85
3.3	<i>Scénario de sensibilité : dérèglement climatique et changements réglementaires</i>	86
3.3.1	Politique de souscription avec prise en compte du changement climatique	86
3.3.2	Politique de souscription avec prise en compte de changements réglementaires	88
	Conclusion	92
	Références	94

Introduction

Dans un contexte où le dérèglement climatique s'accélère, les événements météorologiques extrêmes prennent une place de plus en plus importante dans la vie des populations et représentent des défis majeurs pour les systèmes d'assurance, en France notamment. Ces événements ne sont plus considérés comme de simples anomalies, mais comme des occurrences fréquentes et amplifiées, entraînant des impacts significatifs sur les infrastructures, les biens et l'économie mondiale. Parmi ces aléas, la grêle se distingue par sa complexité à être anticipée et modélisée. Bien que localisée dans le temps et l'espace, elle peut causer des dégâts matériels et économiques considérables pour les particuliers, les entreprises et les collectivités. Les records de sinistralité enregistrés en 2022 témoignent de la gravité croissante de ce phénomène en France et renforcent l'urgence de développer des outils performants pour sa gestion.

En France, la grêle représente un risque particulièrement préoccupant en raison de sa fréquence, de son intensité et des dommages qu'elle engendre. Son impact touche de multiples secteurs :

- Les habitations, dont les toitures, fenêtres et autres éléments structurels subissent souvent des dommages sévères,
- Les infrastructures, dont la restauration nécessite des coûts élevés,
- Les cultures agricoles, qui peuvent être anéanties en quelques minutes, mettant en péril les revenus des exploitants,
- Les véhicules, particulièrement vulnérables lors d'épisodes violents.

L'importance des épisodes de grêle est mise en évidence par les chiffres de 2022. Ces événements provoquent non seulement des pertes financières considérables pour les acteurs économiques, mais posent également des défis aux assureurs, responsables d'anticiper ces risques et de les intégrer dans leurs politiques tarifaires et contractuelles.

Les sinistres liés à la grêle en France relèvent principalement de la garantie *Tempête, Grêle, Neige* (TGN), incluse dans les contrats d'assurance dommages. Contrairement à d'autres aléas climatiques, tels que les inondations ou la sécheresse qui relèvent du régime des catastrophes naturelles et qui sont partiellement réassurés par la Caisse Centrale de Réassurance (CCR), la tarification des sinistres grêle repose entièrement sur les assureurs. Cette spécificité confère aux assureurs un contrôle dans la tarification et la gestion de ce risque, tout en complexifiant leur stratégie. Cela exige la mise en place d'outils analytiques performants pour mieux comprendre et modéliser ce phénomène.

Face à ces défis, la création d'un zonier grêle s'impose comme une réponse méthodologique classique et pertinente. Un zonier est un outil de cartographie permettant d'identifier les zones géographiques les plus exposées à un risque. Il offre la possibilité d'affiner les politiques tarifaires et de souscription des assureurs en fonction des spécificités locales. En outre, il constitue un levier essentiel pour renforcer la résilience du secteur assurantiel face à l'intensification des aléas climatiques.

Ce mémoire s'inscrit dans cette démarche et vise à construire un zonier grêle dédié à une région assimilée à l'Auvergne-Rhône-Alpes¹, zone particulièrement vulnérable en raison de sa géographie et de sa climatologie. Cette étude est basée sur les données issues d'un portefeuille d'assurance couvrant cette région, et s'articule autour de trois grandes parties.

La première partie se concentre sur l'exploration du phénomène de la grêle en décrivant les mécanismes à l'origine de la grêle, suivie d'une analyse des tendances en France (fréquence, sévérité, saisonnalité et répartition géographique). Les événements marquants des dernières décennies, comme ceux de 2022, seront mis en lumière, ainsi que leurs conséquences sur le système assurantiel. Enfin, les solutions existantes, incluant les garanties et les propositions d'adaptation du rapport Langrenay [19], seront abordées.

1. Plus précisément, le mémoire s'intéresse aux 12 départements suivants : 1, 3, 18, 38, 42, 43, 58, 63, 69, 71, 73 et 74. Neuf d'entre eux appartiennent à la région Auvergne-Rhône-Alpes. Toutefois, afin de simplifier la présentation, l'ensemble des 12 départements sera désigné sous le terme de région Auvergne-Rhône-Alpes dans la suite.

La deuxième partie s'oriente vers la construction du zonier grêle, objectif principal de ce mémoire. Cette étape repose sur l'utilisation de plusieurs approches méthodologiques : une méthode itérative, des modèles linéaires généralisés (GLM) et la théorie de la crédibilité. Chacune de ces méthodes sera présentée théoriquement, appliquée et comparée en termes de pertinence et de performance. Une comparaison des trois méthodes permettra de choisir celle qui offre la meilleure performance, à la fois en termes de précision et de simplicité d'utilisation sachant que l'objectif est de concevoir un zonier fiable permettant une meilleure évaluation et gestion du risque.

Enfin, la dernière partie de ce mémoire se focalisera sur l'application pratique du zonier dans la politique de souscription des assureurs, ainsi que sur des scénarios de sensibilité liés aux évolutions climatiques et réglementaires. Cette section abordera notamment la tarification, l'adaptation des stratégies commerciales, et l'intégration de deux enjeux fondamentaux : d'une part, l'impact croissant du dérèglement climatique sur la fréquence et l'intensité des sinistres, d'autre part, les évolutions réglementaires visant à limiter le désengagement potentiel des assureurs dans les zones les plus exposées. Une analyse prospective permettra d'identifier les limites du système actuel et de proposer des pistes concrètes d'amélioration.

Ce mémoire aspire ainsi à apporter une contribution utile à la gestion du risque de grêle, en combinant des méthodes de modélisation avancées et une réflexion stratégique approfondie. L'objectif final est de fournir des outils opérationnels et prospectifs permettant au secteur assurantiel de relever les défis croissants posés par les aléas climatiques dans un monde en constante évolution tout en construisant une résilience collective et durable.

1 La grêle : du phénomène naturel aux solutions assurantielles

1.1 Présentation du phénomène grêle

1.1.1 Définition et présentation du phénomène grêle

Définition de la grêle

La grêle est un phénomène météorologique caractérisé par la formation de précipitations solides sous forme de morceaux de glace, connus sous le nom de grêlons. Il ne faut pas confondre la grêle avec le grésil ou avec la neige. Le grésil se présente sous forme de particules de glace de moins de 5 mm de diamètre, plutôt opaques, qui tombent d'un nuage dit cumulonimbus² en rebondissant au sol sans se briser. Quant à la neige, elle est douce et légère, se formant quand la vapeur d'eau se cristallise directement en flocons dans des conditions de faible température. À l'inverse, les grêlons sont durs et denses, d'un diamètre supérieur à 5 mm, avec une forme souvent irrégulière, bien que généralement circulaire. Lorsqu'ils fusionnent, les grêlons peuvent adopter des formes variées et parfois inhabituelles. Un grêlon est typiquement composé de couches concentriques de glace, opaques ou transparentes, qui se sont formées au fil des différents cycles de croissance à l'intérieur du nuage. Ils varient en taille, depuis de petits morceaux de glace de quelques millimètres de diamètre jusqu'à des boules de glace qui peuvent atteindre plusieurs centimètres de diamètre.



FIGURE 15 – Grêlons, source : [4]

Présentation du phénomène

La formation de grêle se produit principalement pendant les périodes de forte montée d'air chaud dans des conditions météorologiques instables. À mesure que l'air chaud et léger s'élève en altitude, il rencontre de l'air plus frais et plus dense, qui descend ensuite vers le sol. Cette rencontre provoque de la condensation, produisant des nuages épais appelés cumulonimbus, souvent associés à de violents orages.

2. Un **cumulonimbus** est un type de nuage à développement vertical, souvent associé à des conditions météorologiques instables. Il est caractérisé par sa grande hauteur, qui peut s'étendre de quelques centaines de mètres au-dessus du sol jusqu'à la haute troposphère, voire parfois jusqu'à la stratosphère. Ce nuage est connu pour engendrer des phénomènes météorologiques violents tels que des orages, des averses de pluie, de grésil ou de grêle, des éclairs, et parfois des tornades. Les cumulonimbus ont une base relativement sombre et plate, mais leur sommet peut s'étendre en forme d'enclume, marquant la limite où l'air ascendant ne peut plus monter plus haut à cause des températures plus froides de la troposphère supérieure.

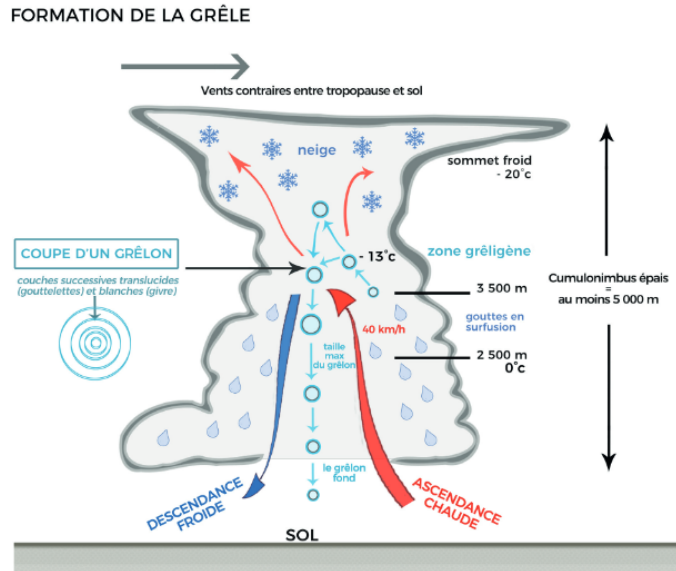


FIGURE 16 – Formation de la grêle, source : [6]

Une fois ces cumulonimbus formés, les conditions sont idéales pour le développement de la grêle, en raison des forts courants ascendants et descendants qui se produisent entre la base chaude et humide et le sommet très froid du nuage. Lorsqu'un courant ascendant est particulièrement puissant, il transporte des gouttelettes d'eau au-delà de l'isotherme 0°C, les rendant extrêmement adhésives. Ces gouttelettes s'agglomèrent autour de cristaux de glace et gèlent, formant ainsi une couche de glace translucide autour des cristaux.

Les grêlons continuent de croître par accumulation successive de ces couches de glace, contribuant ainsi à leur formation. Ils peuvent également grossir par givrage, qui crée une couche de glace blanche. Lorsque l'accumulation de glace translucide et le givrage alternent, les grêlons développent une structure en couches concentriques. Cela se produit lorsque le grêlon monte et descend plusieurs fois dans le nuage, ou lorsqu'il traverse différentes zones de température pendant son ascension. Les grêlons peuvent également augmenter de taille en fusionnant avec d'autres cristaux de glace ou grêlons, formant des agrégats. Ils finissent par tomber au sol lorsque leur poids excède la force des courants ascendants, ce qui explique pourquoi seules les ascendances les plus rapides permettent la croissance des grêlons.

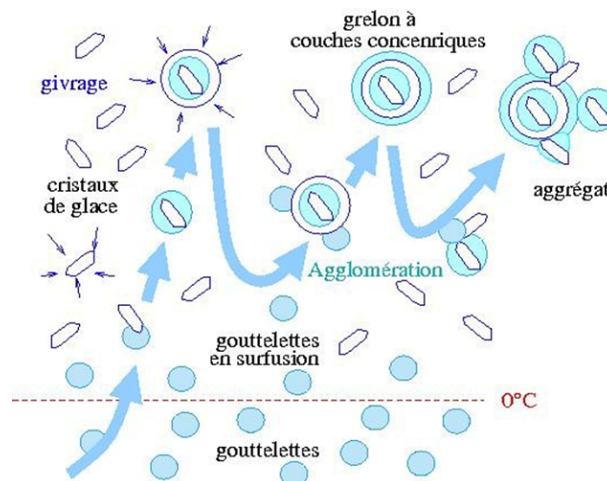


FIGURE 17 – Formation de grêlons par agglomération, source : [3]

L'incidence du phénomène

La grêle est un phénomène météorologique susceptible d'avoir des impacts majeurs sur les biens et les personnes, causant des dommages importants, notamment :

- **Aux cultures** : La grêle peut détruire les récoltes, entraînant des pertes financières significatives pour les agriculteurs.
- **Aux véhicules** : Les grêlons peuvent endommager la carrosserie, les vitres, et les éléments extérieurs des voitures, nécessitant des réparations coûteuses, ainsi que causer des accidents de la route.
- **Aux bâtiments** : Les habitations sont également vulnérables face à la grêle, car elle peut simultanément endommager les toitures, les fenêtres, les façades et les équipements extérieurs. Les toitures abîmées peuvent provoquer des infiltrations d'eau, aggravant les dégâts à l'intérieur de la maison, tandis que les fenêtres brisées et les systèmes extérieurs endommagés nécessitent souvent des réparations coûteuses.

La vitesse de chute des grêlons et ses conséquences

La vitesse de chute des grêlons est directement proportionnelle à leur taille : plus un grêlon est volumineux, plus sa vitesse de chute est élevée. Cette augmentation de la vitesse accentue les impacts au sol, amplifiant ainsi la gravité des dommages. Par exemple, les petits grêlons peuvent seulement causer des dommages superficiels aux cultures, tandis que les grêlons plus gros ont le potentiel de perforer les toitures, briser les vitres et infliger des blessures graves aux personnes. La diversité des tailles de grêlons entraîne donc une variété de dommages selon les objets ou surfaces touchés.

Pour évaluer ces impacts de manière plus précise, l'échelle ANELFA³ est couramment utilisée. Cette échelle, qui s'appuie sur la taille des grêlons et les dégâts observés, permet de quantifier l'intensité du phénomène et d'estimer les coûts potentiels des réparations nécessaires.

CLASSE ANELFA ▶	A0	A1	A2	A3	A4	A5
Diamètre maximal des grêlons	< 1	1-1.9	2-2.9	3-3.9	4-4.9	>= 5
Terme usuel	grésil, petit pois	bille, grain de raisin, cerise	oeufs de pigeon, pièce de 2 €	noix, balle de ping-pong	oeuf de poule, balle de Golf	pêche, pomme orange, balle de tennis
Energie cinétique moyenne	10 J.m ⁻²	50 J.m ⁻²	200 J.m ⁻²	500 J.m ⁻²	800 J.m ⁻²	
dommages types	accidents de la route, fleurs coupées	dommages aux vignes vergers, tabac	dommages importants aux céréales légumes, arbres	dommages à 100% sur toute culture, vitres cassées, voitures endommagées	paysage d'hiver, animaux tués, personnes blessées, avions au sol endommagés	évènement extrêmement dangereux, risque mortel
couverture moyenne du sol	0.1	0.15		0.35	0.35	

FIGURE 18 – Description des classes ANELFA, source : [1]

3. Association Nationale d'Etude et Lutte contre les Fléaux Atmosphériques : elle agit depuis 1951 pour réduire les dégâts liés à la grêle

1.1.2 La grêle en France : tendances, saisonnalité, géographie, événements marquants

Les chutes de grêle sont un phénomène récurrent en France, comme dans la plupart des pays situés sous des latitudes tempérées. La formation de la grêle nécessite une instabilité atmosphérique significative ainsi qu'une masse d'air modérément chaude. Cela explique la rareté des fortes chutes de grêle sous les latitudes polaires, où les températures sont trop basses, ainsi qu'aux latitudes tropicales et équatoriales, où les grêlons fondent en grande partie avant d'atteindre le sol.

En France, les conditions climatiques sont particulièrement favorables à la formation de la grêle, qui représente ainsi un risque majeur, surtout dans certaines régions où elle se manifeste régulièrement. Cette section explore les tendances, la saisonnalité, la répartition géographique, ainsi que quelques événements marquants liés à la grêle en France.

Tendance et évolutions en France : fréquence et gravité des sinistres

Des études montrent une augmentation non pas de la fréquence des épisodes de grêle, mais plutôt de leur intensité, fortement corrélée à la température nocturne (Berthet, Dessens et Sánchez, 2011) [26]. Les projections climatiques prévoient une augmentation de 40 % de l'intensité des chutes de grêle d'ici 2040, sans évolution notable de leur fréquence (Dessens, Berthet et Sánchez, 2015) [27]. Récemment, le groupe Covéa a sorti une étude qui anticipe une hausse de 40% des orages de grêle sur le territoire français dans le cadre du scénario RCP8.5, accompagnée d'une augmentation de 20% de la charge de sinistres (hors inflation). [20] Toutefois, il convient de noter que peu d'études ont été réalisées sur ce sujet, ce qui invite à interpréter ces résultats avec prudence.

En effet, les tendances récentes suggèrent que le changement climatique pourrait accroître l'intensité des orages de grêle. L'élévation des températures de surface et les modifications des régimes de vent semblent créer des conditions plus propices à la formation de grêlons plus gros et plus nombreux, entraînant ainsi des dommages accrus sur les cultures, les bâtiments et les véhicules.

Les travaux de Keraunos⁴, qui a développé un indicateur national de la fréquence et de la sévérité des chutes de grêle en France, confirment ces observations. Cet indicateur, calculé quotidiennement depuis le 1er janvier 2006, attribue à chaque journée un score d'activité grêle, permettant ainsi d'établir des statistiques comparatives à différentes échelles, qu'elles soient locales, départementales ou nationales.

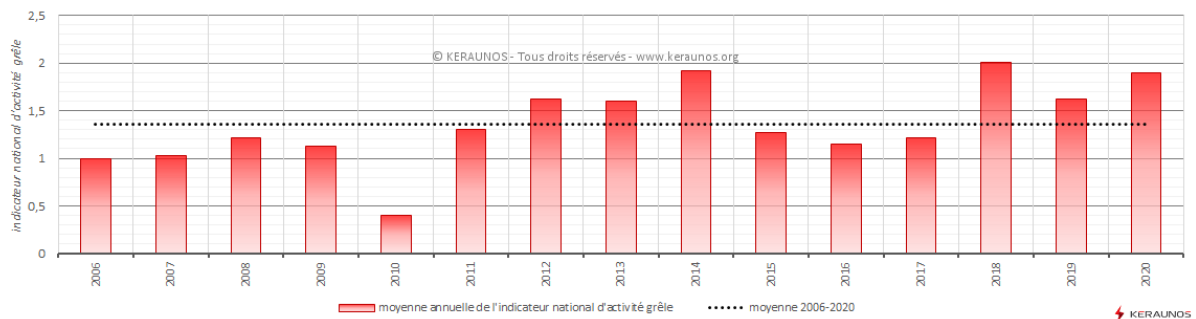


FIGURE 19 – Evolution de l'activité grêle en France de 2006 à 2020, source : [5]

L'évolution de l'activité grêle en France entre 2006 et 2020 montre une tendance à la hausse. De 2006 à 2011, aucune année n'a dépassé la moyenne sur les années de 2006 à 2020. En revanche, l'année 2018 est celle qui a connu le plus d'épisodes de grêle, suivie de près par 2014 et 2020. Absentes sur ce graphique, les années 2021 et surtout 2022, qui est l'année la plus grêleuse depuis plus de 15 ans, confirment cette tendance.

4. Keraunos, aussi appelé observatoire français des tornades et orages violents est le premier bureau d'études français spécialisé dans la prévision et la gestion des risques liés aux phénomènes de la grêle, de la foudre, des pluies diluviennes et des rafales de vent.

Saisonnalité et répartition géographique

En France, la grêle présente une **forte saisonnalité**. Contrairement à l'idée reçue selon laquelle il s'agirait d'un phénomène hivernal, la grande majorité des épisodes se produit entre mai et septembre, avec une fréquence particulièrement élevée en **mai, juin et juillet, période propice aux orages**. Les statistiques montrent que les chutes de grêle sont rares en dehors de ces mois, bien qu'elles restent possibles. Le graphique ci-dessous illustre cette tendance, en présentant le nombre de chutes enregistrées par mois au cours des 20 dernières années par l'ANELFA, classées selon le diamètre des grêlons.

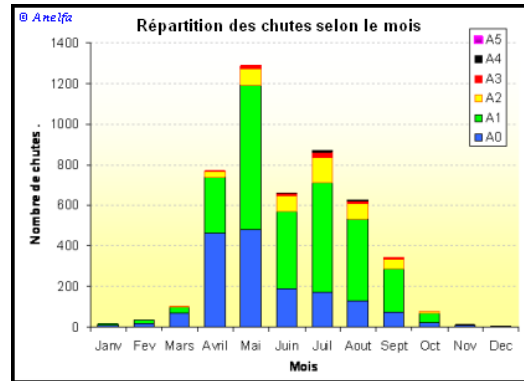


FIGURE 20 – Répartition mensuelle grêle, source : [2]

En termes de **répartition géographique**, la grêle en France ne touche pas toutes les régions de manière uniforme. Une carte des régions les plus exposées, élaborée par Freddy Vinet dans son ouvrage *Le risque grêle en agriculture*, distingue les zones selon la fréquence moyenne des épisodes de grêle estivale, leur intensité et les conditions météorologiques favorables à la formation d'orages grêligènes. De plus, la Mission Risques Naturels (MRN) a également publié une carte mettant en évidence les régions les plus à risque, confirmant ces observations. Les **régions du sud-est intérieur, du sud-ouest intérieur, ainsi que la région Auvergne-Rhône-Alpes sont particulièrement vulnérables**. Le mémoire se portera sur cette dernière région, qui est donc très vulnérable face au risque grêle.

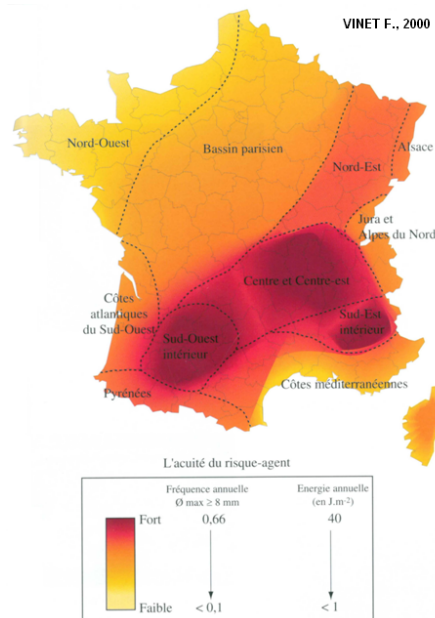


FIGURE 21 – Répartition géographique grêle, source : [2]

Evènements marquants

La France a connu plusieurs épisodes de grêle particulièrement destructeurs qui ont laissé une empreinte durable et ont conduit à des adaptations, notamment au niveau assurantiel. Parmi les plus notables [5] :

- **25 mai 2009** : Un épisode orageux accompagné de grêlons géants a frappé la France sur un axe Nord-Pas-de-Calais/Pyrénées, causant des dégâts considérables, notamment sur les habitations et les véhicules. Des grêlons d'un diamètre compris entre 6 et 12 cm ont été repérés dans le département du Nord.
- **6 au 10 juin 2014** : la tempête de grêle Ela s'est imposée comme l'une des plus violentes depuis 2006, frappant particulièrement l'Île-de-France et le Loiret. Les grêlons géants ont causé d'importants dégâts aux infrastructures. Selon la Fédération Française de l'Assurance, les dommages aux bâtiments ont été estimés à 420 millions d'euros, représentant 43% du total des sinistres TGN liés aux bâtiments. Le secteur automobile n'a pas été épargné, avec des pertes atteignant 430 millions d'euros, portant ainsi le coût total de cet événement à 850 millions d'euros [7].
- **4 juillet 2018** : Des chutes de grêle violentes ont affecté une large partie du sud-ouest de la France, provoquant des dégâts massifs sur les toitures et les réseaux électriques.
- **4 juin 2022** : Cette journée a été marquée par des chutes de grêle exceptionnelles, des Pyrénées jusqu'en Normandie et en Bourgogne. Des supercellules grêligènes très actives ont provoqué des dégâts majeurs sur les bâtiments et les véhicules, établissant un nouveau record de l'indicateur national de grêle.

À l'échelle mondiale, quelques records homologués par l'Organisation météorologique mondiale (OMM) et le National Oceanic and Atmospheric Administration (NOAA) sont à noter :

- **Le plus gros grêlon**, avec un diamètre de 20,3 cm, est tombé à Vivian, Dakota du Sud, le 23 juillet 2010.
- **Le grêlon avec la plus grande circonférence**, mesurant 47,6 cm, est tombé à Aurora, Nebraska, le 22 juin 2003.
- **Le grêlon le plus lourd**, pesant 1,02 kg, a été enregistré à Gopalganj, Bangladesh, le 14 avril 1986.

1.2 La grêle en assurance

Après avoir examiné les caractéristiques de la grêle en France, notamment sa répartition géographique et sa saisonnalité, il est essentiel de se pencher sur les dispositifs possibles pour assurer les biens contre ce risque. Les assurances jouent un rôle important dans la gestion des conséquences de ce phénomène météorologique, offrant différentes garanties adaptées aux besoins des agriculteurs, des automobilistes, des propriétaires immobiliers, ainsi que des professionnels.

La suite de cette section présente les principales formes de couvertures disponibles pour se prémunir contre les dommages causés par la grêle, en commençant par l'assurance agricole, automobile, et la multirisque professionnelle (MRP), avant de s'attarder sur la Multirisque Habitation (MRH).

1.2.1 Les garanties existantes

L'assurance agricole et la grêle

Pour les agriculteurs, la grêle représente l'un des risques climatiques les plus redoutés. En quelques minutes, une tempête de grêle peut ruiner une récolte entière, entraînant des pertes financières considérables. Afin de se prémunir contre de telles catastrophes, l'assurance multirisque climatique agricole inclut généralement une couverture spécifique contre la grêle. Cette assurance prend en **charge les dommages causés aux cultures (perte d'exploitation), et dans certains cas, aux installations agricoles** (serres, systèmes d'irrigation, etc.).

La souscription à ce type de couverture a longtemps été soutenue par des subventions publiques, dans le cadre de politiques visant à encourager la résilience des exploitations agricoles face aux aléas climatiques. Toutefois, face à la multiplication des aléas météorologiques, liés au changement climatique, le gouvernement de Jean Castex, avec en particulier le ministère de l'Agriculture et de la Souveraineté alimentaire dirigé par Julien Denormandie, a réformé le dispositif d'assurance récolte à travers la loi du 2 mars 2022.

La réforme de l'assurance récolte repose désormais sur la solidarité nationale et le partage du risque entre l'État, les agriculteurs et les assureurs. En effet, une distinction a été faite entre trois types d'aléas :

- **Aléas courants** : assumés par les agriculteurs, qui peuvent s'appuyer sur des outils (Dotation pour Epargne de Précaution),
- **Aléas significatifs** : pris en charge par l'assurance multirisques climatiques subventionnée,
- **Aléas exceptionnels** : déclenchent l'intervention de l'État via le Fonds de Solidarité Nationale :
 - À partir de 50% des pertes de récolte pour les grandes cultures, cultures industrielles, légumes et viticulture.
 - À partir de 30% des pertes de récolte pour l'arboriculture, les petits fruits, les prairies et les cultures spécialisées.

De plus, cette réforme visait également à inciter davantage les agriculteurs à souscrire une assurance car seulement 17 % de la surface agricole utile était assurée en 2022. Pour cela, la subvention sur les primes et cotisations d'assurance a été renforcée. Le taux de subvention des primes et cotisations d'assurance est porté à 70 % en 2023 (contre 62 % en moyenne en 2022), et le périmètre des garanties subventionnables a été élargi, avec notamment un niveau de franchise subventionnable dès 20 % (contre 25 % en 2022). Par ailleurs, en cas de sinistre, les conditions d'indemnisation seront plus favorables pour les agriculteurs ayant souscrit un contrat d'assurance récolte.

La Multirisque Professionnelle (MRP) et la grêle

La MRP est une assurance essentielle, mais non obligatoire, pour les entreprises et les travailleurs indépendants, couvrant un large éventail de risques liés à leur activité. Parmi les garanties offertes, la protection contre les dommages des intempéries sur les biens, y compris la grêle, est particulièrement importante pour les entreprises dont les biens matériels, les locaux ou les stocks sont exposés à ce type de risque. La couverture contre la grêle dans le cadre de la MRP s'applique généralement aux bâtiments professionnels, aux équipements extérieurs, ainsi qu'aux marchandises stockées, qui peuvent être gravement endommagés par la chute de grêlons, entraînant ainsi une interruption de l'activité.

Les entreprises perturbées par ces phénomènes, bénéficiant d'une garantie "pertes d'exploitation" dans leur contrat d'assurance, peuvent également être indemnisées selon les conditions prévues. Cette assurance vise à compenser les difficultés financières engendrées par un sinistre, en couvrant les pertes de revenus consécutives à l'arrêt de l'activité dû à des dommages matériels causés par la grêle.

Par exemple, si une entreprise subit des dégâts sur sa toiture entraînant des infiltrations d'eau, elle devra cesser son activité. En plus des coûts liés à la réparation du toit, l'entreprise subira également un manque à gagner en raison de l'interruption de son activité.

La couverture grêle pour les véhicules

L'assurance automobile offre également des protections contre les dommages causés par la grêle, particulièrement dans le cadre des contrats d'assurance "tous risques". La grêle peut endommager la carrosserie, briser les parebrises et causer d'autres types de dégâts coûteux à réparer. Les contrats d'assurance automobile tous risques incluent généralement une garantie spécifique pour couvrir ces dommages.

L'indemnisation dépend de plusieurs facteurs, tels que l'état du véhicule avant le sinistre, sa valeur sur le marché, et les termes précis du contrat. Comme pour d'autres types de garanties, une franchise peut s'appliquer, ce qui signifie que l'assuré devra supporter une partie des coûts de réparation. De plus, certaines polices d'assurance proposent des services additionnels, comme la mise à disposition d'un véhicule de remplacement ou la couverture "valeur à neuf", qui offre une indemnisation correspondant au prix d'achat initial du véhicule, sous certaines conditions. Cela peut s'avérer particulièrement avantageux dans les cas où les dommages sont si importants que le véhicule est considéré comme irréparable.

La Multirisque Habitation (MRH) et la grêle

La MRH est un contrat d'assurance destiné à protéger les habitations contre une variété de risques, incluant notamment la grêle. Les dommages causés par les grêlons sur les toitures ou les fenêtres ne relèvent pas de la garantie catastrophe naturelle fixée par les articles L.125-1 et suivants du Code des Assurances, car ces dégâts sont considérés comme assurables. En effet, ces dommages sont couverts par les contrats d'assurance au titre de la garantie "Tempêtes, Grêle et Neige" (TGN).

Contrairement aux catastrophes naturelles, aucune attente de publication d'un arrêté interministériel n'est nécessaire pour déclarer un sinistre lié à la grêle, une déclaration peut être effectuée dès que les dégâts sont constatés, dans un délai de cinq jours ouvrés.

La liste des dégâts pris en charge par l'assurance MRH dans le cadre d'une chute de grêle varie selon les contrats, mais elle inclut généralement :

- La réparation de la toiture, y compris la couverture provisoire.
- Le remboursement du mobilier endommagé par la grêle, y compris les aménagements extérieurs.
- La remise en état des fenêtres brisées.
- La gestion des infiltrations d'eau survenant dans les heures suivant les dommages.

Les contrats MRH sont modulables, permettant aux assurés d'adapter la couverture en fonction de leurs besoins spécifiques. Par exemple, les propriétaires de maisons équipées de panneaux solaires ou d'installations particulières, comme des vérandas, peuvent souscrire à des options supplémentaires pour renforcer la protection de ces biens contre la grêle. Lorsqu'un sinistre survient, l'indemnisation est calculée sur la base des dommages réels, après déduction d'une franchise, souvent précisée dans le contrat.

Avec l'intensification des changements climatiques et l'augmentation de la fréquence des tempêtes de grêle, certaines compagnies d'assurance revoient les termes de leurs contrats MRH pour mieux adapter les garanties proposées à ces nouveaux défis. Ces ajustements peuvent inclure des augmentations de primes ou des ajustements des franchises, en fonction du risque perçu dans différentes régions.

L'ensemble des études réalisées pour ce mémoire portera sur ce contrat, avec en particulier la garantie TGN.

1.2.2 Bilan 2022 des événements de grêle : étude France Assureurs

L'objectif de cette partie est de mettre en évidence l'augmentation significative du risque de grêle dans le secteur assurantiel en France au cours des dernières années, en portant une attention particulière à l'année 2022, marquée par une sinistralité particulièrement élevée. Les conclusions de l'étude réalisée par France Assureurs, *L'assurance des événements naturels en 2022* [17], seront analysées. Elle fournit des données essentielles sur l'impact des événements de grêle durant cette année. Cette enquête, couvrant 88% des cotisations perçues sur le marché national, nous permettra de mieux comprendre la dynamique de ce risque et ses implications pour les assureurs.

Évaluation des cotisations en 2022

En 2022, la **couverture des événements naturels** a généré un volume total de cotisations de 4,9 milliards d'euros, réparti ainsi : 39 % pour les catastrophes naturelles, **38 % pour la garantie TGN sur les bâtiments**, 9 % pour la garantie TGN sur les véhicules, et 13 % pour les assurances agricoles. Les cotisations liées à la garantie TGN pour les bâtiments sont donc estimées à 1,86 milliard d'euros, dont 1,3 milliard provient des particuliers, au niveau national (hors assurances automobile et agricoles), marquant une hausse de 5,1 % par rapport à 2021, où elles s'élevaient à 1,77 milliard d'euros. Cela représente **8,7 % de l'ensemble des cotisations des assurances de dommages aux biens** (hors assurances récoltes).

Impact économique de la sinistralité TGN

En 2022, la sinistralité de la garantie TGN atteint 4,67 milliards d'euros, marquant le niveau de charges le plus élevé depuis les tempêtes de 1999. Ce montant représente une augmentation spectaculaire de +386 % par rapport à 2021.

Parmi les trois aléas couverts par la garantie TGN, la grêle se classe généralement derrière la tempête en termes de coût annuel. Toutefois, certaines années, comme 2022, se démarquent par leur caractère exceptionnel : la grêle représente deux tiers des charges liées à la sinistralité TGN, soit un total de 3,32 milliards d'euros. Cette seule année pour la grêle représente un tiers des indemnités totales de la garantie TGN sur la période 2018-2022. Ce montant reflète une augmentation de la sinistralité de la grêle spectaculaire de +1 060 % par rapport à 2021 (3 317 millions d'euros en 2022 vs 286 millions d'euros en 2021). La sinistralité due à la grêle en 2022 établit un record absolu depuis 1984, dépassant de 4,7 fois le précédent sommet atteint en 2014 (713 millions d'euros actualisés).

Année	Montant des indemnités (en M€) (estimation marché national)				
	Tempête	Grêle	Neige sur toitures	Total en euros courants	Total en euros constants 2022 *
2018	810	390	75	1 275	1 459
2019	810	530	30	1 370	1 556
2020	884	97	9	990	1 121
2021	631	286	43	960	1 038
2022	1 350	3 317	3	4 670	4 670
Total 2018 – 2022	4 485	4 620	160	9 265	9 845
Total depuis 1990	23 085	7 845	1 170	32 100	49 103
Total depuis 1984	24 320	8 040	1 300	33 660	53 686

* réévalué par l'indice FFB

FIGURE 22 – Historique des indemnités versées depuis 5 ans, source : [17]

Les épisodes de grêle sont souvent localisés, comme ceux observés entre mai et août 2022, qui ont particulièrement affecté les régions de Nouvelle-Aquitaine et d'Auvergne-Rhône-Alpes. Dans ces régions, les coûts moyens sont très élevés, concentrant ainsi plus de la moitié des indemnités versées en 2022.

Régions	Ensemble des assurés			Particuliers	
	Part des indemnités versées (en %)	Coût moyen (en euros)	S/P (en %)	Fréquence des sinistres (en ‰)	Coût moyen (en euros)
Auvergne-Rhône-Alpes	24,1	9 372	525	17,3	6 707
Bourgogne-Franche-Comté	9,2	7 694	492	24,7	5 199
Bretagne	0,7	2 189	33	4,3	1 654
Centre-Val-de-Loire	10,7	7 652	586	31,2	5 117
Corse	1,1	4 814	509	31,2	3 754
Grand-Est	2,1	3 189	77	7,8	2 133
Hauts-de-France	9,9	2 601	349	44,1	2 064
Île-de-France	3,0	5 576	58	2,9	3 609
Normandie	1,3	2 791	70	8,0	2 153
Nouvelle-Aquitaine	30,3	8 215	606	30,3	6 582
Occitanie	4,3	3 493	107	9,5	2 736
Pays de La Loire	1,3	3 799	62	5,2	2 670
Provence – Alpes – Côtes d'Azur	1,9	2 593	68	6,6	2 099
Ensemble des régions	100	5 819	263	15,2	4 332
Rappel France entière	100	5 655	251	15,5	4 275

FIGURE 23 – Paramètres de la sinistralité par région en 2022, source : [17]

Evolution de la fréquence et du coût moyen de la grêle : enjeux croissants pour l'assurance

Le graphique ci-dessous illustre l'évolution de la fréquence et du coût moyen des sinistres liés à la grêle depuis 1990. Il met en évidence une augmentation progressive du coût moyen des sinistres, qui s'élevait autour de 2 200 euros dans les années 1990, avant de passer à environ 3 000 euros dans les années 2000 et 2010. Depuis 2018, cette tendance s'est nettement accélérée, avec les 4 années les plus coûteuses enregistrées depuis 1990, culminant en 2022 avec un coût moyen exceptionnel de 8 830 euros.

En ce qui concerne la fréquence des sinistres, des fluctuations importantes se manifestent d'une année à l'autre, bien qu'une tendance générale à la hausse soit observable. L'année 2022 se distingue particulièrement, enregistrant une fréquence record de 7,3 ‰, la plus élevée depuis 1990.

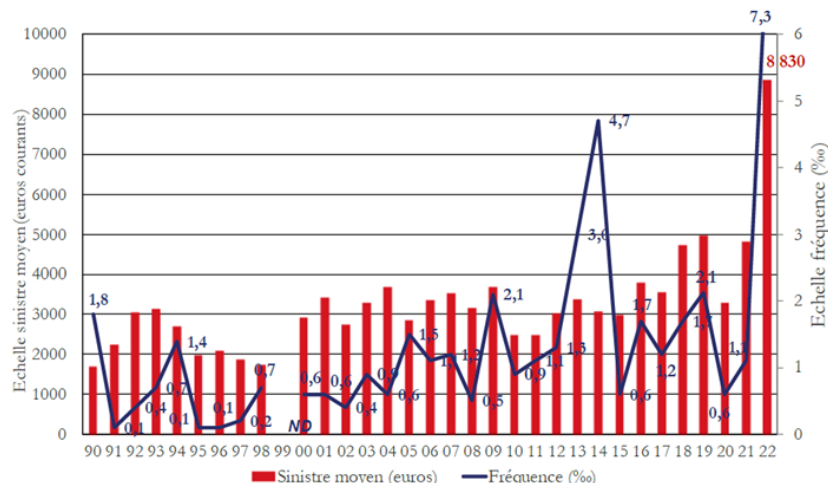


FIGURE 24 – Historique de la fréquence et du coût moyen grêle depuis 1990, source : [17]

Ainsi, il est mis en évidence l'importance croissante de la gestion du risque grêle dans le secteur de l'assurance, ainsi que la nécessité de développer des solutions innovantes pour mieux anticiper et couvrir ce phénomène climatique. Il souligne également l'ampleur des défis auxquels les assureurs sont confrontés, tant en matière de tarification que de gestion du risque (en ayant par exemple recours à la réassurance), dans un contexte de changement climatique qui exacerbe la fréquence et la gravité des épisodes de grêle.

1.2.3 Grêle en France : défis climatiques et réponses assurantielles

L'année 2022 a mis en évidence l'ampleur des impacts économiques liés aux sinistres causés par la grêle, avec des chiffres sans précédent qui soulignent la vulnérabilité croissante face à cet aléa naturel. Cette évolution préoccupante, soulignée par l'étude de France Assureurs, s'inscrit dans un contexte plus large marqué par les changements climatiques. L'augmentation de la fréquence et de l'intensité des épisodes de grêle reflète une tendance générale au dérèglement climatique, posant de nouveaux défis pour le secteur assurantiel, confronté à une charge croissante. Selon la Caisse Centrale de Réassurance (CCR), **la sinistralité liée aux événements climatiques extrêmes pourrait augmenter de 27% à 62% d'ici 2050** [18]. Ce changement impose au système assurantiel de s'adapter pour maintenir son équilibre financier tout en continuant à jouer un rôle sociétal fondamental.

Adaptations économiques et rôle sociétal face aux catastrophes naturelles

Sur le plan économique, la recrudescence des sinistres liés aux catastrophes naturelles a entraîné une hausse des primes d'assurance et une révision des contrats. Les assureurs ont par ailleurs développé des outils d'aide à la modélisation, tels que les zoniers, pour évaluer les risques et tarifier plus justement selon l'exposition des communes, voire de l'adresse. Cette sophistication soulève toutefois des interrogations quant à l'accessibilité, en particulier dans les zones à risque élevé, où les primes peuvent devenir dissuasives, menaçant l'équité d'accès à la couverture assurantielle.

En plus des enjeux économiques, l'assurance joue un rôle sociétal clé en atténuant les impacts financiers des catastrophes. Elle protège les ménages, les agriculteurs et les entreprises contre les pertes subies, contribuant ainsi à la résilience des communautés. En l'absence de compensations, des faillites ou des crises économiques locales pourraient se multiplier, notamment dans les secteurs les plus vulnérables. Ce rôle pousse les assureurs à adapter leurs offres pour proposer des solutions accessibles et mieux alignées sur les besoins de leurs clients.

Le régime d'indemnisation face aux catastrophes naturelles

En France, un régime d'indemnisation spécifique a été mis en place pour protéger les assurés contre les conséquences financières des catastrophes naturelles. Cependant, ce régime présente aujourd'hui un déséquilibre structurel. Un écart estimé à 1,2 milliard d'euros par an entre les ressources du régime et ses charges met en lumière une pression croissante sur le système assurantiel [10]. Ce déséquilibre conduit au retrait progressif des assureurs des zones les plus à risque et complique la capacité de la CCR et de l'État à garantir les financements nécessaires. Bien que la grêle ne soit pas directement intégrée dans ce régime, les problématiques soulevées par ce déséquilibre s'appliquent également au risque grêle car elle constitue également un risque climatique.

Recommandations pour renforcer l'assurabilité des risques climatiques

Dans ce contexte, un rapport a été commandé par les Ministères de l'Economie, des Finances et de la Souveraineté industrielle et numérique et de la Transition écologique pour formuler des recommandations visant à renforcer l'assurabilité des risques climatiques [19]. Ce rapport, nommé également rapport Langrenay, propose 11 objectifs déclinés en 37 recommandations, et s'articule autour de deux axes principaux : l'adaptation du système assurantiel et l'accélération de la contribution de l'assurance aux efforts d'adaptation. Parmi les principales propositions, on trouve :

- Le rééquilibrage financier du régime d'indemnisation, avec une augmentation des primes Cat Nat ;
- Le maintien d'un système d'assurance protecteur avec un partage des risques, la CCR s'associant à la réassurance privée pour les risques de pointe tels que les inondations, les cyclones, les séismes ;
- Un système de bonus-malus fiscal incitant les assureurs à assurer les zones à risque élevé. Les assureurs prenant en charge ces zones bénéficieraient de bonus fiscaux, tandis que ceux se retirant des zones à risque se verraient pénalisés ;
- Une meilleure régulation, avec un renforcement du rôle de l'Autorité de Contrôle Prudentiel et de Résolution (ACPR) pour superviser les plans de transition climatique des assureurs et réassureurs ;
- Un encouragement à la prévention, avec l'élaboration d'un catalogue d'actions d'adaptation pour orienter les mesures de prévention ainsi que la création d'un fonds dédié au financement des mesures de prévention individuelles et collectives, alimenté par des cotisations supplémentaires liées à la garantie catastrophes naturelles ;
- L'atténuation des effets du changement climatique grâce à la mise en place d'offres d'assurance plus respectueuses de l'environnement.

Ce rapport alimentera les réflexions du mémoire dans la 3ème partie, car ces recommandations marquent une avancée importante pour assurer la pérennité des risques climatiques. Toutefois, cet objectif ne pourra être atteint sans un engagement conjoint de l'État, des assureurs et des assurés.

En conclusion, l'adaptation du système assurantiel face aux risques liés à la grêle en France souligne l'interdépendance entre les défis économiques et les enjeux sociaux. L'assurance, en tant qu'outil essentiel de gestion des risques, doit évoluer pour répondre aux bouleversements climatiques, tout en garantissant une protection juste et durable pour l'ensemble de la population.

1.3 Objectifs et enjeux du mémoire

Après avoir exploré le phénomène de la grêle dans le contexte de l'assurance, avec une présentation des garanties existantes, du bilan de l'année 2022, des défis climatiques à venir et des réponses assurantielles associées, il convient de définir les objectifs et enjeux de ce mémoire.

La gestion d'un péril naturel tel que la grêle repose sur une compréhension approfondie de ses mécanismes, de sa fréquence, de son impact sur les expositions assurées, et sur les moyens de l'anticiper. Ce mémoire s'inscrit dans cette démarche en développant un outil concret et opérationnel : **un zonier de risque grêle**, conçu pour la région Auvergne-Rhône-Alpes dans le cadre de l'assurance MRH. Cet outil servira de base à des applications pratiques en **tarification, souscription, et gestion des risques climatiques**.

Cette section présente les **objectifs** et les **enjeux** de ce mémoire.

1.3.1 Objectifs

Les objectifs de ce mémoire sont :

- **Créer un zonier sur le risque grêle** : En assurance IARD⁵, la zone géographique constitue un critère de segmentation fondamental. La sinistralité dépend fortement de l'environnement géographique, météorologique et sociodémographique. Le découpage du territoire en zones de risque homogènes, appelé zonier, permet une gestion différenciée et optimisée des portefeuilles.
- **Améliorer la précision tarifaire** : Utiliser le zonier pour adapter les primes d'assurance à la fréquence et à la gravité des sinistres spécifiques à chaque zone. L'objectif est d'assurer une équité tarifaire et de réduire les déséquilibres techniques.
- **Élaborer une politique de souscription** : Tester et proposer des stratégies de souscription basées sur le zonier. Cela inclut la gestion des expositions, le contrôle des coûts et le ciblage de zones spécifiques pour équilibrer le portefeuille.
- **Anticiper des évolutions climatiques et réglementaires** : Simuler des scénarios prospectifs pour évaluer l'impact du dérèglement climatique et des modifications réglementaires sur la sinistralité et la gestion du portefeuille.

1.3.2 Enjeux

Après avoir défini les objectifs de ce mémoire, il est essentiel de mettre en lumière les enjeux qu'il aborde, afin de mieux en apprécier l'importance et la pertinence dans le contexte actuel.

Les enjeux de ce mémoire s'inscrivent dans un contexte où la problématique de la grêle, bien qu'essentielle, reste largement sous-explorée, en particulier dans le cadre des politiques assurantielles. Pourtant, les données récentes et les projections climatiques mettent en évidence une tendance alarmante : une augmentation probable de la fréquence et surtout de l'intensité des épisodes de grêle [22]. Cette évolution constitue un défi majeur pour le secteur assurantiel : comment anticiper et adapter ses pratiques à une réalité climatique en constant mouvement, tout en garantissant la viabilité économique de ses opérations ?

Un des enjeux centraux de ce mémoire réside dans la prise en compte de l'engagement tripartite entre assureurs, assurés et régulateurs. Cela est indispensable pour garantir une couverture durable et équitable des risques climatiques, en particulier dans les zones à forte sinistralité. Les assureurs doivent trouver un juste équilibre entre la gestion des risques et la maîtrise des coûts, les assurés doivent avoir accès à une protection financière abordable, et les régulateurs doivent veiller à prévenir tout désengagement des assureurs dans les régions les plus vulnérables. Ce mémoire propose une réflexion stratégique pour concilier ces trois dimensions et contribuer à un modèle de gestion des risques plus résilient.

5. Incendie, Accident et Risques Divers

Par ailleurs, ce mémoire intègre les contraintes économiques et réglementaires auxquelles fait face le secteur assurantiel. Alors que les coûts liés aux sinistres de grêle augmentent considérablement, il devient crucial de concevoir des solutions qui soient à la fois financièrement viables pour les assureurs et socialement acceptables pour les assurés.

Enfin, ces travaux ont été réalisés dans le cadre d'une mission pour un acteur du secteur assurantiel. Cette dimension opérationnelle confère au mémoire une portée concrète, avec des analyses et des recommandations directement applicables dans la pratique. Cependant, pour des raisons de confidentialité, aucune information relative à ce client ne sera donnée.

2 Construction d'un zonier grêle

Après avoir étudié les spécificités du phénomène de grêle et son impact en assurance, la deuxième partie se concentre sur la construction d'un zonier grêle. Ce dernier correspond à l'élaboration d'un score de risque, représenté graphiquement en tenant compte des contraintes géographiques. Dans le cadre de ce mémoire, le score est basé sur les primes pures et permet une classification des zones selon leur niveau d'exposition.

Cette partie débute par une présentation et un traitement des données utilisées, avant d'explorer trois approches méthodologiques : la méthode par itération, les modèles linéaires généralisés et la théorie de la crédibilité. Chaque approche est analysée en détail, avec une comparaison finale permettant de déterminer le zonier le plus adapté.

2.1 Présentation et traitement des données

2.1.1 Sources des données : historiques des sinistres, portefeuilles, météorologiques, INSEE

Dans le cadre des travaux, 3 bases sont mises à disposition :

La base de données "sinistralité" de la garantie TGN

La base de données des sinistres contient uniquement les sinistres de la compagnie d'assurance relatifs à la garantie TGN. Chaque ligne, présente un unique sinistre avec ses caractéristiques détaillées. Un tableau récapitulatif des variables utiles de cette base de données est présenté ci-dessous :

Variables	Description	Exemple
DEPT	Département du sinistre	42
CINSEE	Code INSEE de la commune de survenance	42095
année survenance	Année de survenance du sinistre	2019
description sinistres	Courte description du sinistre	GRELE DPT 42
date survenance	Date de survenance (et non de déclaration) du sinistre	15/06/2019
NB pièces	Nombre de pièces du logement (De 1 à 15 puis plus de 16)	6
Charge totale	Montant total du sinistre	4465,38
Période Const Bâtiment	Intervalle sur l'année de construction du bâtiment	De 1951 à 2000
Exposition	Valeur comprise dans l'intervalle $]0; 1]$. Exposition de l'assuré durant l'année de survenance. Une exposition de 0,5 signifie que l'assuré victime du sinistre a été assuré 6 mois dans l'année de survenance.	1
Estimation cotis annuel	Prime de l'assuré pour ce contrat durant l'année du sinistre	183,6

TABLE 1 – Variables de la base "sinistres" de la compagnie d'assurance

Le portefeuille d'assurés lié à cette garantie

Ensuite, la base de données du portefeuille contient l'ensemble des logements inclus dans le portefeuille d'assurance habitation de la compagnie d'assurance. Cette base est présentée sous forme d'agrégat des biens ayant les mêmes caractéristiques par année. Les caractéristiques se retrouvent dans les variables listées dans le tableau ci-dessous :

Variables	Description	Exemple
Année	Année d'exposition du portefeuille	2019
NB pièces	Nombre de pièces du logement (De 1 à 15 puis plus de 16)	6
CREG6 051	Code INSEE de la commune	42095
nbcontimp	Nombre d'habitations assurées ayant ces caractéristiques sur l'année. Il faut noter que ce nombre n'est pas forcément entier si l'habitation n'était pas assurée pendant toute l'année	5,81
cotis imp	Somme des primes des assurés ayant ces caractéristiques sur l'année	2574,62
Nature Habitation	Nature de l'habitation des assurés	Maison

TABLE 2 – Variables de la base "portefeuille"

Sachant que les données du portefeuille couvrent les années 2019 à 2023, voici quelques chiffres sur ces différentes années avant retraitement :

Année	Nombre de lignes	Dont maison	Nombre de logements assurés	Dont maison
2019	231 149	168 655 (73%)	472 698	269 957 (57%)
2020	232 876	167 959 (72%)	494 526	270 397 (55%)
2021	231 776	165 854 (72%)	500 642	263 881 (53%)
2022	227 212	161 461 (71%)	505 776	256 741 (51%)
2023	220 900	157 080 (71%)	502 944	247 574 (49%)

TABLE 3 – Évolution des données d'assurés sur les années 2019-2023

Tout d'abord, le portefeuille est relativement stable au fil du temps. Ensuite, ce tableau montre qu'environ la moitié du portefeuille est constituée de maisons, sur lesquelles se concentreront exclusivement les travaux. En effet, les maisons sont particulièrement exposées à la grêle : sur la période 2019-2023, 25 922 sinistres sont enregistrés au titre de la garantie TGN, dont 23 784 sur des maisons. Ce chiffre correspond à environ 2% des maisons assurées ayant subi un sinistre.

A noter que le dénombrement des biens assurés est non cumulatif d'une année à une autre. En d'autres termes, une maison assurée sur la période de 2019 à 2021 est recensée trois fois dans le tableau ci-dessus.

Les données météorologiques

Enfin, la dernière base recense des données météorologiques répertoriant les phénomènes de grêle. Elle est gérée par l'Observatoire français des tornades et orages violents, également appelé Keraunos. L'objectif principal de Keraunos est de collecter, analyser et diffuser des informations sur les événements météorologiques violents tels que les tornades, les orages, les tempêtes, les grêlons, et autres phénomènes extrêmes en France. Cette base de données vise à améliorer la connaissance et la prévision de ces phénomènes pour mieux protéger les personnes et les biens.

Concernant la grêle, les bases de données Keraunos répertorient les épisodes de grêle survenus par commune de 2006 à 2023. Elles recensent plus de 100 000 chutes de grêle provenant de différentes sources :

- **Observations directes** : témoignages, photos et vidéos d'épisodes de grêle
- **Données radar** : informations radar pour analyser les structures orageuses
- **Données satellitaires** : images et informations provenant de satellites météorologiques

Cependant, il est important de noter que les événements recensés avant 2019 ne sont pas jugés fiables par le client. Les données collectées et disponibles dans la base de données sont détaillées dans le tableau suivant :

Variables	Description	Exemple
Date	Date de survenance de l'évènement	12/07/2023
Code INSEE	Code INSEE de la commune	38004
Gros grêlons	Taille de plus de 2 cm (1 = Oui, 0 = Non)	1
Valeur basse estimée	Diamètre minimal des grêlons observé (en cm)	1
Valeur haute estimée	Diamètre maximal des grêlons observé (en cm)	3
Classe	De A1 à A5 en fonction de la classification des grêlons proposée par l'Anelfa	A3

TABLE 4 – Variables de la base Keraunos

À noter que la base Keraunos ne répertorie pas les épisodes A0. Cela n'est pas problématique pour la réalisation des travaux puisque ces grêlons ont un diamètre inférieur à 1 mm. Par conséquent, les épisodes A0 ne font que très peu de dégâts sur les habitations, et sont supposés dans le cadre de cette étude comme ne causant aucun sinistre.

Des bases de données en open source, principalement issues du site data.gouv.fr, sont également utilisées. Bien que leur contenu ne soit pas présenté de manière détaillée, elles fournissent des informations précieuses telles que le nombre de maisons et d'appartements par commune, les coordonnées géographiques des localités, ainsi que les regroupements administratifs résultant de fusions de communes. Ces données constituent un complément utile à l'analyse.

2.1.2 Retraitements

Après la présentation des trois principales bases de données disponibles, plusieurs retraitements classiques ont été réalisés. Parmi eux, l'homogénéisation des départements, certains étant indiqués par leur numéro et d'autres par leur nom, ou encore la modification du code INSEE des communes disparues, remplacées par de nouvelles. L'un des retraitements les plus importants a concerné la base des sinistres. À l'origine, cette base contenait l'ensemble des sinistres de la région pour la garantie TGN, incluant les sinistres liés à la grêle, aux tempêtes et à la neige.

Traitement et sélection des données sinistres

Une première étape consiste à exploiter la colonne description sinistres pour ne conserver que ceux mentionnant le mot "grêle". Toutefois, une grande partie de ces sinistres est décrite comme "Tempête, Grêle, Neige", rendant complexe l'extraction des sinistres spécifiquement liés à la grêle. Pour isoler ces derniers, il est supposé qu'un sinistre relevant de la garantie TGN est dû à un épisode de grêle s'il contient le mot "grêle" dans sa description et s'il correspond à un événement de grêle recensé dans la base Keraunos le même jour ou jusqu'à cinq jours avant. Ce délai de cinq jours a été déterminé en collaboration avec le client, en tenant compte du délai légal de déclaration de sinistre dans le cadre de la garantie TGN et de leur process de gestion.

Par exemple, un sinistre survenu le 28 mai 2023 dans la commune 01034 est retenu comme sinistre grêle s'il y a eu un événement de grêle dans cette commune entre le 24 et le 28 mai 2023. En cas de plusieurs événements survenus dans ce laps de temps, l'événement le plus proche de la date déclarée par l'assuré a été considéré comme celui ayant causé le sinistre.

Retraitement des valeurs extrêmes et aberrantes

En plus de cela, les valeurs extrêmes et aberrantes sont retraitées afin, d'une part, de garantir la fiabilité des données pour les valeurs aberrantes, et, d'autre part, de répondre au besoin du client dans le temps imparti. La modélisation spécifique des valeurs extrêmes dans le cadre d'un zonier constitue une étape complexe qui, sur demande du client, dépasse le périmètre de ce mémoire. Deux situations principales sont identifiées :

- **Sinistres avec un coût élevé :** Ces sinistres, considérés comme des valeurs extrêmes, peuvent biaiser les résultats des régressions. Pour éviter cet effet, ils sont isolés et ne sont pas inclus dans les modèles présentés dans la suite.
- **Sinistres mal attribués :** Certains sinistres liés à des tempêtes peuvent être incorrectement catégorisés comme des sinistres de grêle. Par exemple, un sinistre dont la description est "Tempête, Grêle, Neige" est classé comme grêle si un événement répertorié par Keraunos se produit dans un intervalle de 5 jours. Cependant, il est possible qu'il s'agisse en réalité d'un sinistre dû à une tempête. Par exemple, un sinistre de 100 000 € avec des grêlons de taille A1 pourrait en réalité être attribué à une tempête, même si des grêlons étaient présents le même jour. Ces cas sont considérés comme aberrants et sont donc exclus des analyses.

Identification des valeurs aberrantes à partir de régressions linéaires simples

Tout d'abord, l'identification des valeurs aberrantes est réalisée à partir d'une appréciation des boîtes à moustaches. Toutefois, cette méthode n'est pas suffisamment robuste pour tirer des conclusions fortes. Ainsi, une seconde option est retenue – le recours aux régressions linéaires. Elles sont effectuées sur différentes variables telles que le nombre de pièces, la classe de l'événement, et le département, afin de déterminer des variables explicatives du coût d'un sinistre grêle. Deux variables explicatives significatives sont identifiées : le nombre de pièces et la classe de l'événement de grêle.

Passage à une régression linéaire multiple

Pour tenir compte de ces deux variables explicatives, une régression linéaire multiple a été réalisée. Cette régression permet de prédire le coût du sinistre en fonction du nombre de pièces de la maison et de la taille des grêlons responsables. Les résultats obtenus sont les suivants : $\beta_{classe} = 2084,97$ et $\beta_{nbpieces} = 729,10$. Cela signifie qu'une pièce supplémentaire entraîne une augmentation de 729,10 € dans la prédiction du coût du sinistre, tandis qu'une augmentation d'une classe de sinistre se traduit par une hausse de 2084,97 € dans la prédiction. Les résultats complets de la régression sont présentés dans la capture d'écran ci-dessous.

OLS Regression Results						
=====						
Dep. Variable:	charge total	R-squared:	0.038			
Model:	OLS	Adj. R-squared:	0.038			
Method:	Least Squares	F-statistic:	181.0			
Date:	Mon, 16 Sep 2024	Prob (F-statistic):	8.05e-78			
Time:	12:17:59	Log-Likelihood:	-1.0181e+05			
No. Observations:	9107	AIC:	2.036e+05			
Df Residuals:	9104	BIC:	2.037e+05			
Df Model:	2					
Covariance Type:	nonrobust					
=====						
	coef	std err	t	P> t	[0.025	0.975]

const	-1583.1857	699.595	-2.263	0.024	-2954.549	-211.823
intensite_encoded	2084.9683	115.108	18.113	0.000	1859.331	2310.605
NB pieces	729.0998	101.157	7.208	0.000	530.810	927.390
=====						
Omnibus:	13976.212	Durbin-Watson:	0.117			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	18901725.887			
Skew:	9.211	Prob(JB):	0.00			
Kurtosis:	225.425	Cond. No.	25.4			
=====						

FIGURE 25 – Résultats de la régression linéaire multiple

Détection des valeurs aberrantes avec la Distance de Cook

À partir de cette régression, les résidus, représentant la différence entre le coût réel et la valeur prédite, sont calculés pour chaque sinistre. Afin d'identifier les valeurs aberrantes, la distance de Cook est utilisée. Cette mesure est couramment employée en statistique pour évaluer l'influence de chaque point de données sur le modèle global. Elle combine l'effet de l'écart résiduel et la position de l'observation dans l'espace des prédictors. Une valeur élevée de la distance de Cook signale qu'une donnée exerce une forte influence sur le modèle, suggérant ainsi qu'elle pourrait être une valeur aberrante. Introduite par le statisticien R. Dennis Cook [25], cette approche s'est imposée comme un outil standard dans l'analyse de régression.

La distance de Cook est calculée selon la formule suivante :

$$D_i = \frac{\sum_{j=1}^n \left(\hat{Y}_j - \hat{Y}_j^{(i)} \right)^2}{(p+1)\sigma^2}$$

Où :

- $\hat{Y}_j$ représente la prédiction du modèle de régression complète pour l'observation j ,
- $\hat{Y}_j^{(i)}$ est la prédiction pour l'observation j à partir d'un modèle de régression ajusté sans l'observation i ,
- p est le nombre de coefficients de régression,
- σ^2 est l'erreur quadratique moyenne.

Concrètement, la distance de Cook est calculée en excluant le i -ème point de données du modèle et en recalculant la régression. Les valeurs du modèle de régression sont ensuite examinées pour détecter les changements provoqués par la suppression du point. Cette méthode permet d'évaluer l'influence de chaque observation sur le modèle.

Enfin, pour déterminer si une observation est une valeur aberrante, une règle empirique, classiquement utilisée dans le cas de la distance de Cook est appliquée : tout point dont la distance de Cook est supérieure à $\frac{4}{n}$, où n est le nombre d'observations, est considéré comme une valeur aberrante.

Ainsi, parmi les 9182 sinistres analysés, 338 ont été retirés (soit 3,68%) car jugés non pertinents pour cette étude. Bien que ce nombre de sinistres ne soit pas négligeable, il reste acceptable dans ce contexte. Un travail spécifique sur ces sinistres à l'aide de la théorie des valeurs extrêmes pourrait être intéressant pour compléter l'étude. Toutefois, une telle approche dépasse le cadre de ce mémoire et n'a donc pas été menée.

2.1.3 Qualité et fiabilité des données

Une fois les retraitements effectués, les données obtenues sont fiables et de qualité. En effet, l'assureur client veille rigoureusement à leur intégrité grâce à des processus de contrôle réguliers, conçus pour détecter et corriger les erreurs, notamment les doublons. Ces bases, enrichies par des sources externes telles que la base Keraunos, offrent une base solide pour le développement d'un zonier grêle.

Fiabilité et pertinence des données de la base Keraunos

La base Keraunos occupe une place centrale dans cette étude et nécessite une évaluation rigoureuse de sa fiabilité. Depuis 2006, Keraunos recense systématiquement les chutes de grêle à l'échelle nationale grâce à un réseau croissant d'observateurs, composé de chasseurs d'orages, de passionnés de météorologie, et de contributeurs occasionnels. Ces observateurs mobiles fournissent des données détaillées sur les épisodes de grêle, incluant les communes touchées, le diamètre des grêlons (mesuré ou estimé à partir de photographies), ainsi que des informations complémentaires telles que les accumulations au sol ou les dégâts observés.

Ces rapports sont enrichis par des articles de presse et des analyses d'images radar, permettant de retracer la trajectoire des cellules orageuses et d'extrapoler les communes potentiellement affectées. Bien que certaines zones montagneuses soient moins documentées en raison du manque d'observateurs, la couverture demeure étendue et précise.

La base Keraunos couvre la période 2006-2023 avec près de 50 000 événements recensés entre 2006 et 2020. Les méthodes uniformes appliquées garantissent l'homogénéité des données, malgré une amélioration progressive de l'exhaustivité des observations, particulièrement pour les grêlons de moins de 2 cm. Des biais potentiels liés au sous-recensement peuvent exister mais ils sont jugés marginaux et n'altèrent pas significativement les analyses climatologiques effectuées à des échelles plus larges, telles que l'arrondissement ou le département. Malgré ses limites, la base Keraunos est l'une des sources les plus complètes et fiables disponibles pour caractériser l'exposition au risque grêle en France, contribuant à une meilleure compréhension et gestion de ce phénomène.

Enfin, la fiabilité des données de Keraunos est attestée par leur utilisation dans l'étude de la MRN sur le risque grêle. La MRN, qui regroupe 15 assureurs représentant 88 % du marché assurantiel français, a pour mission d'évaluer le niveau de risque des communes pour divers périls naturels (inondations, sécheresses, grêle, etc.). Si une organisation aussi rigoureuse et reconnue que la MRN s'appuie sur cette base pour ses analyses, cela confirme son sérieux et sa pertinence dans le cadre d'études sur le risque grêle. [21]

2.2 Première approche : construction d'un zonier par itération

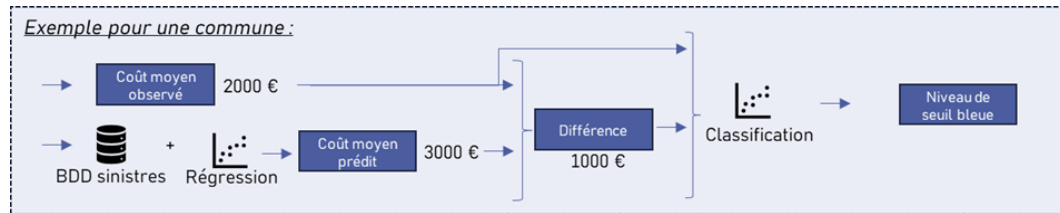
Après la présentation et le retraitement des données, une première approche simpliste est testée afin de poser les bases du projet et de mieux appréhender les outils disponibles. Cette étape initiale a pour objectif d'explorer une approche préliminaire et rapide permettant de combiner différents indicateurs, tout en restant adaptable aux contraintes du calendrier du client.

2.2.1 Principe de la méthode

La première approche consiste à créer un zonier combinant l'intensité et la fréquence des épisodes de grêle. Ce zonier est spécifiquement élaboré pour le département 42, en s'inspirant directement de la méthodologie employée par MRN pour concevoir un zonier dédié au risque grêle. Cette étude, non disponible en open source, a été fournie par le client assureur. Elle révèle que la cartographie du seuil de vigilance grêle de la MRN résulte du croisement des indicateurs "fréquence" et "intensité". [9].

Zonier intensité

Tout d'abord, le zonier intensité est basé sur la taille maximale des grêlons observée par commune. Cette information est croisée avec un calcul d'incertitude sur le coût moyen. Ce calcul repose sur la différence entre le coût moyen réel d'une commune et le coût moyen prédit selon les caractéristiques des biens de cette commune, à l'aide d'une régression linéaire sur la base de données des sinistres. Les résultats sont ensuite classés par quantile pour attribuer un niveau de risque allant de très prudent à très risqué.



Une fois le score d'une commune établi en fonction de la taille maximale du grêlon d'un côté, et du résidu du coût moyen d'un autre côté, le tableau suivant est utilisé pour attribuer une note à la commune dans le cadre du zonier intensité :

		Taille des grêlons					
		A0	A1	A2	A3	A4	A5
		$x < 1\text{cm}$	$1\text{cm} \leq x < 2\text{cm}$	$2\text{cm} \leq x < 3\text{cm}$	$3\text{cm} \leq x < 4\text{cm}$	$4\text{cm} \leq x < 5\text{cm}$	$5\text{cm} \leq x$
Incertitude du coût moyen	Très prudent	TRES BAS	TRES BAS	BAS	MOYEN	HAUT	TRES HAUT
	Prudent	TRES BAS	TRES BAS	BAS	MOYEN	HAUT	TRES HAUT
	Moyen	TRES BAS	TRES BAS	BAS	MOYEN	HAUT	TRES HAUT
	Risqué	TRES BAS	TRES BAS	MOYEN	HAUT	HAUT	TRES HAUT
	Très risqué	TRES BAS	TRES BAS	MOYEN	HAUT	TRES HAUT	TRES HAUT
	Absence	TRES BAS	TRES BAS	MOYEN	HAUT	TRES HAUT	TRES HAUT

FIGURE 26 – Grille pour attribuer la note du zonier intensité

Ce tableau reprend celui de la MRN concernant la taille maximale des grêlons, auquel une variable d'incertitude du coût moyen est ajoutée. L'incertitude du coût moyen augmentait la notation préconisée par la MRN pour les cases en bleu ciel dans le tableau. Cette justification, bien que qualitative, repose sur des discussions avec le client, considérées comme des jugements d'experts fondés sur leur connaissance du portefeuille.

Zonier fréquence

Le zonier intensité présenté ci-dessus est complété par un zonier fréquence. Ce dernier s'intéresse au nombre d'événements de grêle par commune enregistrés annuellement dans la base Keraunos, en distinguant ceux ayant généré des sinistres. À partir de ces données, la grille suivante a été établie :

			Nombre d'évènements enregistré dans Keraunos/an				
			Inférieur à 1	Entre 1 et 2	Entre 2 et 3	Entre 3 et 5	Plus de 5
			Tres bas	Bas	Moyen	Haut	Tres haut
Nombre d'évènements ayant causé au moins un sinistre dans notre portefeuille/an	0	Tres bas	Tres bas	Bas	Bas	Moyen	Moyen
	Entre 0 et 1	Bas	Tres bas	Bas	Moyen	Moyen	Haut
	Entre 1 et 2	Moyen		Bas	Moyen	Haut	Haut
	Entre 2 et 3	Haut			Moyen	Haut	Tres haut
	Plus de 3	Tres haut				Haut	Tres haut

FIGURE 27 – Grille pour attribuer la note du zonier fréquence

La fréquence du phénomène grêle est déterminée par le nombre moyen d'événements recensés par an dans les communes, conformément à la méthode de calcul utilisée par la MRN. Afin de mieux refléter les spécificités du portefeuille, les seuils sont ajustés en tenant compte du nombre d'événements ayant effectivement impacté les assurés.

Zonier final

Le zonier final est obtenu en combinant les zoniers d'intensité et de fréquence décrits précédemment. Cette combinaison s'appuie sur la grille de croisement utilisée par la MRN, présentée ci-dessous :

		Intensité				
		TRES BAS	BAS	MOYEN	HAUT	TRES HAUT
Fréquence	Très bas	Très bas	Bas	Moyen	Haut	Très haut
	Bas	Très bas	Bas	Moyen	Haut	Très haut
	Moyen	Très bas	Bas	Moyen	Haut	Très haut
	Haut	Très bas	Bas	Moyen	Haut	Très haut
	Très haut	Bas	Moyen	Haut	Très haut	Très haut

FIGURE 28 – Grille pour combiner intensité et fréquence et attribuer la note du zonier final

La grille montre que l'intensité influe de manière prépondérante sur le zonier final, la fréquence n'intervenant que lorsqu'elle est classée en "Très haut". Une des principales limites de cette méthode réside dans le fait qu'une commune ayant enregistré un seul grêlon de taille A5 depuis 2019 sera directement classée en "Très haut", indépendamment de la survenance de sinistres ou de la fréquence des événements.

2.2.2 Résultats obtenus

Cette première approche, réalisée sur le département 42, permet de produire une première cartographie. Ce premier zonier offre une meilleure compréhension du phénomène de la grêle ainsi que des bases de données disponibles. Les résultats croisés entre le client et la MRN étant cohérents, il est donc pertinent d'utiliser les bases de données du client pour faire des études plus poussées. Le zonier réalisé pour le département 42 est présenté ci-dessous.

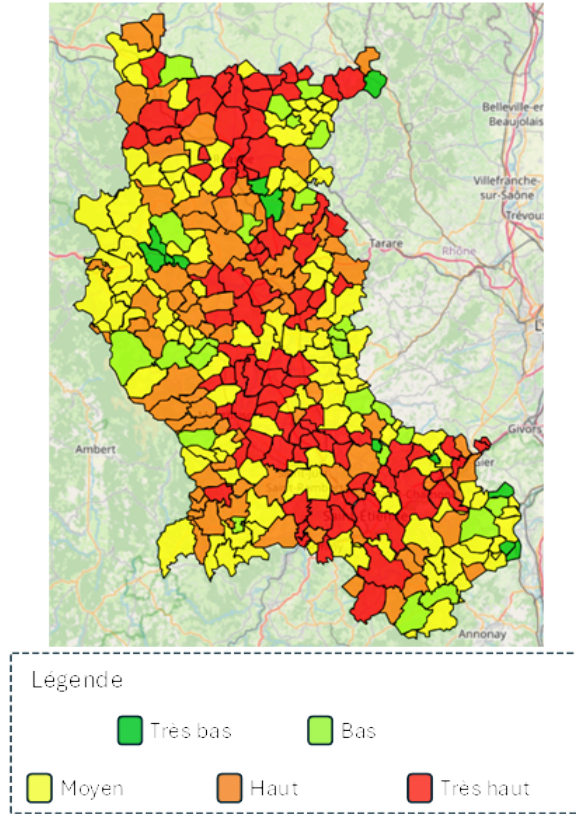


FIGURE 29 – Zonier finale du département 42 avec la première méthode

Les résultats de cette méthode mettent en évidence une certaine disparité entre les communes. De nombreuses communes voisines affichent des notes sensiblement différentes, ce qui est peu logique. Cependant, une zone centrale traversant le département du nord au sud semble regrouper de nombreuses communes classées en rouge, un phénomène également identifié par la MRN.

Cette approche, bien qu'informatrice, présente un nombre trop important de limites et d'hypothèses. Elle n'est pas retenue pour la suite des travaux, qui porteront sur l'ensemble de la région.

2.3 Deuxième approche : construction d'un zonier par GLM

2.3.1 Introduction au GLM

À la suite des résultats obtenus sur le zonier précédent, il est nécessaire d'adopter une méthodologie plus robuste sur le plan mathématique. C'est dans cette optique qu'une construction d'un zonier à partir de **Modèles Linéaires Généralisés (GLM)** est retenue.

Cette section se base sur le cours d'actuariat de l'assurance non-vie dispensé par Nicolas Baradel à l'ENSAE [12], ainsi que le cours de Juliette Chevallier sur les GLM [13].

Les GLM sont une généralisation des modèles de régression linéaire classiques. Ils permettent de **modéliser des relations** entre une **variable dépendante** et une ou plusieurs **variables explicatives**, lorsque les hypothèses des modèles linéaires ne sont pas respectées. En particulier, lorsque la variable dépendante suit une distribution différente de la normale. En actuariat, cet outil de modélisation est souvent utilisé pour expliquer la fréquence de survenance de sinistres ou encore le montant des coûts moyens.

Définitions et Principes de Base

Un GLM se décompose en trois éléments principaux :

- **La variable réponse, composante aléatoire** : La variable réponse Y est une variable aléatoire à modéliser, généralement pour ensuite estimer son espérance. Par exemple, en assurance, le coût des sinistres peut être considéré comme une variable aléatoire dont l'espérance doit être estimée afin de déterminer un tarif approprié pour l'ensemble du portefeuille. Les GLM reposent sur l'hypothèse que cette variable réponse suit une distribution appartenant à la **famille exponentielle**. Soit $(P_\theta)_{\theta \in \mathbb{R}}$ une famille de mesures de probabilité. Le modèle appartient à cette famille exponentielle si la densité peut s'écrire sous la forme :

$$f(y; \theta, \phi) = \exp \left(\frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi) \right)$$

Où a , b , et c sont des fonctions mesurables, a est strictement positive, et ϕ est un paramètre de dispersion. Les distributions couramment utilisées sont : la normale, la binomiale, la Poisson et la binomiale négative. Cette dernière est fréquemment employée dans les modèles de comptage avec sur-dispersion, comme pour la modélisation des sinistres.

- **Les variables explicatives, composante déterministe** : Les variables explicatives sont les variables permettant de décrire et de prédire la variable réponse. Le choix des variables explicatives est crucial pour la pertinence et l'utilisabilité du modèle. Il est essentiel de sélectionner un nombre limité de variables afin de garantir la simplicité d'utilisation tout en veillant à ce que celles-ci soient suffisamment explicatives pour capturer la variabilité de la variable réponse. Ces variables peuvent être quantitatives et/ou qualitatives et pourraient, par exemple, inclure des facteurs liés à l'exposition au risque, comme les caractéristiques des biens assurés ou des données géographiques.
- **La fonction de lien** : La fonction de lien joue un rôle clé dans le modèle, en établissant une relation entre l'espérance de la variable réponse et une combinaison linéaire des variables explicatives. Plutôt que de modéliser directement la moyenne, une fonction de lien monotone et dérivable, notée g , est utilisée pour relier ces deux composantes. La relation est exprimée comme suit :

$$g(\mathbb{E}[Y | X]) = g(\mu) = X\beta \iff \mathbb{E}[Y | X] = g^{-1}(X\beta)$$

où :

- X est la **matrice des variables explicatives**,
- β est le **vecteur des coefficients du modèle**.

Ces trois éléments, la variable réponse, les variables explicatives et la fonction de lien, confèrent aux GLM une grande flexibilité, permettant de s'adapter à diverses formes de distributions de la variable réponse (qu'elles soient binaires, de comptage ou continues) tout en capturant efficacement la structure des données.

Estimation des paramètres du modèle par maximum de vraisemblance

L'estimation des paramètres du modèle GLM (notés X ci-dessus) se fait généralement par la méthode du maximum de vraisemblance. Étant donné que la densité de la variable réponse Y , issue de la famille exponentielle (comme la loi de Poisson, Gamma ou Binomiale Négative), est connue, il est possible de maximiser la log-vraisemblance pour obtenir les estimations des paramètres du modèle.

Soit Y une variable aléatoire conditionnée sur X , suivant la loi $P_{\theta(X)}$. En particulier, en observant n réalisations indépendantes $(y_i, x_i)_{1 \leq i \leq n}$, la log-vraisemblance dans le cas de la fonction de lien canonique s'écrit :

$$\begin{aligned} l(\beta_0, \beta; (y_i | x_i)_{1 \leq i \leq n}) &= \sum_{i=1}^n \log(f(y_i; \beta_0 + x_i' \beta, \phi)) \\ &= \sum_{i=1}^n \frac{y_i(\beta_0 + x_i' \beta) - b(\beta_0 + x_i' \beta)}{a(\phi)} + c(y_i, \phi) \end{aligned}$$

L'estimation par maximum de vraisemblance de (β_0, β) est indépendante du paramètre de dispersion ϕ , qui peut être estimé ultérieurement.

La résolution des équations du premier ordre résultant de la maximisation de la log-vraisemblance se présente comme suit :

$$(\hat{\beta}_0, \hat{\beta}) \in \arg \max_{(\beta_0, \beta) \in \mathbb{R}^{d+1}} l(\beta_0, \beta; (y_i | x_i)_{1 \leq i \leq n})$$

Ces équations ne fournissent généralement pas de solution analytique. Ainsi, des **méthodes numériques** sont nécessaires pour estimer β_0 et β . Des outils comme **Python** permettent d'appliquer des algorithmes itératifs, notamment la méthode de Newton-Raphson ou les moindres carrés pondérés itératifs (IWLS), pour obtenir les paramètres du modèle.

Application dans le cadre d'un zonier grêle

Dans le cadre de l'assurance, et plus spécifiquement pour la modélisation du risque grêle via un zonier, les GLM permettent de comprendre et de modéliser les relations complexes entre divers facteurs explicatifs, tels que la fréquence et la sévérité des événements de grêle, les caractéristiques des habitations, ou encore la localisation géographique.

Dans le cadre de ce mémoire, l'utilisation d'un GLM permet de modéliser la fréquence des sinistres, et leur coût moyen en fonction de variables comme la taille des grêlons ou l'exposition géographique, tout en prenant en compte la surdispersion des données, un phénomène fréquent dans le contexte d'événements rares mais aux conséquences financières significatives.

En combinant les résultats de ces deux modèles il devient possible de calculer une prime pure pour chaque commune. Cet indicateur comparatif permet de construire un zonier où les communes les moins exposées sont celles ayant les plus faibles primes pures et inversement.

Prime pure

La prime pure représente le montant que l'assureur doit collecter pour couvrir le coût des sinistres prévus, sans inclure les frais de gestion, les bénéfices ou d'autres charges. Mathématiquement, la prime pure, notée π dans la suite, est équivalente à l'espérance des sinistres et peut être calculée en multipliant la fréquence des sinistres par le coût moyen d'un sinistre :

$$\pi = \text{Fréquence} \times \text{Coût moyen}$$

2.3.2 Structure des modèles et choix des variables explicatives

Pour obtenir une prime pure par commune à partir des données disponibles, il est nécessaire d'appliquer **deux modèles GLM** distincts : l'un pour modéliser le **coût moyen** et l'autre pour la **fréquence des sinistres**.

Variables explicatives

Le choix des variables explicatives dans un GLM est déterminant pour la qualité du modèle. Dans le contexte de la modélisation du risque grêle en assurance pour un produit MRH, des variables telles que **la fréquence et l'intensité des événements de grêle**, ainsi que les **caractéristiques des habitations**, apparaissent comme des choix naturels.

Les données de **Keraunos** fournissent des informations détaillées sur la **fréquence des épisodes de grêle et la taille des grêlons**, permettant ainsi d'incorporer ces éléments dans le modèle. Concernant les caractéristiques des maisons, seul le **nombre de pièces** est disponible comme variable, qui sera également intégré dans la modélisation. Ce choix est pertinent, car le nombre de pièces peut servir d'indicateur de la superficie couverte, offrant ainsi une meilleure estimation du risque d'endommagement potentiel lié aux événements de grêle.

Fonction de lien logarithmique

Dans la continuité du choix des variables explicatives, le type de fonction de lien choisi est essentiel pour modéliser correctement la relation entre ces variables et la variable réponse. Dans les deux modèles, la **fonction de lien logarithmique** est couramment employée. Elle permet de représenter les **relations** entre les **variables explicatives** et la **variable réponse** sous forme **multiplicative**, grâce aux **propriétés** de la **fonction exponentielle**. L'espérance conditionnelle des sinistres se modélise alors ainsi :

$$\begin{aligned} \log(\mathbb{E}[Y \mid X]) &= \beta_0 + \sum_{j=1}^p \beta_j X_j \\ \Leftrightarrow \mathbb{E}[Y \mid X] &= \exp(\beta_0) \cdot \exp(\beta_1 X_1) \cdots \exp(\beta_p X_p) \end{aligned}$$

Cette formulation permet de décomposer l'effet des différentes variables explicatives X_i , telles que la localisation géographique, les caractéristiques des habitations ou encore la taille des grêlons, en facteurs multiplicatifs. Chaque coefficient β traduit alors un effet multiplicatif sur la variable réponse, facilitant ainsi l'interprétation des résultats et l'ajustement des primes selon les caractéristiques spécifiques de chaque commune.

Différentes bases de données et leur structure

Une fois les variables explicatives et la fonction de lien établies, il est nécessaire de construire la base de données la plus adaptée pour la modélisation. Celle-ci renseigne **le nombre d'assurés**, **le nombre de sinistres** et **le nombre d'événements classés par taille de grêlons**, associés à des caractéristiques comme le **code INSEE**, l'**année**, ou le **nombre de pièces assurées**.

Pour fournir ces informations, trois bases distinctes sont construites et étudiées :

- **Base 1** : Une ligne par commune (environ 4 500 lignes) ;
- **Base 2** : Une ligne par commune et par année, avec un maximum de cinq lignes par commune correspondant aux années 2019 à 2023 (environ 22 000 lignes) ;
- **Base 3** : Une ligne par commune, année et nombre de pièces assurées, introduisant un niveau de détail supplémentaire (environ 170 000 lignes).

Chaque base de données présente **un niveau d'informations** distinct, ce qui influence la richesse des données disponibles par ligne. Par exemple, bien que la **première base** soit **plus condensée**, elle fournit une **quantité d'informations plus riche par ligne**. En revanche, la troisième base, qui contient **un nombre de lignes plus important**, offre des **données moins pertinentes**, car chaque ligne présente un nombre réduit d'assurés et de sinistres. Il est également important de souligner que toutes les lignes ne renseignent pas forcément de sinistres. Cette différence de granularité peut affecter la modélisation et la qualité des résultats. Le choix de la base de données sera effectué en fonction de plusieurs critères, qui seront détaillés dans la prochaine partie.

2.3.3 Modélisation de la fréquence des sinistres par assuré et par commune

La modélisation de la fréquence des sinistres s'appuie sur des distributions de comptage, telles que la loi de Poisson classiquement, ou la binomiale négative si une sur-dispersion est observée dans les données. Le nombre de sinistres prédit est ensuite divisé par le nombre d'assurés dans chaque commune pour obtenir une fréquence par assuré.

Moyennes et variances des sinistres par base

Avant d'aborder la modélisation, les moyennes et variances du nombre de sinistres ont été calculées pour chaque base :

Base	Moyenne	Variance
Base 1	1.948	39.40
Base 2	0.39	7.43
Base 3	0.05	0.24

Comme attendu, ces résultats mettent en évidence une **forte sur-dispersion** dans les données, rendant la **loi de Poisson inadaptée**. Un modèle de **binomiale négative** ou **quasi-Poisson** est donc plus approprié pour mieux appréhender cette variabilité.

Mesures d'évaluation des modèles : Log-vraisemblance, AIC et F-score

- **Log-vraisemblance** : Elle évalue l'ajustement des modèles. Une **log-vraisemblance plus élevée** (moins négative) indique que le **modèle explique mieux les données**, bien qu'elle ne prenne pas en compte la complexité du modèle. Pour cette raison, un critère comme l'AIC est utilisé en complément.
- **Akaike Information Criterion (AIC)** : L'AIC combine la log-vraisemblance avec une pénalisation liée au nombre de paramètres estimés, afin de minimiser la perte d'information tout en évitant les modèles trop complexes. La formule est :

$$AIC = 2k - 2 \log(L)$$

où k est le nombre de paramètres dans le modèle et L est le maximum de la fonction de vraisemblance du modèle. Le **modèle ayant l'AIC le plus faible est préféré**.

- **F-score** : Principalement utilisé pour évaluer les performances des modèles de classification, le F-score est une mesure qui combine la précision ($\text{précision} = \frac{\text{vrai positifs}}{\text{vrai positifs} + \text{faux positifs}}$) et le rappel ($\text{rappel} = \frac{\text{vrai positifs}}{\text{vrai positifs} + \text{faux négatifs}}$) en une seule métrique. Il est défini comme la moyenne harmonique de ces deux mesures :

$$F_1 = 2 \cdot \frac{\text{précision} \cdot \text{rappel}}{\text{précision} + \text{rappel}}$$

Un F-score élevé indique un bon équilibre entre précision et rappel.

Ces critères sont utilisés pour **comparer** plusieurs modèles et choisir celui qui offre le **meilleur compromis** entre **ajustement** et **simplicité**.

Résultats des modèles

Les modèles de GLM ont été testés sur les différentes bases, avec les résultats suivants :

Base de données	Variables explicatives	Distribution	P-valeur des coefficients	R ²	Log-vraisemblance	AIC
Base 1	- Nombre d'assurés - Nombre d'évènements par intensité ⁶	Négative binomiale	Toutes < 0.1%	62%	-11328.60	22671
Base 2	- Nombre d'assurés - Nombre d'évènements par intensité et par année	Négative binomiale	Toutes < 0.1%	47%	-11328.60	22671
Base 3	- Nombre d'assurés - Nombre de pièces - Nombre d'évènements par intensité et par année	Négative binomiale	< 0.1%, sauf pour le nombre de pièces (1.9%)	13%	-23626.23	47268
Base 3	- Nombre d'assurés - Nombre de pièces - Nombre d'évènements par intensité et par année	Quasi-Poisson	< 0.1%, sauf pour le nombre de pièces (1.7%)	9%	-29207.78	58431
Base 3	- Nombre d'assurés - Nombre de pièces - Nombre d'évènements par intensité et par année	Poisson	< 0.1%, sauf pour le nombre de pièces (0.6%)	12%	-21907.69	43831

³ L'intensité étant déterminée par des classes qui correspondent à celles proposées par l'échelle ANELFA, à savoir : A1, A2, A3, A4 et A5

TABLE 5 – Comparaison des résultats des modèles en fonction des bases de données

Les cases colorées en vert indiquent des résultats pertinents, tandis que ceux en rouge représentent les valeurs les moins satisfaisantes pour les colonnes R^2 , $\log\text{-vraisemblance}$ et AIC .

F-score

Le calcul du F-score est ajouté à ces résultats. Trois classes ont été définies pour les bases de données initiales : "Pas de sinistre", "Un sinistre", et "Plus d'un sinistre". Les mêmes classes sont appliquées aux prédictions du nombre de sinistres en utilisant les règles suivantes :

- "Pas de sinistre" pour des prédictions inférieures à 0,5,
- "Un sinistre" pour des prédictions comprises entre 0,5 et 1,5,
- "Plus d'un sinistre" pour des prédictions supérieures à 1,5.

Cette méthode permet de calculer le F-score global du modèle GLM, mais aussi de l'évaluer par classe. Ainsi, il devient possible d'identifier si le GLM est performant pour prédire certaines classes (par exemple, lorsqu'il n'y a pas de sinistre) tout en étant moins précis pour d'autres (lorsqu'il y en a plusieurs), ou inversement. Les résultats obtenus sont synthétisés dans le tableau suivant :

Base de données	Distribution	F-score global	F-score pas de sinistre	F-score un sinistre	F-score plus d'un sinistre
Base 1	Négative binomiale	0.517	0.59	0.622	0.277
Base 2	Négative binomiale	0.47	0.509	0.546	0.163
Base 3	Négative binomiale	0.446	0.586	0.502	0.285
Base 3	Quasi-Poisson	0.303	0.569	0.176	0.237
Base 3	Poisson	0.303	0.569	0.176	0.237

TABLE 6 – F-scores des modèles selon les bases de données et les distributions

Analyse des résultats

Les modèles avec une distribution binomiale négative ont été testés sur l'ensemble des bases de données, cette distribution étant particulièrement adaptée aux données sur-dispersées. Pour la base 3, les distributions quasi-Poisson et Poisson ont également été explorées, étant donné la moindre sur-dispersion observée dans cette base.

Les tests réalisés sur ces différents modèles fournissent plusieurs **indicateurs de performance** qui permettent de comparer leur pertinence en fonction des bases de données. L'objectif est de déterminer le modèle offrant le meilleur **compromis** entre la **qualité de l'ajustement**, la **simplicité** et la **capacité explicative**. Pour ce faire, plusieurs critères ont été analysés : la significativité des variables, le coefficient R^2 , la log-vraisemblance, l'AIC et le F-score.

Voici un récapitulatif de ces critères ainsi que leur rôle dans la sélection du modèle le plus approprié :

- **P-valeurs** : Les P-valeurs montrent la **significativité des variables explicatives**. Elles sont toutes inférieures à 0.1%, sauf pour la variable "nombre de pièces" dans la base 3, où elles varient entre 1.9%, 1.7% et 0.6%, ce qui reste en-dessous du seuil de 5%. Cependant, la p-valeur possède certaines limites et doit donc être complétée par d'autres indicateurs.
- **R^2** : Ce coefficient indique la **proportion de la variance expliquée par le modèle**. La base 1 présente un R^2 de 62%, suggérant un modèle solide. La base 2 est à 47%, tandis que la base 3 présente des R^2 entre 9% et 13%, signifiant une plus faible capacité explicative. Cependant, comme pour la p-valeur, il doit être complété par d'autres indicateurs.
- **Log-vraisemblance et AIC** : La log-vraisemblance est plus élevée pour les bases 1 et 2, tandis que la base 3 affiche des résultats inférieurs, notamment avec le modèle quasi-Poisson. De même, l'AIC est plus faible pour les modèles utilisant les bases 1 et 2, confirmant une meilleure qualité d'ajustement pour ces bases.
- **F-score** : Comme prévu par le fait que la base 1 est la plus agrégée, elle affiche le meilleur F-score global. De plus, les résultats par classe pour cette base sont également les plus performants.

Choix du modèle

Le modèle optimal dépend de plusieurs facteurs :

- **Significativité des variables** : Les variables explicatives doivent être significatives. Les **P-valeurs très faibles** trouvées confirment la **pertinence** des variables explicatives.
- **Capacité explicative** : Le modèle basé sur la première base se distingue par un R^2 de 62 %, ce qui en fait le modèle avec le R^2 le plus satisfaisant parmi les différents modèles évalués.
- **Qualité et complexité** : L'équilibre entre la log-vraisemblance et l'AIC met en avant le modèle basé sur la **base 1** comme le **meilleur compromis**, car c'est lui qui est la **log-vraisemblance la plus élevée** et l'AIC la **plus faible**. De plus, elle présente également le **F-score le plus élevé**.

Ainsi, le modèle utilisant la base de données avec une ligne par commune, une distribution négative binomiale, le nombre d'assurés et le nombre d'événements (A1 à A5) comme variables explicatives, sera utilisé pour modéliser le nombre de sinistres par commune. Ce choix résulte de l'élimination des autres modèles jugés non satisfaisants.

Il est naturel que ce modèle, basé sur la base la plus agrégée, présente les meilleurs résultats. Cependant, ce travail avait également pour objectif d'explorer différentes bases de données pour évaluer si une autre granularité aurait pu produire de meilleures performances. Cette approche a permis de tester plusieurs échelles d'analyse afin d'identifier celle qui était la plus pertinente pour la problématique étudiée.

2.3.4 Modélisation du coût moyen d'un sinistre par commune

Le **coût moyen des sinistres** peut être modélisé à l'aide d'un **GLM** en utilisant une **distribution Gamma ou Normale**, en fonction de la distribution observée des sinistres. En cas d'**asymétrie importante**, caractérisée par une queue longue, la **loi Gamma** est généralement privilégiée. Dans le cadre de cette étude, la queue longue est associée aux communes ayant enregistré des coûts moyens élevés.

Tout d'abord, les trois bases de données utilisées précédemment sont reprises, en ne conservant que les lignes contenant des sinistres. Une nouvelle colonne sera créée pour représenter le coût moyen d'un sinistre, en divisant la somme des charges des sinistres par le nombre de sinistres correspondant.

Distribution des coûts moyens

Une fois ces trois bases de données obtenues, la distribution des coûts moyens est tracée afin de déterminer si elle se rapproche davantage d'une distribution Gamma ou Normale.

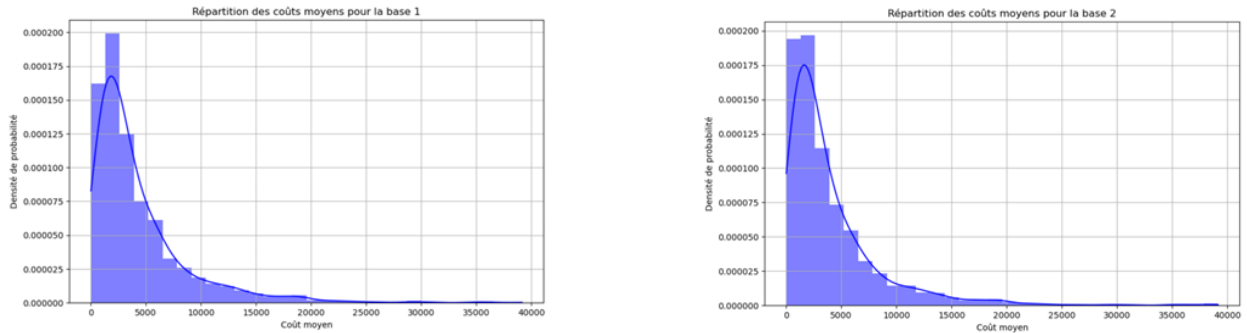


FIGURE 30 – Répartition des coûts moyens pour la base 1 à gauche et base 2 à droite



FIGURE 31 – Répartition des coûts moyens pour la base 3

Ces trois graphiques montrent clairement que les **distributions suivent une loi Gamma** plutôt qu'une loi Normale, pour l'ensemble des bases de données.

Résultats des modèles

La modélisation du coût moyen des sinistres présente des défis en raison des importantes variations possibles des coûts individuels. De plus, étant donné que la base de données contient un nombre limité de sinistres, l'analyse s'appuie principalement sur les p-valeurs pour évaluer la significativité des variables explicatives, sur le R^2 pour mesurer la proportion de variance expliquée, et sur le F-score pour évaluer la performance globale et l'équilibre entre précision et rappel dans la prédiction des coûts moyens.

Base de données	Variables explicatives	Distribution	P-valeur par coefficient	R ²	F-score global
Base 1	- Nombre d'évènements et intensité	Gamma	A1 : 0.723 A2 : 0.129 A3 : 0.232 A4 : < 0.1% A5 : < 0.1%	4.4%	0.466
Base 2	- Nombre d'évènements et intensité	Gamma	A1 : 0.244 A2 : 0.279 A3 : 0.929 A4 : < 0.1% A5 : < 0.1%	3.9%	0.466
Base 3	- Nombre d'évènements et intensité - Nombre de pièces	Gamma	A1 : 0.019 A2 : 0.079 A3 : 0.167 A4 : < 0.1% A5 : < 0.1% Pièces : < 0.1%	6.2%	0.466
Base 3	- Nombre d'évènements et intensité - Nombre de pièces	Gamma	A1 : 0.021 A2 : 0.0084 A4 : < 0.1% A5 : < 0.1% Pièces : < 0.1%	6.2%	0.466

TABLE 7 – Résumé des résultats des modèles Gamma en fonction des bases de données.

Pour calculer les F-scores dans ce contexte, trois classes sont définies : "Coût moyen nul" pour les lignes sans sinistre, "Coût moyen faible" et "Coût moyen élevé". Les deux dernières classes sont créées en répartissant les lignes restantes (celles avec un coût moyen non nul) en deux groupes égaux : les 50% des coûts moyens les plus faibles ont été attribués à la classe "Coût moyen faible", tandis que les 50% restants ont été affectés à la classe « Coût moyen élevé ».

Pour les prédictions, les observations sont classées par ordre croissant de coût moyen, puis réparties dans les trois classes définies précédemment, telles que chaque classe contienne un nombre de lignes identique à celui de la classification initiale.

Cependant, l'analyse montre que le F-score reste identique quel que soit le GLM utilisé. Par conséquent, cet indicateur, bien qu'intéressant, n'est pas retenu pour l'analyse des résultats qui va suivre.

Analyse des résultats et choix du modèle

Les modèles avec une distribution **Gamma** sont appliqués à l'ensemble des bases de données. Cette distribution est adaptée en raison de la nature des coûts moyens observés, qui présentent souvent une **asymétrie avec une queue longue**, comme ce qui peut être observé sur les graphiques précédemment montrés.

En examinant les **R²** obtenus, il apparaît qu'ils sont **faibles** pour tous les modèles, avec des valeurs comprises entre **3.9% et 6.2%**. Cela est relativement prévisible étant donné la **petite taille de l'échantillon de sinistres** (<10 000 sinistres), ainsi que les importantes **variations de coût** d'un sinistre à l'autre. Le **R²** n'est pas toujours un indicateur fiable pour les modèles Gamma, car il mesure la proportion de variance expliquée par les variables explicatives, mais ne capture pas toujours bien les phénomènes avec des distributions asymétriques.

Dans ce contexte, l'accent est mis sur les **p-valeurs**, qui permettent de déterminer si une variable explicative a un effet **statistiquement significatif** sur le coût moyen d'un sinistre.

Pour les deux premiers modèles (bases 1 et 2), les **p-valeurs des variables explicatives A1, A2 et A3 sont trop élevées** (> 0.1), indiquant qu'elles ne sont pas significatives. Ces modèles sont donc **écartés**.

Le **modèle 3 (sur la base 3)** présente des résultats plus intéressants, avec des **p-valeurs très faibles pour les variables A4, A5, et le nombre de pièces**, confirmant leur **forte significativité**. La variable **A1 est significative**, tandis qu'**A2 est moyennement significative**, et **A3 n'est pas significative** avec une p-valeur de **0.167**. Il faut noter cependant que ces résultats sont attendus du fait de la modélisation d'un phénomène compliqué, qu'est le coût moyen, par un modèle assez simple.

Un ajustement du modèle est tenté en retirant la variable **A3**. Les résultats sont restés assez similaires, tant en termes de **p-valeurs** que de **R²**. Cependant, l'exclusion de la variable **A3**, qui correspond aux grêlons de taille intermédiaire, semble problématique. En effet, la classe **A3 joue un rôle important dans la compréhension des variations de coût**. De plus, cela correspondrait à **supprimer 10% des sinistres**.

Ainsi, le modèle retenu pour la modélisation du coût moyen d'un sinistre est celui utilisant la **base de données 3**, avec une distribution **Gamma**. Les variables explicatives comprennent le **nombre de pièces** ainsi que les **événements de grêle (A1 à A5, A3 inclus)**. Bien que ce modèle soit la **meilleure option** parmi celles évaluées, il n'est pas totalement satisfaisant.

Finalement, pour déterminer le **coût moyen d'un sinistre par commune**, tout en tenant compte du **nombre de pièces** dans la modélisation, le coût moyen d'un sinistre pour un nombre de pièces dans chaque commune est d'abord calculé. Ensuite, une **moyenne pondérée** de ce coût moyen par le **nombre d'assurés** est effectuée, afin d'obtenir le coût moyen d'un sinistre pour chaque commune.

2.3.5 Résultats et interprétations

La réalisation des deux modèles GLM permet d'obtenir, pour chaque commune, une estimation de la fréquence des sinistres ainsi que du coût moyen par sinistre. La prime pure par commune est ensuite calculée en multipliant ces deux valeurs. Les communes sont ensuite classées par ordre croissant de prime pure. Les 50% des communes présentant la prime pure la plus faible sont attribuées à la catégorie de risque faible, celles situées entre 50% et 85% à la catégorie de risque modéré, et les 15% des communes ayant la prime pure la plus élevée sont classées dans la catégorie de risque élevé. Cette classification répond à un critère établi dans la politique de souscription du client. Cette hypothèse est retenue pour la suite du mémoire.

Les classifications possibles

D'autres types de classifications peuvent être envisagés selon les besoins :

- **Découpage par quantiles homogènes** : L'ensemble des primes est divisé en groupes contenant un nombre équivalent de valeurs. Le nombre de groupes de risque est déterminé à l'avance (par exemple, 5 groupes), puis les primes sont découpées en quantiles homogènes (20 % des plus faibles dans le premier groupe, 20 % suivants dans le deuxième groupe, etc.). Cette méthode permet d'avoir des groupes contenant un même nombre de communes.
- **Découpage par intervalles fixes** : Les primes sont regroupées en fonction de plages de valeurs définies, par exemple 20 €. Le premier groupe inclut les primes de 0 à 20 €, le deuxième celles de 21 à 40 €, etc. Cette méthode garantit une échelle uniforme pour la classification.
- **K-means clustering** : Les primes sont regroupées en groupes homogènes en minimisant la variance intra-groupe, grâce à l'algorithme de clustering K-means. Cette approche offre l'avantage de créer des groupes adaptés à la distribution des primes.

Cette liste est non exhaustive et peut être enrichie en fonction des besoins et des spécificités du projet.

Zonier

Le zonier obtenu, basé sur la première classification, est présenté ci-dessous :

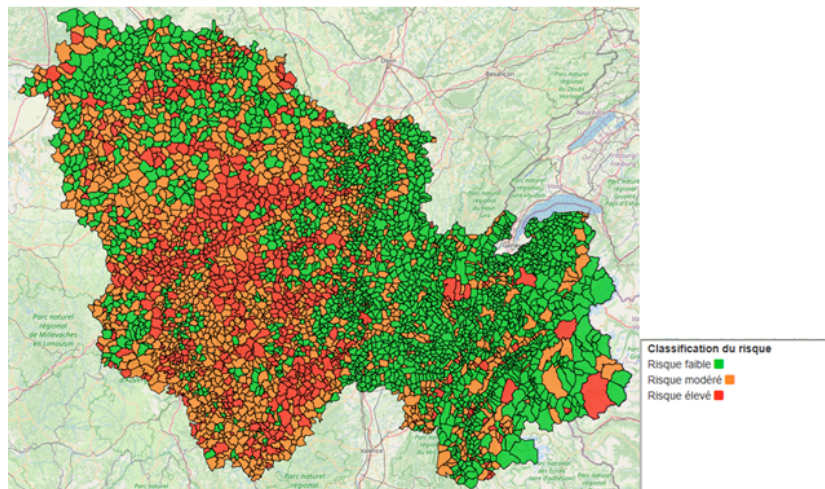


FIGURE 32 – Zonier par GLM de l'exposition de l'assureur MRH au risque grêle pour la région Auvergne-Rhône-Alpes

Ce zonier révèle des **bandes caractéristiques du risque grêle**. Toutefois, il présente une certaine dispersion : certaines communes classées en zone verte (risque faible) sont entourées de communes rouges (risque élevé), et inversement. Ce phénomène peut s'expliquer par un **manque ou un excès de données dans certaines communes**. Pour corriger cette dispersion, un **lissage géographique** a été appliqué.

Lissage géographique

Le lissage est basé sur le **calcul du barycentre** de chaque commune à partir de ses coordonnées géographiques. Pour lisser les résultats d'une commune, la **moyenne pondérée des primes pures des six communes les plus proches** est d'abord calculée. **L'inverse de la distance** sert de **coefficient de pondération**. Le choix des six communes est un choix arbitraire qui s'explique par le fait qu'en moyenne, une commune en Auvergne-Rhône-Alpes possède six communes limitrophes. Ce choix offre un compromis raisonnable entre précision et simplicité dans le lissage.

Ensuite, une seconde **moyenne pondérée** est calculée **entre la prime pure initiale de la commune et la moyenne pondérée des six communes voisines**. Plusieurs tests ont été effectués pour déterminer le poids optimal à attribuer à ces deux composantes. Ces tests ont pris en compte l'impact visuel des résultats (minimisant le nombre de communes entourées par des communes de couleurs différentes), des considérations de prudence (en privilégiant les résultats offrant les primes pures collectées les plus élevées), ainsi que la nécessité de conserver suffisamment d'informations sur la commune elle-même. Le choix d'un poids égal de 0.5 pour chaque composante s'est révélé être un bon compromis, permettant de compenser le manque de données, de maintenir une précision satisfaisante des résultats et de garantir que la prime pure finale reflète une estimation prudente.

Mathématiquement, la moyenne pondérée des 6 communes les plus proches se calcule comme suit :

$$\bar{\pi} = \frac{\sum_{i=1}^6 \rho_i \pi_i}{\sum_{i=1}^6 \rho_i}$$

où $\rho_i = \frac{1}{\text{dist}(\text{commune lissée}, \text{commune } i)}$

Ainsi, la prime pure lissée s'obtient :

$$\pi_{\text{lissée}} = 0.5 \pi_{\text{initiale}} + 0.5 \bar{\pi}$$

Après l'application du lissage, la même classification des risques que précédemment est utilisée pour obtenir le zonier suivant :

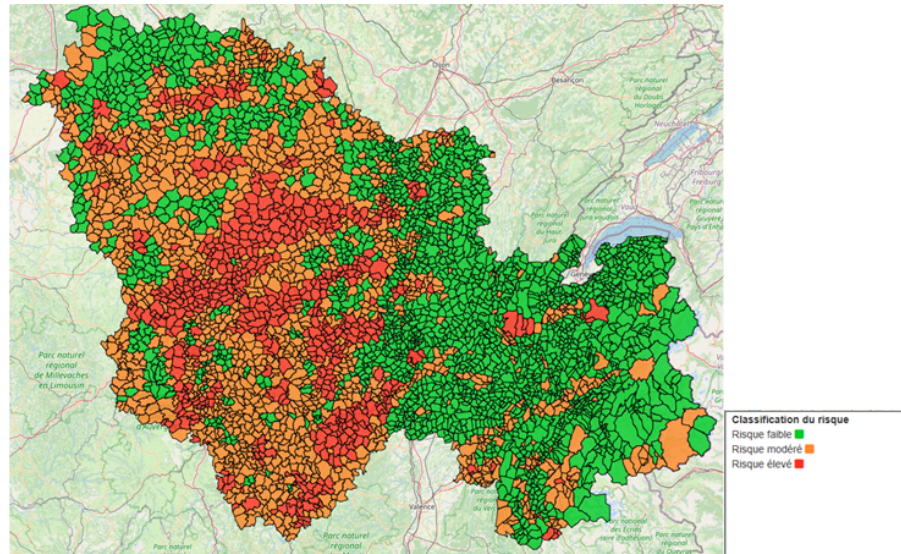


FIGURE 33 – Zonier par GLM de l'exposition de l'assureur MRH au risque grêle pour la région Auvergne-Rhône-Alpes après lissage

Grâce au lissage, les communes "encerclées" par des communes avec des classifications différentes se font beaucoup plus rares, ce qui améliore la cohérence générale du zonier. Les zones de risque élevé restent bien identifiées, renforçant ainsi la pertinence du modèle.

Limites du modèle

Le modèle GLM, bien que couramment utilisé en actuariat, présente plusieurs limites dans le cadre de cette étude. Tout d'abord, le nombre restreint de variables explicatives réduit la capacité du modèle à capturer la complexité du phénomène étudié. Par exemple, des caractéristiques importantes telles que la superficie des communes, qui pourraient influencer la dispersion des sinistres ou l'exposition aux événements climatiques, n'ont pas été prises en compte. De plus, les performances du modèle restent insatisfaisantes, comme en témoigne le coefficient de détermination, R^2 , très faible pour la modélisation du coût moyen des sinistres. Cela reflète une part importante de variabilité non expliquée, limitant ainsi la fiabilité des résultats obtenus. Ces insuffisances justifient la proposition d'une troisième méthode, qui cherche à surmonter ces limites en intégrant des approches plus robustes et adaptées aux spécificités des données disponibles.

2.4 Troisième approche : construction à partir de la théorie de la Crédibilité

Une troisième approche, basée sur la théorie de la crédibilité, est explorée. Cette méthode, particulièrement adaptée pour affiner les prévisions en tenant compte des données historiques spécifiques à chaque zone, est examinée en détail dans la section suivante. Celle-ci se focalise sur la théorie de la crédibilité, avec une attention particulière aux modèles de Bühlmann et Bühlmann-Straub, ainsi qu'à leur application pratique dans le cadre de cette étude.

2.4.1 Présentation de la théorie de la crédibilité

Les origines de la crédibilité

La théorie de la crédibilité est développée au début du 20^e siècle. Son origine remonte à une situation impliquant General Motors et Tucker (un petit constructeur indépendant), tous deux assurés par Allstate pour les accidents du travail. Allstate calculait un taux moyen sur l'ensemble de son portefeuille et appliquait le même tarif à tous ses assurés. Cependant, General Motors a analysé sa propre sinistralité et a démontré qu'elle est inférieure au taux imposé par Allstate. En conséquence, General Motors a demandé de payer un tarif plus bas et "juste" au vu de son historique, en arguant que le nombre élevé de ses employés rend cette demande justifiée. Une question se pose alors : cette approche était-elle applicable à Tucker, qui a un faible nombre d'employés assurés ?

C'est dans ce contexte que les travaux de Mowbray, publiés en 1914, ont émergé. Ceux-ci précisent les conditions nécessaires pour garantir une pleine crédibilité à l'expérience d'un assuré. Ces travaux ont ensuite été étendus par Whitney en 1918, qui introduit le concept de crédibilité partielle. Whitney propose une prime individuelle comme une pondération entre l'expérience collective et l'expérience individuelle :

$$\pi = \alpha X + (1 - \alpha)C$$

Où α est le facteur de crédibilité (présenté et détaillé ultérieurement), X représente l'expérience individuelle, basée sur l'historique des sinistres propres à l'assuré, et C correspond à l'expérience collective, reposant sur l'historique de sinistres de tous les assurés du portefeuille.

Définition de la crédibilité

La théorie de la crédibilité est une méthode statistique utilisée en actuariat pour estimer les primes d'assurance en se basant sur la sinistralité passée. Elle intègre à la fois les informations spécifiques à un groupe particulier et les données globales de l'ensemble des assurés.

Il existe deux types principaux de crédibilité :

- **Crédibilité de stabilité (ou crédibilité américaine)** : Ce type de crédibilité est utilisé lorsque l'assureur prend en compte l'expérience individuelle uniquement si elle est suffisamment significative.
- **Crédibilité de précision** : Ici, l'importance de l'expérience individuelle augmente avec l'hétérogénéité du portefeuille. Cette crédibilité est définie par un facteur de crédibilité, noté α , qui peut prendre les valeurs suivantes :
 - $\alpha = 1 \Rightarrow$ Crédibilité totale
 - $\alpha \in [0, 1[\Rightarrow$ Crédibilité partielle

Lorsque la crédibilité est totale ($\alpha = 1$), la prime pour la période $n+1$ dépend uniquement de la sinistralité propre du contrat observée sur les n dernières années. En revanche, pour une crédibilité partielle ($0 \leq \alpha < 1$), une combinaison entre les données de sinistralité individuelle et celles du portefeuille global doit être faite pour obtenir la prime. Pour déterminer le facteur de crédibilité α approprié, plusieurs modèles sont disponibles, notamment ceux de **Bühlmann** et de **Bühlmann-Straub**, qui seront décrits en détail dans les sections suivantes.

2.4.2 Modèle de Bühlmann

Le modèle de Bühlmann vise à obtenir le meilleur estimateur de la prime individuelle π , c'est-à-dire la prime la plus compétitive et la plus équitable possible. En assurance, la prime à payer par un assuré doit refléter le risque qu'il représente. Cependant, les données historiques concernant un assuré spécifique peuvent être limitées ou insuffisantes pour évaluer ce risque de manière précise. Le modèle de Bühlmann propose une solution en trouvant un équilibre entre les données propres à l'assuré et les données globales de l'ensemble du portefeuille d'assurés.

Cette sous-section, ainsi que la suivante, s'appuient sur plusieurs ressources fondamentales : le livre "*A course in Credibility Theory and its Application*" de Hans Bühlmann et Alois Gisler [28], le cours de Pierre Thérond sur la théorie de la crédibilité [14], les slides d'Olivier Lopez consacrées au même sujet [15], les enseignements d'Adrien Suru sur les principes de base de l'assurance dommage [16], ainsi que le mémoire d'actuariat de Candice Poujade intitulé "*Application de la théorie de la crédibilité dans le contexte de l'assurance des arrêts de travail*" [11].

Soit un portefeuille constitué de I risques similaires. Il est noté $\mathbf{X}_i = (X_{i1}, \dots, X_{in})'$ le vecteur des observations associé au risque i et Θ_i son profil de risque. L'objectif est de déterminer, pour chaque risque individuel i , un estimateur de la prime individuelle. Bühlmann a restreint son analyse aux estimateurs linéaires par rapport aux observations $(X_{ij})_{1 \leq i \leq I, 1 \leq j \leq n}$.

Hypothèses du modèle

Le modèle de Bühlmann repose sur les deux hypothèses suivantes :

- **(H1)** : Les variables X_{ij} , $j = 1, \dots, n$, sont, conditionnellement à $\Theta_i = \theta$, indépendantes et identiquement distribuées selon une loi F_θ avec les moments conditionnels suivants :

$$\pi(\Theta_i) = \mathbb{E}[X_{ij} \mid \Theta_i],$$

$$\sigma^2(\Theta_i) = \text{Var}(X_{ij} \mid \Theta_i).$$

- **(H2)** : Les couples $(\Theta_1, X_1), \dots, (\Theta_I, X_I)$ sont indépendants et identiquement distribués.

Formulation du modèle

Sous ces hypothèses, l'estimateur de crédibilité $\widehat{\pi(\Theta_i)}$ est de la forme :

$$\widehat{\pi(\Theta_i)} = \widehat{a_0^{(i)}} + \sum_{k=1}^I \sum_{j=1}^n \widehat{a_{kj}^{(i)}} X_{kj}.$$

La distribution des X_i étant invariante par permutation des $X_{ij}, j = 1, \dots, n$, alors :

$$\widehat{\pi(\Theta_i)} = \widehat{a_0^{(i)}} + \sum_{k=1}^I \widehat{a_k^{(i)}} \overline{X_k},$$

où $\overline{X_k} = \frac{1}{n} \sum_{j=1}^n X_{kj}$.

Ensuite, pour trouver cet estimateur de crédibilité, l'erreur quadratique moyenne est minimisée :

$$(\widehat{a_0^{(i)}}, \widehat{a_k^{(i)}}) = \arg \min \mathbb{E} \left[\left(\pi(\Theta_i) - a_0^{(i)} - \sum_{k=1}^I a_k^{(i)} \overline{X_k} \right)^2 \right].$$

En annulant les dérivées partielles de l'erreur quadratique moyenne par rapport à $a_0^{(i)}$ et $a_k^{(i)}$, les équations suivantes sont obtenues :

$$\begin{cases} \mathbb{E}[\pi(\Theta_i) - a_0^{(i)} - \sum_{k=1}^I a_k^{(i)} \overline{X_k}] = 0, \\ \mathbb{E}[\overline{X_j}(\pi(\Theta_i) - a_0^{(i)} - \sum_{k=1}^I a_k^{(i)} \overline{X_k})] = 0, \quad \forall 1 \leq j \leq I. \end{cases}$$

C'est un système de $I + 1$ équations à $I + 1$ inconnues.

En injectant le résultat de $a_0^{(i)}$ de la première équation dans les I suivantes, l'équation suivante est obtenue :

$$\begin{aligned} \mathbb{E}[\overline{X_j} \pi(\Theta_i)] - \mathbb{E}[\pi(\Theta_i)] \mathbb{E}[\overline{X_j}] &= \sum_{k=1}^I a_k^{(i)} (\mathbb{E}[\overline{X_j} \overline{X_k}] - \mathbb{E}[\overline{X_j}] \mathbb{E}[\overline{X_k}]), \quad \forall 1 \leq j \leq I \\ \Leftrightarrow \text{Cov}(\overline{X_j}, \pi(\Theta_i)) &= \sum_{k=1}^I a_k^{(i)} \text{Cov}(\overline{X_k}, \overline{X_j}), \quad \forall 1 \leq j \leq I \end{aligned}$$

Or la covariance peut être définie comme telle, avec δ_{jk} le symbole de Kronecker :

$$\text{Cov}(\overline{X_k}, \overline{X_j}) = \delta_{jk} \text{Var}(\overline{X_j})$$

Par conséquent :

$$\text{Cov}(\overline{X_j}, \pi(\Theta_i)) = a_j^{(i)} \text{Var}(\overline{X_j}), \quad \forall 1 \leq j \leq I$$

Aussi, d'après l'hypothèse (H2) du modèle :

$$\text{Cov}(\overline{X_j}, \pi(\Theta_i)) = 0 \quad \text{si } i \neq j$$

Par conséquent, le système suivant est obtenu :

$$\begin{cases} a_j^{(i)} = 0, & \text{si } i \neq j, \\ a_i^{(i)} = \frac{\text{Cov}(\overline{X_i}, \pi(\Theta_i))}{\text{Var}(\overline{X_i})}, \\ a_0^{(i)} = \mathbb{E}[\pi(\Theta_i)] - a_i^{(i)} \mathbb{E}[\overline{X_i}]. \end{cases}$$

Et

$$\text{Cov}(\overline{X_i}, \pi(\Theta_i)) = \mathbb{E}[\overline{X_i} \pi(\Theta_i)] - a_i^{(i)} \mathbb{E}[\overline{X_i}]$$

Par espérance conditionnelle, les égalités suivantes sont vérifiées :

$$\begin{cases} \mathbb{E}[\pi(\Theta_i)] = \mathbb{E}[\mathbb{E}[X_i | \Theta_i]] = \mathbb{E}[X_i] = \mathbb{E}[\overline{X_i}] \\ \mathbb{E}[\overline{X_i} \pi(\Theta_i)] = \mathbb{E}[\mathbb{E}[\overline{X_i} \pi(\Theta_i) | \Theta_i]] = \mathbb{E}[\pi(\Theta_i)^2] \end{cases}$$

D'où

$$\text{Cov}(\overline{X_i}, \pi(\Theta_i)) = \mathbb{E}[\pi(\Theta_i)^2] - \mathbb{E}[\pi(\Theta_i)]^2 = \text{Var}[\pi(\Theta_i)] := \tau^2$$

De plus, le théorème de la variance totale permet de définir la variance ainsi :

$$\begin{aligned} \text{Var}(\overline{X_i}) &= \mathbb{E}[\text{Var}(\overline{X_i} | \Theta_i)] + \text{Var}(\mathbb{E}[\overline{X_i} | \Theta_i]) \\ &= \frac{\mathbb{E}[\sigma^2(\Theta_i)]}{n} + \text{Var}(\pi(\Theta_i)) = \frac{\sigma^2}{n} + \tau^2 \end{aligned}$$

Et comme

$$\mathbb{E}[\pi(\Theta_i)] = \mathbb{E}[\overline{X_i}] = C$$

Les estimateurs suivants sont obtenus :

$$\begin{aligned} \widehat{a_i^{(i)}} &= \frac{n}{n + \frac{\sigma^2}{\tau^2}} \\ \widehat{a_0^{(i)}} &= \left(1 - \widehat{a_i^{(i)}}\right) C \end{aligned}$$

Enfin, en notant $\alpha = \widehat{a_i^{(i)}}$, l'expression de la prime de Bühlmann est retrouvée :

$$\widehat{\pi(\Theta_i)} = \alpha \overline{X_i} + (1 - \alpha)C$$

Cette expression peut être réécrite de manière plus intuitive. En effet, le rapport $\frac{\sigma^2}{\tau^2}$ peut être défini comme suit :

$$\frac{\sigma^2}{\tau^2} = \left(\frac{\sigma/C}{\tau/C}\right)^2$$

Ainsi :

- $\frac{\sigma}{C}$ s'interprète comme la **variabilité interne** du risque au sein d'un groupe homogène de risque. En d'autres termes, cela reflète la dispersion des sinistres dus à la grêle au sein d'un groupe homogène de risque.
- $\frac{\tau}{C}$ s'interprète comme l'**hétérogénéité du portefeuille**, c'est-à-dire la différence du risque de grêle d'un groupe de risque à l'autre.

Le facteur de crédibilité α augmente avec le nombre d'années d'observations n , avec l'hétérogénéité du portefeuille $\frac{\tau}{C}$, et diminue avec la variabilité interne du risque $\frac{\sigma}{C}$.

Le facteur de crédibilité peut donc se réécrire comme suit :

$$\alpha = \frac{n}{n + \frac{\sigma^2}{\tau^2}} = \frac{\tau^2}{\tau^2 + \frac{\sigma^2}{n}} = \frac{\sigma_h^2}{\sigma_h^2 + \sigma_{\overline{X_i}}^2}$$

Où $\sigma_{\overline{X_i}}^2 = \frac{\sigma^2}{n}$ et $\sigma_h^2 = \tau^2$. Ainsi, la prime de crédibilité devient :

$$\widehat{\pi(\Theta_i)} = \frac{\sigma_h^2}{\sigma_h^2 + \sigma_{\overline{X_i}}^2} \overline{X_i} + \frac{\sigma_{\overline{X_i}}^2}{\sigma_h^2 + \sigma_{\overline{X_i}}^2} C$$

2.4.3 Modèle de Bühlmann-Straub

Le modèle de Bühlmann-Straub est une extension du modèle de Bühlmann, qui est l'un des piliers de la théorie de la crédibilité. Alors que le modèle de Bühlmann repose sur l'hypothèse que chaque groupe de risques possède un même niveau d'exposition, le modèle de Bühlmann-Straub permet de généraliser cette approche en prenant en compte des expositions inégales entre les différents groupes. Cela le rend particulièrement adapté aux situations où l'on dispose d'un nombre variable de sinistres ou d'assurés.

Les notations restent identiques à celles du modèle de Bühlmann. Cependant, ω_{ij} le poids associé à l'observation X_{ij} , qui traduit l'importance du risque i pour l'année j , est introduit.

Hypothèses du modèle

Le modèle de Bühlmann-Straub repose sur les hypothèses suivantes :

- **(H1)** : Les variables X_{ij} , $j = 1, \dots, n$, sont, conditionnellement à Θ_i , indépendantes avec les moments conditionnels suivants :

$$\begin{aligned}\pi(\Theta_i) &= \mathbb{E}[X_{ij} \mid \Theta_i], \\ \sigma^2(\Theta_i) &= \text{Var}(X_{ij} \mid \Theta_i).\end{aligned}$$

- **(H2)** : Les couples $(\Theta_1, X_1), \dots, (\Theta_I, X_I)$ sont indépendants et les $\Theta_1, \dots, \Theta_I$ sont indépendantes et identiquement distribuées.

Formulation du modèle

L'objectif est d'estimer, pour chaque risque i , la prime individuelle $\pi(\Theta_i)$. Les risques étant indépendants par (H2), l'estimateur $\widehat{\pi(\Theta_i)}$ dépend uniquement des observations X_{ij} rattachées au risque i :

$$\widehat{\pi(\Theta_i)} = a_0^{(i)} + \sum_j a_j^{(i)} X_{ij}$$

De l'indépendance des risques, on a également :

$$\text{Cov}(\widehat{\pi(\Theta_i)}, X_{kl}) = \text{Cov}(\pi(\Theta_i), X_{kl}) = 0 \quad \forall l, k \neq i$$

Cela signifie que la prime de crédibilité pour un risque i ne dépend que de son propre historique de sinistres, et non de celui des autres risques, ce qui est intuitif. Si la prime collective C est connue, les autres risques n'apportent aucune information supplémentaire, car ils sont indépendants du risque considéré.

Estimateur de crédibilité

Selon (H1) et (H2), les observations X_{ij} sont individuellement non biaisées, et leur variance conditionnelle est inversement proportionnelle aux poids ω_{ij} . En notant $\omega_i = \sum_{j=1}^n \omega_{ij}$, le meilleur estimateur linéaire, individuellement sans biais, est :

$$X_i = \sum_j \frac{\omega_{ij}}{\omega_i} X_{ij}$$

Avec les moments suivants :

$$\begin{aligned}\mathbb{E}[X_i \mid \Theta_i] &= \sum_j \frac{\omega_{ij}}{\omega_i} \mathbb{E}[X_{ij} \mid \Theta_i] = \pi(\Theta_i) \\ \text{Var}[X_i \mid \Theta_i] &= \sum_j \left(\frac{\omega_{ij}}{\omega_i} \right)^2 \text{Var}[X_{ij} \mid \Theta_i] = \frac{\sigma^2(\Theta_i)}{\omega_i}\end{aligned}$$

Comme dans le modèle de Bühlmann, l'estimateur de la crédibilité a la forme :

$$\widehat{\pi(\Theta_i)} = \alpha_i X_i + (1 - \alpha_i)C$$

Et doit satisfaire :

$$\begin{aligned} \text{Cov}(\widehat{\pi(\Theta_i)}, X_i) &= \alpha_i \text{Cov}(X_i, X_i) = \text{Cov}(\pi(\Theta_i), X_i) \\ &\Leftrightarrow \alpha_i \text{Var}(X_i) = \text{Cov}(\pi(\Theta_i), X_i) \end{aligned}$$

Étant donné que :

$$\text{Var}(X_i) = \mathbb{E}[\text{Var}(X_i | \Theta_i)] + \text{Var}(\mathbb{E}[X_i | \Theta_i]) = \frac{\sigma^2}{\omega_i} + \tau^2$$

Et :

$$\begin{aligned} \text{Cov}(\pi(\Theta_i), X_i) &= \mathbb{E}[\text{Cov}(\pi(\Theta_i), X_i | \Theta_i)] + \text{Cov}(\pi(\Theta_i), \mathbb{E}[X_i | \Theta_i]) \\ &= 0 + \text{Var}[\pi(\Theta_i)] = \tau^2 \end{aligned}$$

Il en découle :

$$\alpha_i = \frac{\text{Cov}(\pi(\Theta_i), X_i)}{\text{Var}(X_i)} = \frac{\tau^2}{\frac{\sigma^2}{\omega_i} + \tau^2} = \frac{\omega_i}{\omega_i + \frac{\sigma^2}{\tau^2}}$$

Il reste à montrer que $\widehat{\pi(\Theta_i)}$ est également l'estimateur de crédibilité pour le risque i basé sur toutes les données, ce qui est attendu d'après les propriétés de X_i . Il a déjà été montré que l'estimateur de crédibilité du risque i dépend uniquement des données du risque i . Il reste donc à montrer que :

$$\text{Cov}(\widehat{\pi(\Theta_i)}, X_{ij}) = \text{Cov}(\pi(\Theta_i), X_{ij}) = \tau^2 \quad \forall j = 1, \dots, n$$

Il s'avère que :

$$\begin{aligned} \text{Cov}(\widehat{\pi(\Theta_i)}, X_{ij}) &= \alpha_i \left(\sum_k \frac{\omega_{ik}}{\omega_i} \text{Cov}(X_{ik}, X_{ij}) \right) \\ &= \alpha_i \left(\sum_k \frac{\omega_{ik}}{\omega_i} \left(\frac{\sigma^2}{\omega_{ik}} \delta_{kj} + \tau^2 \right) \right) \\ &= \alpha_i \left(\sum_k \frac{\omega_{ik}}{\omega_i} \tau^2 + \frac{\sigma^2}{\omega_i} \right) \\ &= \frac{\omega_i}{\omega_i + \frac{\sigma^2}{\tau^2}} \left(\sum_k \frac{\omega_{ik}}{\omega_i} \tau^2 + \frac{\sigma^2}{\omega_i} \right) \\ &= \tau^2 \end{aligned}$$

Sachant que $\delta_{kj} = \begin{cases} 1 & \text{si } k = j \\ 0 & \text{sinon} \end{cases}$.

Ainsi, la relation est vérifiée, et il a été prouvé que l'estimateur inhomogène du modèle de Bühlmann-Straub est le suivant :

$$\widehat{\pi(\Theta_i)} = \alpha_i X_i + (1 - \alpha_i)C$$

Où :

$$X_i = \sum_j \frac{\omega_{ij}}{\omega_i} X_{ij}, \quad \omega_i = \sum_j \omega_{ij}, \quad \alpha_i = \frac{\omega_i}{\omega_i + \frac{\sigma^2}{\tau^2}}$$

Estimation des paramètres σ^2 et τ^2

Pour implémenter le modèle, il est nécessaire d'estimer les coefficients de crédibilité σ^2 et τ^2 . Les estimateurs sans biais et convergents de ces paramètres sont :

$$\widehat{\sigma^2} = \frac{1}{I(n-1)} \sum_{i=1}^I \sum_{j=1}^n \omega_{ij} (X_{ij} - X_i)^2$$

$$\widehat{\tau^2} = \frac{\omega}{\omega^2 - \sum_{i=1}^I \omega_i^2} \left\{ \sum_{i=1}^I \omega_i (X_i - \bar{X})^2 - (I-1) \widehat{\sigma^2} \right\}$$

Avec :

$$\omega = \sum_{i,j} \omega_{ij},$$

$$\bar{X} = \frac{\sum_{i,j} \omega_{ij} X_{ij}}{\omega}$$

Cependant, comme $\widehat{\tau^2}$ peut être négatif, l'estimateur corrigé : $\widehat{\tau^2} = \max\{\widehat{\tau^2}, 0\}$ sera utilisé en pratique.

2.4.4 Application de la théorie de la crédibilité pour la construction d'un zonier

Après sa présentation théorique, le modèle de Bühlmann-Straub est mis en application dans cette partie pour proposer une nouvelle construction de zonier. Le modèle de Bühlmann-Straub est particulièrement adapté à ce type de problématique, car il combine les données spécifiques à un groupe de risques avec des données globales afin d'obtenir une estimation plus fiable. Cela permet de pallier un manque de données pour certains groupes en équilibrant la spécificité des données individuelles avec la stabilité des données globales.

Pour la construction du zonier de niveau de risque, le modèle de Bühlmann-Straub est appliqué une première fois pour modéliser la fréquence des sinistres par commune, puis une seconde fois pour modéliser le coût moyen des sinistres par commune.

Modélisation de la fréquence des sinistres

La première étape consiste à créer des groupes de risques homogènes en fréquence, tout en assurant une différenciation significative entre les groupes. Les événements Keraunos et le nombre de pièces sont utilisés, car ces deux variables influencent la fréquence des sinistres.

Concernant les événements Keraunos, le nombre moyen de sinistres par événement est observé par classe afin de calculer le nombre théorique de sinistres par commune.

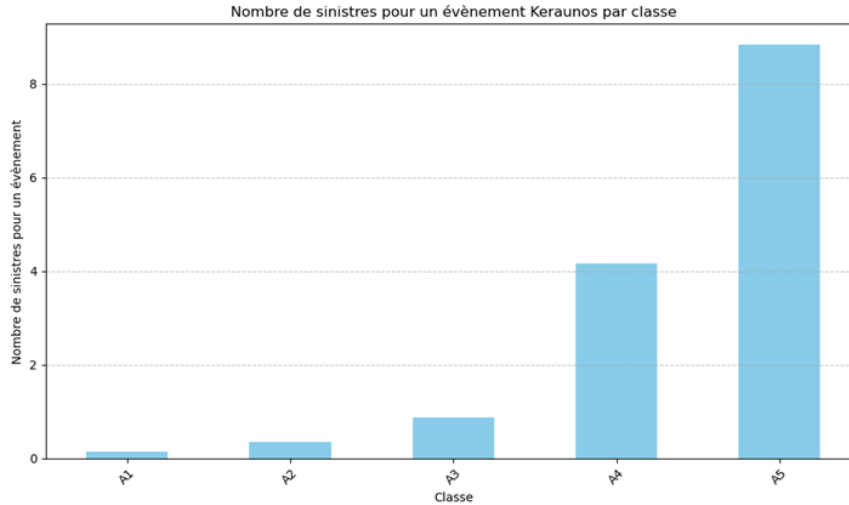


FIGURE 34 – Nombre de sinistres pour un évènement Keraunos par classe

Sur le graphique ci-dessus, le nombre de sinistres pour un événement Keraunos par classe est représenté. Pour obtenir chaque valeur des événements de types A1, A2, A3, A4 et A5, le nombre de sinistres est divisé par le nombre total d'événements de cette classe. Cela permet de calculer le nombre théorique de sinistres qu'une commune aurait dû avoir en fonction des événements de grêle qu'elle a subis. Si une commune a connu deux événements A1 et un événement A3 en 2019, elle aurait théoriquement eu $2 \times 0,15 + 1 \times 0,88 = 1,18$ sinistres.

Pour éviter de pénaliser les grandes communes en superficie, susceptibles d'avoir plus d'événements de grêle, le score obtenu est divisé par la superficie de la commune (en hectares). La superficie de chaque commune a été obtenue à partir des bases de données en libre accès fournies par l'INSEE. Par exemple, pour une commune de 500 hectares, le calcul serait $\frac{1,18}{500} = 0,00236$, représentant le nombre théorique de sinistres par hectare.

À partir de ces calculs, les communes sont réparties suivant leur nombre théorique de sinistres par hectare dans les quatre groupes suivants :

- **Groupe 1** : les communes n'ayant eu aucun événement Keraunos,
- **Groupe 2** : les 50 % des communes avec les scores les plus faibles, mais non nuls,
- **Groupe 3** : les communes dont les scores les situent entre 50 % et 85 %,
- **Groupe 4** : les 15 % avec les scores les plus élevés.

Ce découpage a été retenu après plusieurs essais, notamment celui consistant à répartir 33 % des communes dans chaque groupe, car il offre une meilleure distinction des communes fortement exposées aux événements de grêle. Ensuite, pour affiner la classification, ces groupes sont croisés avec le nombre de pièces des habitations.

Ensuite, pour affiner la classification, ces groupes sont croisés avec le nombre de pièces des habitations.

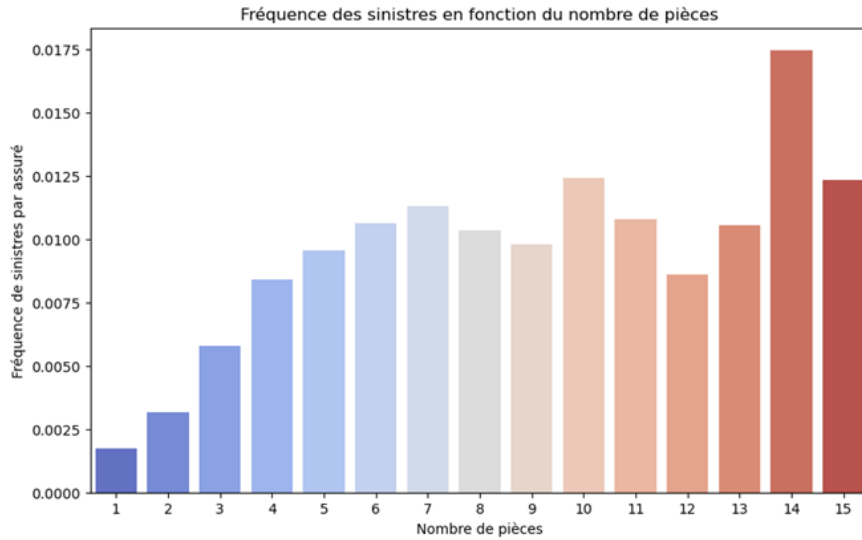


FIGURE 35 – Fréquence des sinistres en fonction du nombre de pièces

La fréquence des sinistres en fonction du nombre de pièces est analysée, et trois groupes supplémentaires sont créés, tout en respectant des critères d'homogénéité au sein des groupes, d'hétérogénéité entre eux et en s'assurant qu'un nombre suffisant d'assurés soit représenté dans chaque groupe.

Nb_pieces	Nb_sinistres	Nb_assurés	Fréquence
1	22	12537.836986	0.001755
2	152	47641.193637	0.003191
3	711	122722.977813	0.005794
4	1959	232332.071368	0.008432
5	2257	236174.432592	0.009556
6	1778	166802.914844	0.010659
7	980	86557.241455	0.011322
8	460	44315.149338	0.010380
9	217	22105.864623	0.009816
10	131	10551.270222	0.012416
11	53	4913.920174	0.010786
12	34	3939.318871	0.008631
13	5	472.890980	0.010573
14	7	400.927861	0.017460
15	3	242.806572	0.012356

FIGURE 36 – Tableau récapitulatif de la fréquence des sinistres selon le nombre de pièces

Ces trois groupes sont définis comme suit :

- **Groupe 1** : maisons de **1 à 3 pièces**, correspondant aux petites maisons avec les fréquences de sinistres les plus faibles,
- **Groupe 2** : maisons de **4 à 7 pièces**, représentant les maisons de taille moyenne, avec des fréquences relativement stables autour de 1%. Au-delà de 8 pièces, le nombre d'assurés chute et la fréquence devient plus volatile,
- **Groupe 3** : maisons de **8 pièces ou plus**, correspondant aux grandes maisons. La fréquence des sinistres dans ce groupe fluctue davantage en raison du faible nombre d'assurés, ce qui accentue l'impact d'un sinistre supplémentaire ou en moins.

En combinant les groupes créés à partir des événements de grêle et ceux basés sur le nombre de pièces, un total de 12 groupes est obtenu. Le modèle de Bühlmann-Straub est ensuite appliqué pour chaque groupe, et les résultats suivants sont obtenus :

Types d'événements \ Nb pièces	1 à 3 pièces	4 à 7 pièces	8 pièces ou plus
Pas d'événements	0,092%	0,101%	0,243%
Basse fréquence	0,291%	0,480%	0,670%
Moyenne fréquence	0,659%	1,303%	1,427%
Haute fréquence	3,509%	7,175%	4,654%

FIGURE 37 – Fréquence prédite par le modèle de Bühlmann-Straub en fonction du nombre de pièces et du nombre d'événements de grêle subis

A noter que le groupe des habitations de 8 pièces ou plus, pour les événements à haute fréquence (4,654% de fréquence), est plus limité en nombre d'assurés. Par conséquent le coefficient du portefeuille global pèse davantage que celui du groupe spécifique dans la valorisation de la fréquence. Cela explique une fréquence moins élevée par rapport aux classes ayant un nombre de pièces inférieur dans le groupe des événements les plus fréquents.

Modélisation du coût moyen d'un sinistre

Un travail similaire est réalisé pour modéliser le coût moyen d'un sinistre. Les groupes sont constitués en fonction des départements et du nombre de pièces. L'analyse commence par le coût moyen d'un sinistre par département.

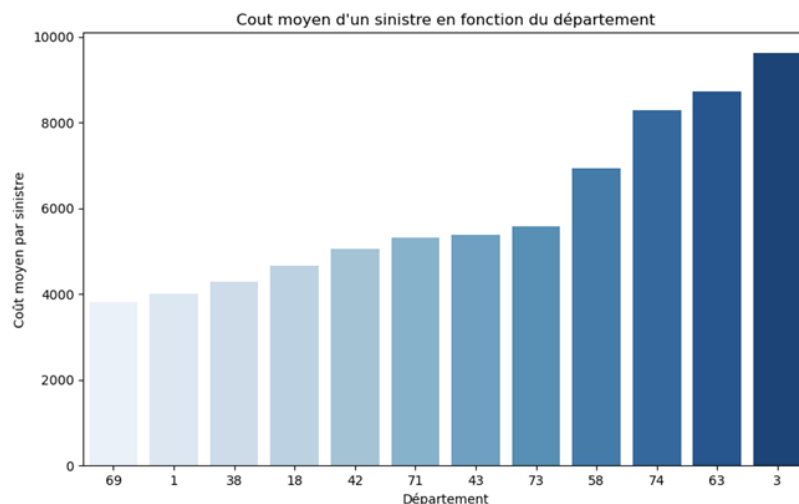


FIGURE 38 – Coût moyen d'un sinistre en fonction du département

- **Groupe 1** : les 4 départements ayant le coût moyen le plus faible (1, 18, 38, 69)
- **Groupe 2** : les 4 départements intermédiaires (42, 43, 71, 73)
- **Groupe 3** : les 4 départements ayant le coût moyen le plus élevé (3, 58, 63, 74)

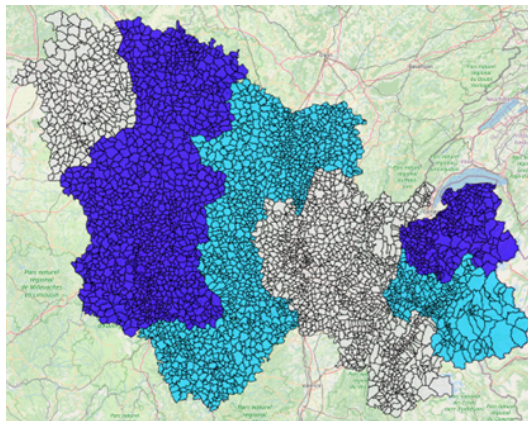


FIGURE 39 – Cartographie avec la répartition géographique des 3 groupes

Une cartographie présentant la répartition géographique des trois groupes est élaborée. Chaque groupe regroupe trois départements frontaliers et un département non frontalier, renforçant ainsi la cohérence de cette classification.

Pour former les groupes finaux, ces regroupements départementaux sont croisés avec le nombre de pièces. Le coût moyen par sinistre est alors analysé en fonction du nombre de pièces.

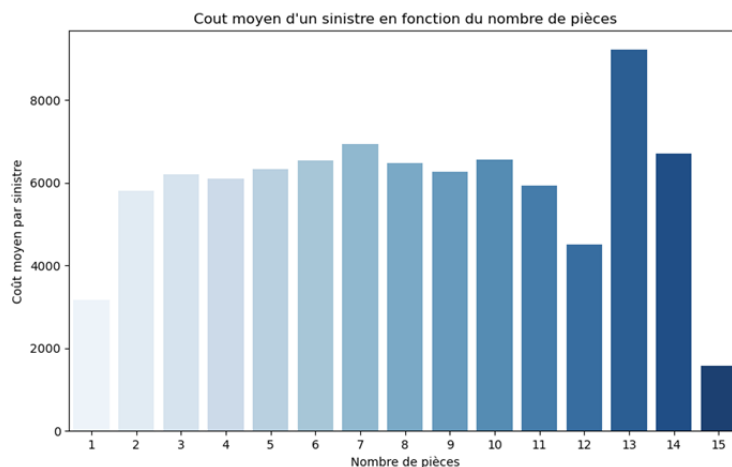


FIGURE 40 – Cout moyen d'un sinistre en fonction du nombre de pièces

À partir de ce graphique, trois groupes sont créés, en tenant compte des trois contraintes suivantes :

1. Le coût moyen au sein de chaque groupe doit être homogène ;
2. Les groupes doivent être suffisamment distincts en termes de coût moyen ;
3. Le nombre d'assurés dans chaque groupe doit être suffisant, sans quoi le portefeuille global influencera trop fortement les résultats de la théorie de la crédibilité.

Les 3 groupes sont définis comme suit :

- **Groupe 1** : maisons de **1 à 4 pièces**, correspondant aux petites maisons avec le coût moyen par sinistre le plus faible ;
- **Groupe 2** : maisons de **5 à 7 pièces**, représentant des maisons de taille moyenne avec des coûts moyens similaires, autour de 6600 € ;
- **Groupe 3** : maisons de **8 pièces ou plus**, correspondant aux grandes maisons. Le coût moyen fluctue fortement dans ce groupe en raison du faible nombre d'assurés, chaque sinistre ayant un impact considérable sur le coût moyen.

En combinant ces groupes par département et par nombre de pièces, 9 groupes distincts sont obtenus. Les résultats finaux sont ensuite déterminés à l'aide du modèle de Bühlmann-Straub.

Départements\Nb pièces	1 à 4 pièces	5 à 7 pièces	8 pièces ou plus
1, 18, 38, 69	4 195,03€	4 464,88€	5 146,84€
42, 43, 71, 73	4 680,05€	5 509,19€	6 300,12€
3, 58, 63, 74	8 485,01€	9 119,58€	7 316,64€

FIGURE 41 – Cout moyen prédit par le modèle de Bühlmann-Straub en fonction du département et du nombre de pièces

Comme pour la fréquence, dans le groupe des maisons de 8 pièces et plus situées dans les départements 3, 58, 63 et 74, le nombre limité d'assurés fait que le coefficient du portefeuille global pèse plus lourdement que celui du groupe spécifique dans la valorisation du coût moyen. Cela explique un coût moyen moins élevé dans ce groupe comparé à celui des maisons ayant un nombre de pièces inférieur dans les départements du 3e groupe.

Combinaison de la fréquence et du coût moyen

À partir des deux tableaux obtenus, chaque assuré se voit attribuer une fréquence de sinistres et un coût moyen. La multiplication de ces deux valeurs permet d'obtenir une prime pure par assuré. La prime pure d'une commune est supposée correspondre à la moyenne des primes des assurés. En d'autres termes, ce montant détermine la prime pure d'un assuré type dans cette commune.

Les communes sont ensuite classées par ordre croissant de prime pure, avec une attribution de notes de risque en utilisant la même classification que celle appliquée dans le GLM, conformément aux critères définis par le client. La classification utilisée est la suivante :

- $X \leq q(0.5)$: **risque faible** pour toute prime inférieure ou égale au quantile d'ordre 0.5 ;
- $X \in]q(0.5); q(0.85)]$: **risque modéré** pour toute prime strictement supérieure au quantile d'ordre 0.5 et inférieure ou égale au quantile d'ordre 0.85 ;
- $X > q(0.85)$: **risque élevé** pour toute prime strictement supérieure au quantile d'ordre 0.85 ;

2.4.5 Résultats et interprétations

Le zonier obtenu, basé sur cette classification, est illustré ci-dessous :

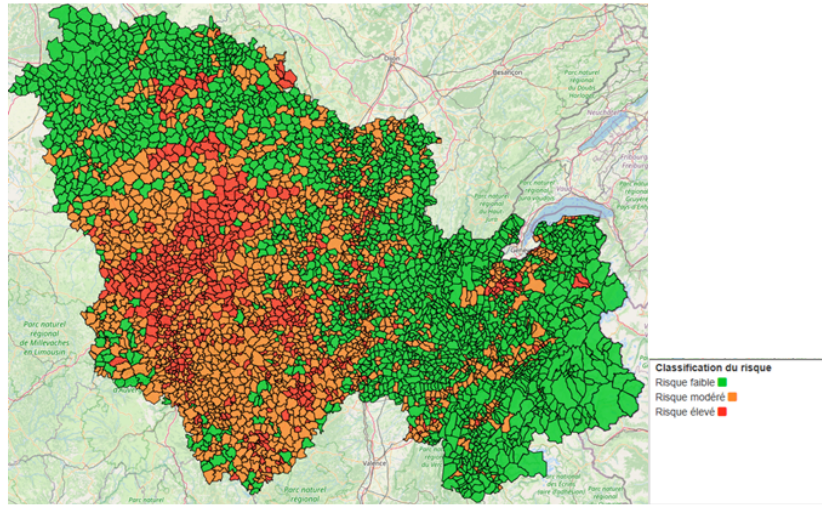


FIGURE 42 – Zonier basé sur la théorie de la crédibilité de l'exposition de l'assureur MRH au risque grêle pour la région Auvergne-Rhône-Alpes

Ce zonier met en évidence des zones distinctes caractérisant le risque grêle, mais présente une dispersion : certaines communes classées en zone verte (risque faible) sont entourées de communes rouges (risque élevé), et inversement. Ce phénomène peut résulter d'un manque ou d'un excès de données dans certaines communes. Pour atténuer cette dispersion, le même lissage géographique que celui appliqué au GLM a été utilisé.

Après l'application de ce lissage, la même classification des risques est reconduite, permettant d'obtenir le zonier suivant :

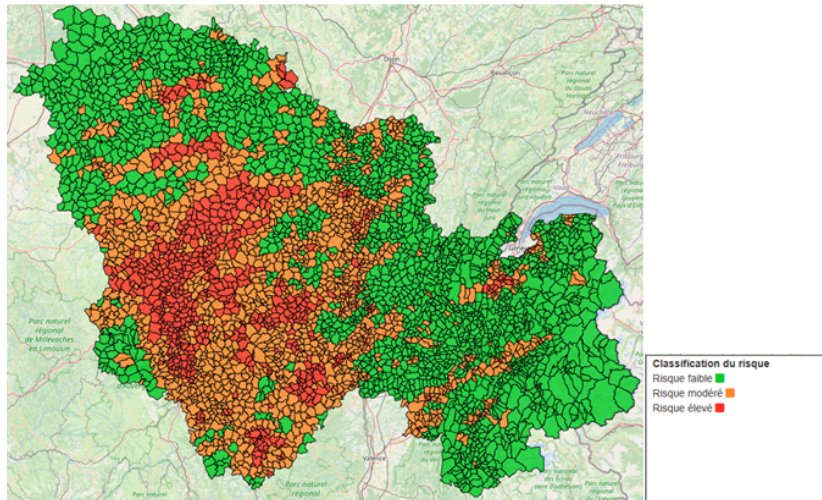


FIGURE 43 – Zonier après lissage basé sur la théorie de la crédibilité de l'exposition de l'assureur MRH au risque grêle pour la région Auvergne-Rhône-Alpes

Grâce au lissage, les communes entourées par des communes ayant des classifications de risque différentes deviennent beaucoup plus rares, ce qui améliore la cohérence générale du zonier. Les zones de risque élevé restent bien identifiées, renforçant ainsi la pertinence du modèle.

2.5 Comparaison des 3 méthodes

Après avoir exploré trois approches distinctes pour la construction d'un zonier grêle, par itération, via GLM et à partir de la théorie de la crédibilité, il est essentiel de les comparer en termes de méthodologie, de robustesse des résultats et d'applicabilité opérationnelle. Cette comparaison permet de mieux comprendre les avantages et les limites de chaque méthode afin d'éclairer le choix final du zonier, utilisé pour la tarification du produit MRH et la définition de la politique de souscription. Les critères de comparaison incluent notamment la simplicité de mise en œuvre, la précision des estimations, et la capacité à capturer les spécificités locales du risque grêle.

2.5.1 Avantages et inconvénients de chaque méthode

Zonier par itération

L'approche par itération a l'avantage d'être relativement simple à mettre en œuvre. Elle repose sur des choix qualitatifs et des analyses basées sur des tables de croisement, ce qui la rend intuitive et facilement compréhensible pour des acteurs non experts. Cette méthode permet de structurer les données de manière progressive et offre une visualisation accessible des zones à risque.

Cependant, elle présente plusieurs limites. Sa précision peut être insuffisante lorsque les données sont complexes ou présentent une grande variabilité. De plus, elle intègre faiblement les spécificités du portefeuille, s'appuyant principalement sur les données météorologiques. Ce manque de prise en compte des particularités des biens assurés peut réduire la pertinence des résultats obtenus dans notre cas.

Zonier par GLM

La construction du zonier à l'aide de modèles linéaires généralisés repose sur une approche statistique rigoureuse. En intégrant différentes variables explicatives, cette méthode permet de combiner des données météorologiques et des caractéristiques des biens, ce qui améliore la capacité à identifier les facteurs augmentant le risque de sinistres grêle. Les résultats produits pourraient non seulement être précis, mais également interprétables, car chaque variable explicative a un effet mesurable sur la variable cible, offrant ainsi des résultats très exploitables.

Cependant, cette méthodologie exige une certaine expertise technique, notamment pour le choix des variables explicatives et la validation des modèles. Dans le cadre de ce mémoire, le nombre limité de variables explicatives disponibles restreint la qualité des prédictions, et l'ajout de variables supplémentaires aurait pu améliorer les performances. Par ailleurs, les résultats des tests de validation, notamment pour la modélisation du coût moyen, ont révélé des limites, principalement dues à la forte volatilité des coûts des sinistres. Cette variabilité reste un défi pour l'application des GLM dans ce contexte.

Zonier par théorie de la crédibilité

La méthode fondée sur la théorie de la crédibilité, et plus particulièrement sur le modèle de Bühlmann-Straub, se distingue par sa capacité à produire des estimations fiables dans des contextes où les données sont limitées. Elle combine habilement les informations spécifiques à chaque assuré avec des données globales, ce qui permet d'obtenir des résultats robustes, même pour des zones peu exposées ou disposant de peu de données historiques. Une fois appliquée, cette approche fournit des résultats faciles à interpréter, notamment grâce aux regroupements effectués par classes de risques.

Cependant, cette méthode peut paraître plus abstraite que les autres et nécessite une bonne maîtrise des concepts sous-jacents pour une mise en œuvre correcte. La qualité des résultats dépend en grande partie de la pertinence des groupes de risques définis en amont, rendant cette étape cruciale. Enfin, bien que la méthode permette une grande adaptabilité, son efficacité est sensible à la cohérence et à la fiabilité des données utilisées.

2.5.2 Comparaison avec les cartographies d'évènements Keraunos

Après avoir présenté les avantages et inconvénients des trois méthodes, la comparaison se concentre sur les résultats obtenus par le GLM et la théorie de la crédibilité, la première méthode ayant été écartée. Cette analyse s'appuie sur une comparaison des cartographies générées avec celles représentant le nombre d'évènements de grêle recensés par Keraunos, classés par taille de grêlons.

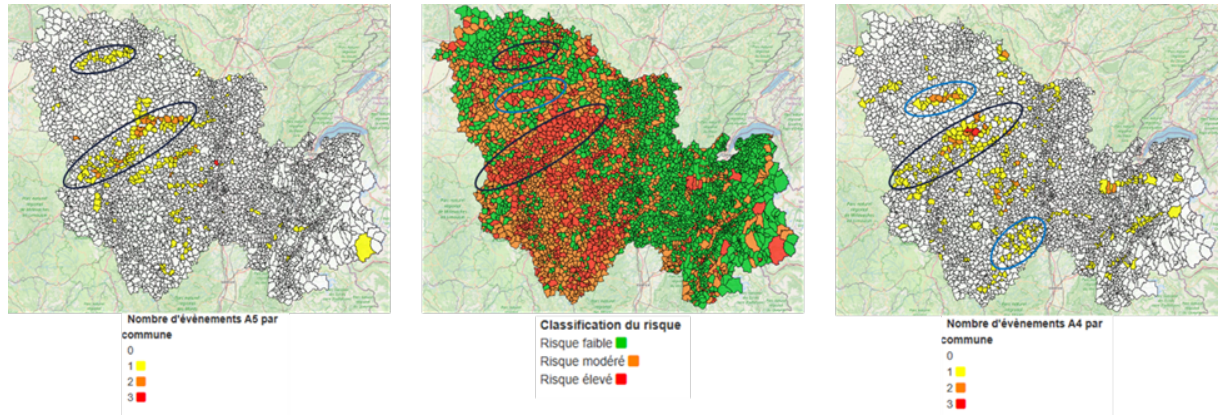


FIGURE 44 – Cartes du nombre d'évènements A5 à gauche, A4 à droite, et le zonier par GLM de l'exposition de l'assureur MRH au risque grêle pour la région Auvergne-Rhône-Alpes au centre

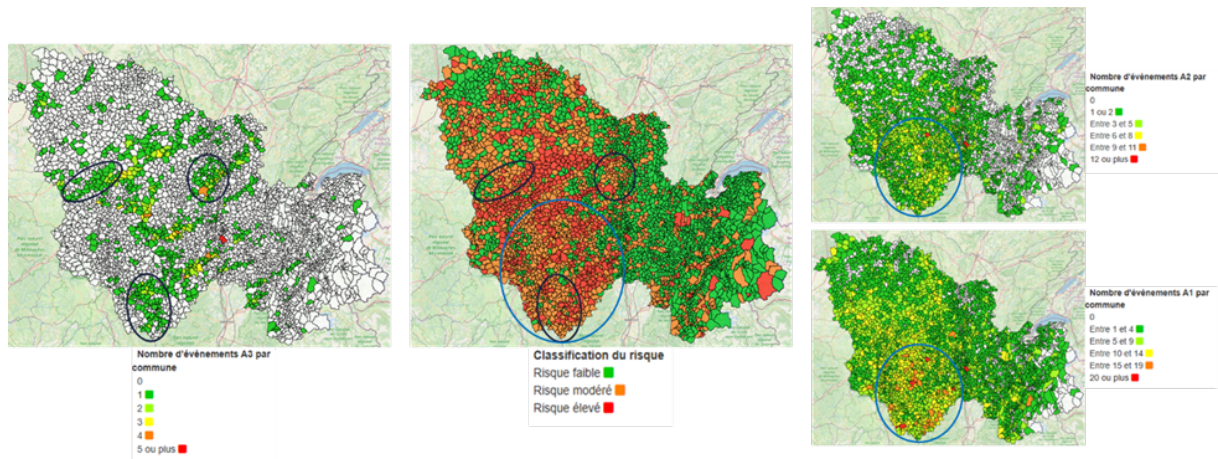


FIGURE 45 – Cartes du nombre d'évènements A3 à gauche, A2 en haut à droite, A1 en bas à droite et le zonier par GLM de l'exposition de l'assureur MRH au risque grêle pour la région Auvergne-Rhône-Alpes au centre

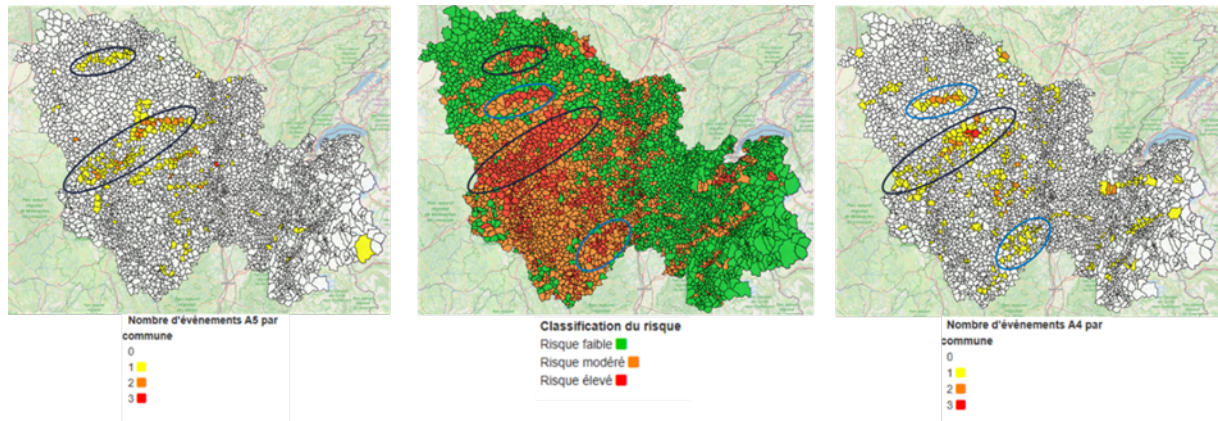


FIGURE 46 – Cartes du nombre d'évènements A5 à gauche, A4 à droite et le zonier par théorie de la crédibilité de l'exposition de l'assureur MRH au risque grêle pour la région Auvergne-Rhône-Alpes au centre

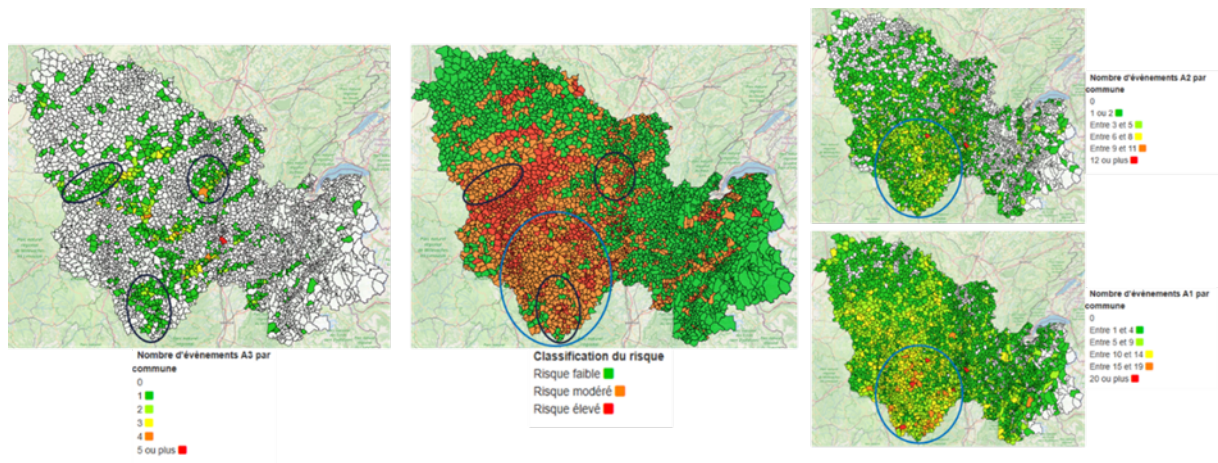


FIGURE 47 – Cartes du nombre d'évènements A3 à gauche, A2 en haut à droite, A1 en bas à droite et le zonier par théorie de la crédibilité de l'exposition de l'assureur MRH au risque grêle pour la région Auvergne-Rhône-Alpes au centre

Sur ces différentes cartes, les mêmes zones ont été encerclées, tant sur le zonier construit par GLM que sur celui basé sur la théorie de la crédibilité, afin de mettre en évidence les régions affectées par des événements de grêle recensés par Keraunos. Globalement, les événements de grêle de Keraunos correspondent bien aux zones à risque identifiées dans les deux zoniers. Cependant, des divergences notables apparaissent entre les deux méthodes. Le zonier construit par GLM montre une disparité plus marquée avec des communes en risque modéré et élevé un peu partout dans la région, tandis que celui basé sur la théorie de la crédibilité produit des zones plus homogènes.

Le choix final du zonier sera effectué dans la sous-section suivante, en tenant compte de ces observations et des besoins opérationnels de l'assureur.

2.5.3 Choix final du zonier

Après avoir comparé les résultats avec les cartographies des événements recensés par Keraunos, le choix du zonier final peut être effectué. Celui-ci, qui sera utilisé dans la politique de souscription décrite dans la prochaine partie, est le zonier lissé issu des résultats de la théorie de la crédibilité. Ce choix repose sur plusieurs raisons :

- Les événements de grêle identifiés par Keraunos sont bien captés par la cartographie du risque grêle.
- La théorie de la crédibilité se révèle particulièrement adaptée lorsque les données disponibles sont limitées, comme c'est le cas dans cette étude.
- La cartographie obtenue est plus homogène que celle produite par le GLM, critère recherché dans ce contexte.
- Les étapes de validation des modèles GLM n'ont pas donné des résultats suffisamment significatifs, notamment pour le coût moyen.
- Le lissage appliqué permet de combler les lacunes liées à l'absence d'informations dans certaines communes, qu'il s'agisse des événements recensés par Keraunos ou des données sur les habitations.

La cartographie du zonier final est rappelée ci-dessous :

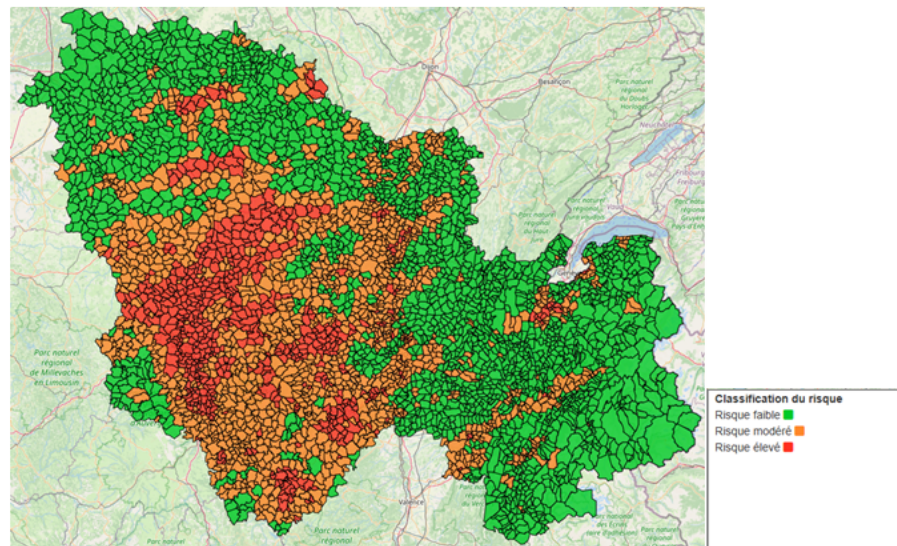


FIGURE 48 – Zonier après lissage basé sur la théorie de la crédibilité de l'exposition de l'assureur MRH au risque grêle pour la région Auvergne-Rhône-Alpes

3 Définition d'une politique de souscription à partir du zonier

Après avoir défini les contours du phénomène de la grêle et élaboré un zonier pertinent grâce à des approches méthodologiques variées, cette troisième partie se concentre sur la mise en application du zonier dans la stratégie assurantielle. Les objectifs principaux sont multiples :

- **Améliorer la précision tarifaire**, en adaptant les primes d'assurance à la fréquence et à la gravité des sinistres spécifiques à chaque zone pour assurer une équité tarifaire et réduire les déséquilibres techniques,
- **Elaborer une politique de souscription** en testant et proposant des stratégies basées sur le zonier, incluant la gestion des expositions, le contrôle des coûts et le ciblage de zones pour équilibrer le portefeuille,
- **Anticiper les évolutions climatiques et réglementaires** en testant des scénarios prospectifs pour évaluer leur impact sur la sinistralité et la gestion du portefeuille.

Pour atteindre ces objectifs, une première section présente le cycle de vie d'un produit d'assurance, où la politique de souscription joue un rôle déterminant. Son rôle est ensuite approfondi par le spectre de la tarification d'une garantie d'assurance grêle, qui utilise le zonier pour proposer des primes adaptées à chaque zone. Enfin, des scénarios prospectifs intègrent les défis posés par le dérèglement climatique et les changements réglementaires, afin de tester la robustesse et la pertinence des stratégies proposées.

3.1 Cycle de vie d'un produit d'assurance

Dans cette section, le cycle de vie d'un produit d'assurance est exploré à travers le schéma ci-dessous, qui met en lumière les étapes clés, de la conception initiale jusqu'au suivi opérationnel. Cette présentation du cycle de vie d'un produit d'assurance permet de mieux comprendre les interactions entre ces différentes étapes et leur rôle dans la mise en œuvre d'une politique de souscription efficace et durable.

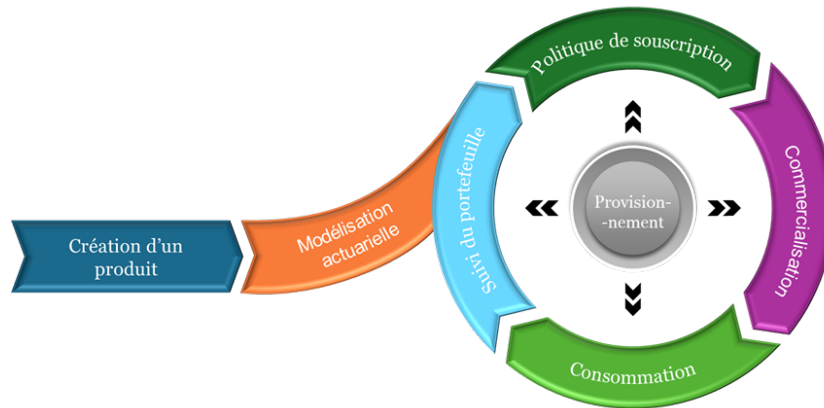


FIGURE 49 – Cycle de vie d'un produit d'assurance

Ensemble, ces étapes forment un processus cyclique et dynamique, permettant à l'assureur de s'adapter continuellement aux évolutions du marché, aux changements réglementaires, ainsi qu'aux variations dans les comportements des assurés et les conditions climatiques. Ce cycle de vie fournit ainsi un cadre structuré pour optimiser la gestion des produits d'assurance et répondre aux défis actuels du secteur.

Dans la suite, chaque étape est présentée, avec une attention particulière portée sur la politique de souscription.

3.1.1 Création d'un produit

Le cycle débute par la création du produit, une étape où l'assureur identifie les besoins du marché et conçoit une offre adaptée. La création d'un nouveau produit permet à l'assureur de :

- **Développer son activité** : Le développement de nouveaux produits permet d'élargir le portefeuille et de se positionner sur le marché. Si le produit correspond à un segment niche ou un marché peu exploité alors, il offre l'opportunité d'attirer une clientèle spécifique, souvent négligée par les produits standards. Sinon, le produit permet de se positionner sur des offres dites « classiques » qui permettent de lui augmenter d'une part la taille de sa clientèle et d'autre part son chiffre d'affaires.
- **Répondre aux évolutions du marché** : En étudiant les tendances et en anticipant les risques émergents, l'assureur peut concevoir des produits innovants et pertinents pour sa clientèle. Ceux-ci peuvent résulter de l'identification d'un besoin spécifique ou de la volonté de susciter un nouveau besoin.
- **Diversifier son portefeuille** : Un nouveau produit permet de diversifier l'exposition de l'assureur à ses risques. Cela permet de générer du capital. En effet, sous le cadre réglementaire de Solvabilité II, la diversification des risques permet de réduire le capital requis pour couvrir ces risques. En répartissant les expositions sur des segments variés, l'assureur bénéficie d'une meilleure mutualisation, ce qui diminue les corrélations entre les risques et réduit le montant du capital de solvabilité requis. Cette optimisation des fonds propres libère des ressources pour des investissements stratégiques et améliore la rentabilité globale de l'entreprise.
- **Renforcer son image de marque** : La création d'un produit s'inscrit également dans la vision et les valeurs de l'entreprise. Proposer un produit en cohérence avec les engagements éthiques ou environnementaux de l'assureur consolide l'image de marque et attire les clients sensibles à ces valeurs. Dans un monde en perpétuelle évolution, où les entreprises sont jugées sur leur capacité à répondre aux préoccupations sociétales, cette approche devient un véritable levier de différenciation.

En somme, la création d'un produit d'assurance repose sur une analyse des besoins, une innovation constante, et une cohérence avec la stratégie globale de l'entreprise. Cette phase inclut également des premières analyses stratégiques pour évaluer la faisabilité commerciale et technique du produit, ainsi que la construction d'un business plan détaillé. Celui-ci projette le chiffre d'affaires espéré, le ratio sinistres à primes prévisionnel, le retour sur investissement, et d'autres indicateurs clés pour aligner les objectifs de rentabilité sur les capacités opérationnelles et le positionnement de l'assureur.

Dans ce mémoire, le produit MRH existe et est connu par le marché. L'intérêt est essentiellement porté sur la couverture contre les sinistres liés à la grêle dans le cadre de cette assurance.

3.1.2 Modélisation

La modélisation suit cette phase et vise à valider la faisabilité du produit et le business plan associé. Cette étape comprend une analyse approfondie du marché afin d'assurer la pertinence du produit vis-à-vis de l'offre existante, tout en intégrant les aspects juridiques et réglementaires nécessaires pour garantir la conformité du produit. Ainsi, elle permet de définir clairement les spécificités du produit :

- **Faisabilité technique et financière du produit**

L'objectif de cette étape est de garantir la durabilité et la rentabilité du produit tout en s'inscrivant dans un business model cohérent. Ce dernier définit les principes économiques qui soutiendront le produit, en s'appuyant sur des projections de revenus, des hypothèses de coûts, une stratégie de rentabilité à long terme et des études de marché.

- **Inscription dans le cadre réglementaire**

La modélisation intègre également une dimension juridique et réglementaire pour garantir la conformité de la proposition commerciale. Chaque nouveau produit d'assurance doit respecter les normes locales et européennes, un prérequis essentiel pour obtenir l'agrément des autorités de régulation. Ce processus peut nécessiter des ajustements dans le business model, comme l'adaptation des garanties, des exclusions ou des critères d'éligibilité, afin de répondre aux attentes des régulateurs tout en maintenant l'attractivité commerciale du produit.

- **Définition des spécificités du produit**

Lors de la modélisation, le périmètre d'application de la couverture est défini. Ces critères incluent par exemple la segmentation des risques selon des facteurs tels que la localisation géographique, la typologie des biens assurés ou encore l'historique de sinistralité.

Dans ce mémoire, le produit MRH en général et la couverture grêle en particulier sont déjà modélisés. Toutefois, les études réalisées, notamment la création d'un zonier, permettent d'apporter des éléments de réflexion à l'assureur quant au raffinement du périmètre de couverture établi (population et zones couvertes, plafond de garantie défini, éventuelle intégration d'une franchise, etc.).

3.1.3 Politique de souscription

L'un des objectifs de ce mémoire est d'utiliser le zonier afin de définir une politique de souscription propre à la région étudiée. Ainsi, cette partie présente cette étape de manière détaillée.

La politique de souscription suit la modélisation et représente un pilier central du cycle de vie. Elle vise à définir les critères de sélection des assurés et à établir les modalités de tarification qui respectent les objectifs de rentabilité et le modèle économique de l'entité. Elle a pour objectif de traduire et concrétiser le produit modélisé de telle sorte à garantir la compétitivité des produits tout en assurant leur rentabilité. Pour cela, la politique de souscription doit :

- **Définir un tarif juste**

La tarification doit s'appuyer sur des modèles mathématiques tels que les GLM, la théorie de la crédibilité ou le machine learning. Ces modèles permettent d'estimer les primes pures en fonction du risque associé à chaque assuré. Ensuite, la prime commerciale, est définie de telle sorte à être suffisamment compétitive pour attirer les clients, tout en garantissant une couverture adéquate des risques. Une prime bien calibrée permet de séduire un large éventail de profils d'assurés, contribuant ainsi à une mutualisation efficace du risque. Dans le cadre de ce mémoire, le tarif est élaboré à partir d'un zonier, et est ajusté selon les spécificités géographiques et les caractéristiques de chaque bien, afin de refléter le coût réel des sinistres liés à la grêle.

- **Être en cohérence avec les tendances du marché**

Cela implique une veille régulière des offres concurrentielles, permettant d'ajuster les tarifs et de positionner les produits de manière attractive.

- **Traduire les spécificités du produit (déterminées lors de la modélisation) en critères de sélection**

Les critères de sélection doivent être soigneusement définis, en tenant compte d'éléments tels que les caractéristiques de la propriété, l'historique des sinistres et le profil de l'assuré afin de cibler le bon profil de clients. Un équilibre approprié entre le risque accepté et les primes perçues est essentiel pour garantir la rentabilité du produit.

- **Concevoir une offre compétitive sur le marché et équilibrée entre les différents produits et les différentes garanties.**
- **Être alignée avec le business plan validé par l'ACPR pour l'obtention de l'agrément.**

3.1.4 Commercialisation

Après avoir établi une politique de souscription solide, il est essentiel de mettre en œuvre une stratégie de commercialisation et de distribution efficace pour positionner le produit sur le marché. La distribution joue un rôle central dans le cycle de vie du produit, influençant directement son adoption par les assurés et sa pérennité. Elle englobe les canaux de vente, les relations avec les intermédiaires, et les outils mis à disposition des équipes commerciales pour promouvoir le produit auprès des clients cibles.

Au-delà de la distribution, la commercialisation doit intégrer des initiatives marketing ciblées pour attirer et fidéliser les clients. Par exemple, des campagnes de publicité numérique et des offres promotionnelles peuvent être mises en place, telles que des réductions pour les nouveaux clients ou des avantages pour ceux qui recommandent des proches. Ces incitations contribuent à renforcer l'attractivité des produits d'assurance, à améliorer la compétitivité de l'offre et à fidéliser les clients de l'assureur, avec la consommation d'autres produits.

Un autre aspect de la commercialisation réside dans le fait de rester en ligne avec les indications de l'ACPR et du business plan établi. La mutualisation du risque constitue une stratégie clé, consistant à rassembler un mélange de "bons" et de "mauvais" profils au sein du portefeuille d'assurances. En intégrant des assurés présentant des caractéristiques de risque variées, il est possible d'optimiser la rentabilité globale du produit. Ce mélange permet de compenser les sinistres potentiels des profils à risque élevé par les primes versées par les profils plus sûrs.

En somme, la commercialisation d'un produit d'assurance repose sur une approche stratégique combinant une distribution efficace, des initiatives marketing pertinentes et une gouvernance alignée sur les réglementations. En optimisant ces éléments, il devient possible de créer une offre attractive, compétitive et rentable, répondant aux attentes des assurés tout en assurant la durabilité du portefeuille d'assurances.

Dès lors que le produit est sur le marché, la consommation représente le recours aux prestations par les assurés, incluant le versement des prestations en cas de sinistres.

3.1.5 Suivi du portefeuille

A mesure que le produit est consommé, le suivi du portefeuille devient indispensable. Cette étape permet de surveiller les indicateurs de performance, tels que le S/P réalisé, et de réajuster les politiques de souscription si nécessaire pour maintenir l'équilibre financier du produit. Concrètement, cela requiert de :

- **Analyser les affaires nouvelles et le chiffre d'affaires**

L'analyse des affaires nouvelles permet de mesurer le volume des contrats souscrits, la répartition géographique des nouveaux assurés et le profil des risques intégrés dans le portefeuille. Ces indicateurs sont essentiels pour évaluer l'attractivité du produit et identifier d'éventuelles disparités par rapport aux objectifs fixés lors de sa conception.

Parallèlement, le chiffre d'affaires réalisé est comparé aux prévisions afin de vérifier si les hypothèses initiales, notamment en termes de primes perçues et de part de marché visée, se confirment dans la réalité. Toute déviation significative par rapport au business plan peut indiquer des problèmes dans la stratégie de commercialisation, une inadéquation de la tarification ou une concurrence plus forte que prévu. Cette comparaison permet également d'anticiper d'éventuelles tensions sur la rentabilité et d'ajuster les efforts commerciaux ou les politiques tarifaires en conséquence.

- **Analyser les résiliations et les dérives**

L'analyse et la compréhension des résiliations peuvent aider à développer des stratégies de fidélisation plus ciblées pour retenir les assurés à risque de résiliation. En complément, l'étude des dérives manifeste une augmentation inattendue de la fréquence ou de la gravité des sinistres, des primes moyennes inférieures à la couverture inposée ou une concentration excessive de risques dans certaines zones ou segments.

- **Suivre des indicateurs clés comme le rapport S/P et le ratio combiné**

Le ratio S/P offre une vision de la rentabilité technique des produits d'assurances. Le ratio combiné le complète en intégrant les frais et les commissions.

- **Anticiper et ajuster l'offre**

Les actions précédentes orientent les mesures de pilotage que l'assureur pourrait mettre en place. Celui-ci peut, par exemple, indexer ses primes ou réviser sa politique de souscription. Une politique de souscription flexible et réactive, intégrée dans une démarche globale de pilotage, permet de maintenir la compétitivité et la rentabilité des produits tout en s'adaptant aux contraintes du marché.

À plus long terme, ces analyses justifient des révisions du business plan. Qu'il s'agisse de réaligner les objectifs de rentabilité, de rééquilibrer le portefeuille ou d'explorer de nouvelles niches de marché, le pilotage basé sur ces ratios garantit une gestion proactive et durable du portefeuille.

Dans le cadre de ce mémoire, le suivi du portefeuille de la garantie n'est pas exploré.

3.1.6 Provisionnement

Enfin, le provisionnement constitue le cœur du cycle de vie en garantissant que l'assureur dispose des réserves financières nécessaires pour honorer ses engagements, et ce, tout au long du cycle de vie du produit. Ce provisionnement doit être ajusté en continu pour faire face aux évolutions des risques et aux nouvelles données de sinistralité.

Le provisionnement se justifie par l'inversion du cycle de production en assurance : le produit est vendu avant que son coût final ne soit connu. Cela nécessite de réserver une part des primes perçues pour couvrir les engagements futurs liés aux sinistres qui surviendront.

Le provisionnement diffère selon les cadres normatifs : normes comptables locales, directive Solvabilité II et normes IFRS 17. Chacun impose des exigences méthodologiques précises, et les entités doivent s'assurer de leur conformité aux règles applicables.

En normes sociales, le provisionnement repose sur la constitution de deux principales catégories de provisions :

- **Provisions pour primes** : Elles couvrent les engagements futurs relatifs aux primes encaissées mais non encore acquises ou non encore consommées pour la couverture des sinistres potentiels.
- **Provisions pour sinistres** : Elles incluent les montants nécessaires pour régler les sinistres survenus, qu'ils soient déclarés ou non, ainsi que ceux en cours de règlement ou sujets à aggravation.

Sous Solvabilité II, les provisions techniques sont calculées sur une base économique et actualisée, visant à refléter la réalité des engagements. Elles incluent :

- **La meilleure estimation (best estimate)** : montant actualisé des flux futurs de trésorerie liés aux contrats, fondé sur des hypothèses probabilistes.
- **Une marge de risque** : coût du capital que devrait lever le cessionnaire pour couvrir son exigence de capital jusqu'à l'extinction des passifs.

Sous IFRS 17, norme comptable pour les contrats d'assurance, à visée internationale propres à l'assurance, les provisions se distinguent entre :

- **Liability for Remaining Coverage (LRC)** : provision pour couverture future, qui reflète les flux associés aux primes encaissées pour des périodes non encore écoulées.
- **Liability for Incurred Claims (LIC)** : provision pour sinistres survenus.

Chacune des composantes LRC et LIC inclut un "best estimate". De plus, une spécificité notable d'IFRS 17 est la **Contractual Service Margin (CSM)**, qui représente les profits futurs attendus des contrats d'assurance. Cette marge, propre à la LRC, s'amortit progressivement en fonction de la prestation des services assurantiels, permettant une reconnaissance étalée des bénéfices.

Le provisionnement, qu'il s'agisse de constituer des provisions pour primes ou pour sinistres, représente un enjeu stratégique majeur. La cohabitation de ces cadres impose une gestion rigoureuse et une adaptation des pratiques aux exigences spécifiques de chaque norme. En respectant ces cadres réglementaires, les assureurs assurent la sécurité des engagements vis-à-vis des assurés tout en répondant aux attentes des régulateurs et des parties prenantes. Cela contribue également à renforcer la durabilité et la résilience de l'activité dans un environnement économique et normatif complexe.

Cette étape n'est pas traitée dans le cadre de ce mémoire.

3.2 Définir une politique de souscription à partir du zonier

Après avoir exploré les différentes étapes de la création d'un produit d'assurance, cette troisième partie du mémoire se concentre sur un élément clé : **la politique de souscription**, en mettant particulièrement l'accent sur la **tarification** des contrats face au risque grêle. La tarification des contrats d'assurance peut être abordée comme un programme d'optimisation sous contrainte, où il s'agit de proposer un tarif compétitif, attractif pour les assurés, tout en garantissant une rentabilité suffisante pour couvrir les sinistres et les charges de l'assureur. Cette section étudie la construction d'une grille tarifaire basée sur le zonier de la partie précédente, un système de segmentation du territoire en zones de risques. Dans un contexte de risque climatique croissant, tel que le risque grêle, la nécessité d'une tarification ajustée et évolutive prend une importance accrue.

3.2.1 Présentation de la tarification dans le cadre actuariel

Lors de la souscription d'un contrat, l'assureur s'engage à indemniser l'assuré en cas de sinistre contre le paiement d'une prime. Contrairement aux entreprises classiques qui connaissent leurs coûts de production avant de fixer leurs prix, un assureur établit une prime sans connaître le montant exact des sinistres à venir. La tarification repose donc sur des méthodes mathématiques sophistiquées visant à estimer la prime requise pour couvrir le risque tout en assurant la rentabilité de l'assureur.

Principes fondamentaux en assurance

La tarification en assurance peut être envisagée comme un programme d'optimisation sous contrainte, combinant la mutualisation et la segmentation des risques. La mutualisation, qui repose sur le regroupement d'un grand nombre de contrats, réduit la variabilité des sinistres en s'appuyant sur la loi des grands nombres. En parallèle, la segmentation des risques permet d'adapter les primes aux profils des assurés en formant des groupes homogènes, chaque groupe se voyant appliquer une prime correspondant à son niveau de risque. Cette approche cherche à équilibrer la diversité des risques tout en préservant les avantages de la mutualisation et l'équité entre assurés.

Objectifs de la tarification en assurance

La tarification en assurance dommages poursuit plusieurs objectifs pour satisfaire les besoins des assurés tout en assurant la rentabilité pour l'assureur :

- **Un tarif suffisant** : La prime doit couvrir l'ensemble des sinistres, assurer la pérennité de l'entreprise et garantir une marge de sécurité contre le risque de ruine.
- **Un tarif adapté** : La prime doit refléter les caractéristiques et la nature spécifiques des risques (franchises, plafonds de garantie, risques de pointe) afin de minimiser les asymétries d'information.
- **Un tarif compétitif** : La prime doit rester concurrentielle pour fidéliser les assurés ou en trouver de nouveaux, tout en assurant la rentabilité des fonds propres investis.

Notations et hypothèses du modèle de tarification simple

Soit un ensemble homogène de N contrats d'assurance offrant les mêmes garanties en échange d'une prime identique. En notant n le nombre de sinistres observés pendant l'exercice et C_i le coût du $i^{\text{ème}}$ sinistre pour $i = 1, \dots, n$, le coût total des sinistres s'exprime comme :

$$\sum_{i=1}^n C_i$$

En supposant que les risques assurés sont de même nature, identiques et indépendants, les variables aléatoires X_i représentant la charge annuelle des sinistres pour chaque assuré sont i.i.d avec espérance $\mathbb{E}(X_i) = \mu$ et variance $\text{Var}(X_i) = \sigma^2$. Ainsi, pour un grand nombre de sinistres n , le théorème central limite donne :

$$\frac{\sum_{i=1}^n X_i}{n} \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$$

La prime pure, correspondant au coût moyen par contrat, est définie par la loi des grands nombres comme :

$$\pi = \frac{\sum_{i=1}^n C_i}{N} \rightarrow \mathbb{E}(X_i)$$

En d'autres termes, la prime pure se calcule comme le produit du nombre moyen de sinistres par contrat et du coût moyen par sinistre :

$$\pi = \frac{n}{N} \cdot \frac{\sum_{i=1}^n C_i}{n}$$

Cependant, il est essentiel de distinguer deux types de primes : la prime pure et la prime commerciale.

Prime pure et prime commerciale

Contrairement à la prime pure, la prime commerciale tient compte de l'ensemble des frais de l'assureur, dont :

- **Frais d'acquisition** : commissions aux réseaux de distribution, frais des services chargés de la conception des contrats, marketing, etc.
- **Frais de règlement des sinistres** : gestion des sinistres, contentieux, recours, etc.
- **Frais d'administration** : gestion comptable, actuariat, juridique, direction générale, etc.
- **Frais financiers** : gestion des placements, courtage, conservation, etc.

Les actuaires utilisent un modèle mathématique pour estimer la prime pure qui correspond au montant minimal à exiger des assurés pour couvrir le risque. Ensuite, la prime pure est augmentée par un taux de chargement θ pour obtenir la prime commerciale PC, effectivement payée par les assurés :

$$PC = \pi + CH = \pi + \theta PC$$

Où θ représente le taux de chargement nécessaire pour couvrir les frais. Ce modèle permet à l'assureur de proposer une prime suffisante pour répondre aux exigences financières, tout en maintenant la compétitivité et l'équité tarifaire.

Après avoir examiné les bases actuarielles et les principes fondamentaux qui régissent le calcul des primes d'assurance, il est essentiel de se pencher sur des méthodes spécifiques de tarification. En particulier, la méthodologie de tarification basée sur le zonier permet d'affiner cette estimation en tenant compte des facteurs locaux et des caractéristiques des logements.

Ainsi, la tarification d'une garantie comme celle de la grêle, nécessite une évaluation précise du coût du risque, intégrant divers critères tels que la nature du logement et la zone géographique. Dans la section suivante, la proposition de tarification pour la garantie grêle en région Rhône-Alpes Auvergne sera présentée, en s'appuyant sur le zonier grêle établi précédemment.

3.2.2 Tarification de la garantie à partir du zonier

La tarification d'un produit d'assurance repose sur une estimation précise du coût du risque. Pour la garantie MRH, plusieurs critères influencent la tarification, tels que :

- **Nature du logement** : maison, appartement, mobil-home, etc.
- **Usage de l'habitation** : résidence principale, résidence secondaire, etc.
- **Nombre de pièces et/ou superficie du logement.**
- **Niveau du logement** : dans le cas des appartements, des surprimes peuvent être appliquées, par exemple pour les rez-de-jardin, qui sont plus exposés aux vols et aux inondations notamment.
- **Caractéristiques de l'assuré** : propriétaire, locataire, etc.
- **Zone géographique.**

Dans le cadre de ce mémoire, et en tenant compte des données disponibles, les éléments suivants décrivent la proposition de tarification :

- Proposition d'un tarif pour la garantie grêle dans le cadre d'une assurance MRH pour la région Rhône-Alpes Auvergne,
- Détermination des zones de risque basée sur le zonier grêle établi précédemment,
- Tarification de la garantie en fonction du nombre de pièces du logement et de la zone géographique.
- A destination des propriétaires de maisons.

L'importance d'un zonier dans la tarification d'une garantie grêle

La zone géographique est l'un des critères de tarification les plus couramment utilisés en assurance dommages, notamment pour les couvertures contre les événements naturels. Le zonier constitue un levier de différenciation essentiel, garantissant à la fois compétitivité et fiabilité des tarifs, grâce à une segmentation précise et performante. Étant donné que de nombreuses variables se recoupent entre assureurs, une segmentation efficace permet également de lutter contre l'antisélection, phénomène selon lequel l'assurance devient plus avantageuse pour les assurés présentant des risques plus élevés.

Un zonier bien conçu apporte une sophistication supplémentaire à la segmentation, facilitant ainsi la détection des comportements d'antisélection liés à la zone géographique. Toutefois, un compromis doit être établi entre la précision de la segmentation et la mutualisation juste du risque ; il est donc nécessaire de regrouper les communes en zones de risque homogènes pour maintenir l'équilibre.

Calcul de la prime pure par zone

Le zonier a été établi en attribuant une prime pure à chaque commune, calculée à partir de la sinistralité observée et des événements de grêle recensés par Keraunos. Ces primes ont été déterminées en appliquant la théorie de la crédibilité, comme expliqué précédemment. Pour consolider ces résultats à l'échelle des zones de risque, une moyenne pondérée des primes pures par commune a été calculée pour chaque zone, en utilisant le nombre d'assurés comme pondération. La prime pure moyenne par zone se calcule ainsi :

$$\pi_{\text{moyenne de la zone}} = \frac{\sum_{\text{communes}} (\pi_{\text{commune}} \times \text{Nombre d'assurés}_{\text{commune}})}{\sum_{\text{communes}} \text{Nombre d'assurés}_{\text{commune}}}$$

Cette approche permet de refléter avec précision la sinistralité des communes les plus peuplées tout en évitant que les communes à faible effectif n'introduisent un biais. Les résultats obtenus, qui correspondent aux primes annuelles, sont les suivants :

Zones de risque	Prime pure
Bas	8.88€
Moyen	50.17€
Haut	231.12€

Construction de la grille tarifaire

Une fois les primes pures moyennes par zone déterminées, une grille tarifaire est construite. Elle se fonde sur deux variables :

- **Zones de risque** : issues du zonier, représentées en colonne dans la grille.
- **Nombre de pièces de la maison** : facteur clé influençant l'exposition au risque, représenté en lignes dans la grille.

La tarification finale pour chaque assuré est déterminée en combinant la prime pure de la zone de risque et le nombre de pièces du logement. Cette tarification doit respecter le critère d'équilibre, c'est-à-dire que les primes pures doivent couvrir le coût total des sinistres.

Pour ce faire, la prime pure moyenne par zone est fixée comme la prime pure pour une maison de 5 pièces dans la zone de risque correspondante. Ce choix est motivé par le fait que, pour chaque zone, le nombre médian, mais aussi moyen, de pièces par assuré est de 5. Un coefficient α est ensuite appliqué pour chaque variation d'une pièce selon la formule suivante :

$$\pi_j^i = \pi_5^i + \alpha_i(j - 5)$$

Où i est la zone de risque, j le nombre de pièces, π_j^i la prime pure pour une maison de j pièces dans la zone de risque i , et α_i le coefficient d'augmentation ou de diminution par pièce pour la zone de risque i .

Un ajustement linéaire est utilisé pour modéliser l'impact du nombre de pièces, car cette approche permet de capturer de manière simple et intuitive l'augmentation de l'exposition au risque avec la taille du logement. En effet, le risque de sinistralité croît avec la surface habitable et le nombre de pièces, car ces caractéristiques augmentent à la fois la valeur des biens couverts et l'exposition potentielle aux événements climatiques. De plus, une approche linéaire garantit une cohérence et une facilité d'application dans la grille tarifaire, évitant des sauts brutaux ou des incohérences dans les primes entre les différentes configurations de logement.

Ajustement du coefficient α

Des calculs préliminaires avec un α commun aux trois zones ont montré des résultats trop approximatifs et peu représentatifs de la réalité des risques. Chaque zone de risque nécessite donc un α_i spécifique. L'ajustement des α_i se fait en veillant à ce que les primes perçues couvrent exactement le coût des sinistres dans chaque zone. Cela revient à résoudre l'équation suivante pour chaque zone i :

$$\sum_{j=1}^{15} S_j^i = \sum_{j=1}^{15} \pi_j^i A_j^i \quad \forall i \in \{1, 2, 3\}$$

Où S_j^i est la somme des sinistres pour une maison de j pièces dans la zone de risque i , et A_j^i est le nombre d'assurés avec une maison de j pièces dans la zone de risque i .

En substituant π_j^i dans l'équation, on obtient :

$$\begin{aligned} \sum_{j=1}^{15} S_j^i &= \sum_{j=1}^{15} (\pi_5^i A_j^i) + \sum_{j=1}^{15} (\alpha_i (j-5) A_j^i) \quad \forall i \in \{1, 2, 3\} \\ \Leftrightarrow \sum_{j=1}^{15} S_j^i - \sum_{j=1}^{15} \pi_5^i A_j^i &= \alpha_i \sum_{j=1}^{15} (j-5) A_j^i \quad \forall i \in \{1, 2, 3\} \\ \Leftrightarrow \alpha_i &= \frac{\sum_{j=1}^{15} S_j^i - \sum_{j=1}^{15} \pi_5^i A_j^i}{\sum_{j=1}^{15} (j-5) A_j^i} \quad \forall i \in \{1, 2, 3\} \end{aligned}$$

Cette formule explicite permet de calculer les coefficients α_i pour chaque zone de risque, en fonction des valeurs observées de S_j^i , π_5^i et A_j^i . Les coefficients obtenus sont les suivants :

Zones de risque	α_i
Bas	1.15€
Moyen	5.60€
Haut	22.43€

Ces coefficients permettent de construire la grille tarifaire suivante, où toutes les primes correspondent à des montants annuels :

Zone de risque Nombre de pièces	Bas	Moyen	Haut
1	4,28€	27,77€	141,40€
2	5,43€	33,37€	163,83€
3	6,58€	38,97€	186,26€
4	7,73€	44,57€	208,69€
5	8,88€	50,17€	231,12€
6	10,03€	55,77€	253,55€
7	11,18€	61,37€	275,98€
8	12,33€	66,97€	298,41€
9	13,48€	72,57€	320,84€
10	14,63€	78,17€	343,27€
11	15,78€	83,77€	365,70€
12	16,93€	89,37€	388,13€
13	18,08€	94,97€	410,56€
14	19,23€	100,57€	432,99€
15	20,38€	106,17€	455,42€

Cette méthodologie garantit une tarification cohérente et adaptée, prenant en compte à la fois le risque climatique local et les caractéristiques spécifiques de chaque logement.

Interprétation des résultats

Les coefficients obtenus montrent une augmentation significative des primes en fonction de la zone de risque, reflétant la nécessité d'une tarification adaptée à la réalité des sinistres. Par exemple, pour une maison de **10 pièces**, le tarif est cinq fois supérieur dans une zone de risque moyenne par rapport à une zone basse, et quatre fois plus élevé dans une zone haute par rapport à une zone moyenne. Ces écarts importants s'expliquent principalement par la concentration des communes en zone rouge, qui regroupent celles où le produit du coût moyen des sinistres et de leur fréquence est le plus élevé. Cette caractéristique reflète non seulement la gravité des sinistres, entraînant des dommages plus coûteux, mais aussi la fréquence plus importante des événements causant des dégâts dans ces zones.

Par ailleurs, ces écarts sont également liés à la répartition des communes : seules 15% d'entre elles se situent en zone rouge. Si ce pourcentage était plus élevé, les primes moyennes dans ces zones seraient naturellement réduites en raison de la mutualisation plus importante des risques.

Bien que cette tarification soit cohérente actuariellement, elle crée une différence importante entre les zones verte et rouge, pouvant être perçue comme injuste pour les assurés des zones exposées et limiter l'incitation à souscrire. Un tarif équilibré doit rester compétitif pour garantir l'adhésion des assurés. Pour évaluer cette compétitivité, un benchmark est réalisé.

Benchmark

Ce benchmark est réalisé sur la même région à l'aide d'un comparateur d'assurances et compare les tarifs à ceux des contrats d'assurance MRH globaux incluant le risque grêle, sans pour autant isoler spécifiquement la part attribuée à cette garantie. Les critères de comparaison ont été définis en supposant une moyenne de 15 m² par pièce et en testant diverses communes situées dans des zones de risque variées.

Les résultats montrent une absence notable de différenciation tarifaire entre les communes des zones verte et rouge du zonier. Au global, les tarifs des zones à risque faible se révèlent compétitifs. Dans les zones à risque moyen, bien qu'un peu plus élevés, ils demeurent acceptables. En revanche, dans les zones à risque élevé, malgré une justification actuarielle solide, les niveaux tarifaires atteints pourraient s'avérer difficilement commercialisables. Ces constats pourraient s'inscrire dans une logique d'équilibre global du portefeuille, visant à répartir équitablement les charges entre les différents segments d'assurés. En appliquant ces tarifs, l'assureur peut garantir une politique tarifaire à la fois actuariellement rigoureuse et alignée sur ses objectifs commerciaux et stratégiques.

Il est néanmoins essentiel de rappeler que la comparaison est réalisée au niveau des ordres de grandeur, car les tarifs recensés dans l'étude correspondent à l'intégralité des contrats MRH et non spécifiquement la garantie grêle. En pratique, la tarification d'une sous-garantie comme la grêle est intégrée dans le tarif global du produit et est ajustée en fonction de son poids relatif dans le contrat global, ainsi que des corrélations existantes avec les autres garanties incluses.

En effet, un tarif MRH tient compte de multiples expositions, telles que les cambriolages, les tempêtes ou la sécheresse. Chaque type de risque est pondéré selon sa probabilité de survenance et son impact potentiel, influençant ainsi la prime globale. Par exemple, dans une zone où le risque de grêle est faible mais où l'exposition au cambriolage est élevée, le tarif global serait ajusté pour refléter ce déséquilibre tout en restant compétitif sur le marché. De manière similaire, une légère réévaluation des tarifs en zone à faible risque de grêle pourrait permettre de maintenir une compétitivité générale tout en absorbant les ajustements nécessaires dans les zones à risque élevé.

A noter que les moyennes relevées dans les zones étudiées sont plus élevées que celles observées à l'échelle nationale, puisque les simulations se concentrent sur des zones géographiques particulièrement exposées.

3.2.3 Définition d'une politique de souscription

Dans la continuité des résultats de la grille tarifaire et du benchmark réalisé, l'introduction du zonier pour le risque grêle permet de définir une politique de souscription. En effet, cet outil a permis d'affiner la stratégie tarifaire et de mieux cibler les zones de risque.

Avant l'implémentation du zonier, les zones à risque étaient simplement celles qui coûtaient le plus cher à l'assureur. Grâce à une méthodologie mathématique robuste, une délimitation précise des zones de risque est possible, prenant en compte à la fois les événements de grêle et les spécificités du portefeuille.

Aussi, l'outil permet d'améliorer la tarification, désormais définie par zone de risque. Cette approche est plus équitable et mieux adaptée aux risques réels identifiés, contrairement à l'ancien système qui se base sur un tarif global avec des majorations dans les zones coûteuses, qui peut être perçu comme injuste et approximatif.

Commercialement, avant l'introduction du zonier, une zone entière pouvait être pénalisée, même si certains endroits ont un risque modéré. Désormais, les actions commerciales peuvent être ciblées, incitant par exemple à développer les zones à faible risque. Cela permet un déploiement plus stratégique des efforts commerciaux.

De plus, l'implémentation du zonier apporte des éléments de réflexion autour de l'introduction des mesures préventives, comme la sensibilisation et l'accompagnement des assurés dans les zones à risque élevé, complétant les ajustements tarifaires et renforçant l'efficacité de la politique de souscription.

En répondant aux objectifs fixés, le zonier permet de :

- **D'appréhender les expositions du portefeuille :** Identification précise des zones à risque pour une gestion proactive.
- **Contrôler les coûts :** Ajustement des tarifs par zone, permettant une répartition plus équitable des charges.
- **Cibler des zones spécifiques pour équilibrer le portefeuille et développer l'activité commerciale.**

Malgré les limites que le zonier représente, il est une avancée en permettant une tarification plus ciblée et adaptée, et constitue aussi une première étape vers une politique de souscription plus précise et équitable.

Cependant, bien que cette vision soit efficace à court terme, elle doit être complétée par une perspective à plus long terme, car deux problématiques majeures émergent :

- **Les évolutions climatiques :** Les phénomènes météorologiques extrêmes tendent à augmenter, rendant nécessaire une réévaluation régulière des zones de risque et des tarifs.
- **Les injustices sociales :** La segmentation tarifaire basée sur le zonier peut renforcer des inégalités d'accès à l'assurance pour certains assurés.

Ces considérations mettent en lumière la nécessité d'une approche évolutive et proactive. La prochaine partie développera ces axes pour proposer des solutions durables et inclusives.

3.3 Scénario de sensibilité : dérèglement climatique et changements réglementaires

Le contexte climatique et réglementaire en France connaît une évolution rapide, nécessitant une adaptation des pratiques assurantielles. La grêle, phénomène de plus en plus fréquent et intense sous l'effet du changement climatique, impose aux assureurs de repenser leur approche tarifaire et de souscription pour couvrir ce risque grandissant. Ce scénario de sensibilité explore deux aspects essentiels : d'une part, l'impact des changements climatiques sur la fréquence et l'intensité des sinistres liés à la grêle et, d'autre part, les modifications réglementaires visant à mieux encadrer l'assurance de ces risques climatiques. Ensemble, ces éléments dessinent un futur où la gestion des risques et l'équité de l'accès aux assurances doivent s'adapter aux réalités d'une société confrontée à des aléas plus fréquents et intenses.

3.3.1 Politique de souscription avec prise en compte du changement climatique

Dans un contexte de dérèglement climatique marqué, l'assurance doit s'adapter pour mieux appréhender les risques croissants liés à la grêle et proposer des solutions équilibrées face à ces nouveaux défis.

Le rapport du groupe 1 du sixième rapport d'évaluation du GIEC (2021) indique qu'en dépit de conditions atmosphériques globales de plus en plus propices à la grêle en Europe, le niveau de confiance dans l'attribution directe de cette augmentation au changement climatique reste faible. Ce faible niveau de confiance est attribué au manque de données d'observation, à la résolution encore insuffisante des modèles climatiques et à la représentation paramétrique de certains processus physiques dans ces modèles.

Cependant, des recherches plus récentes depuis la publication du rapport du GIEC renforcent la fiabilité de cette attribution [23]. Elles associent ainsi l'augmentation des épisodes de grêle entre 1950 et 2021 à une humidité accrue dans les basses couches de l'atmosphère lors des orages. En France, cette tendance se manifeste par une augmentation statistiquement significative des épisodes de grêle avec des grêlons de plus de 5 cm sur une grande partie du territoire métropolitain [19].

Augmentation de la fréquence et de l'intensité des tempêtes de grêle

Le changement climatique intensifie les événements climatiques extrêmes, influençant non seulement leur fréquence, mais aussi leur répartition géographique. La grêle, en particulier, a vu sa fréquence et son intensité augmenter, nécessitant une refonte des stratégies de souscription et de tarification des assurances pour s'adapter à ces nouvelles réalités.

Une étude menée par un groupe de recherche en météorologie et climatologie de l'Université de l'Illinois du Nord [24] projette qu'avec les scénarios actuels de réchauffement climatique, les tempêtes de grêle produiront des grêlons plus gros d'ici au milieu ou à la fin du siècle. Ce phénomène s'explique par une instabilité atmosphérique croissante qui génère des courants ascendants plus puissants au sein des orages. Selon cette étude, la taille des plus gros grêlons pourrait croître de 15% à 75%, en fonction de l'intensité des émissions de gaz à effet de serre.

En parallèle, des projections climatiques et hydro-climatiques réalisées par le groupe Covéa à l'horizon 2050 dans le cadre du scénario RCP8.5 anticipent une hausse de 40% des orages de grêle sur le territoire français, accompagnée d'une augmentation de 20% de la charge de sinistres (hors inflation) [20]. Ce phénomène illustre une dérive structurelle, un changement profond et durable difficile à inverser, nécessitant une adaptation en profondeur des stratégies assurantielles face aux risques climatiques.

Ajustement de la grille tarifaire en tenant compte de ces évolutions

La révision des grilles tarifaires s'impose pour refléter ces évolutions, mais le risque de modèle persiste pour le risque grêle, notamment en raison de l'incertitude associée aux périodes de retour de certains événements extrêmes. Par exemple, l'épisode exceptionnel de grêle de juin 2022 voit sa fréquence de survenance estimée, selon les acteurs, entre 10 et 200 ans, certains modèles de marché élargissant même cette plage jusqu'à 5 000 ans, révélant une forte incertitude [19].

Dans cette analyse, le scénario le plus prudent a été retenu, en se basant sur une période de retour de 10 ans pour l'épisode exceptionnel de 2022. Avec un modèle de données couvrant un historique de 5 ans, incluant 2022, une période d'observation supplémentaire de 5 ans est prévue avant d'intégrer une augmentation progressive des primes, visant à atteindre une hausse totale de 20% d'ici à 2050. Cette évolution tient compte de la hausse anticipée de la charge de sinistres liée à la grêle, projetée pour 2050 par les modélisations de Covéa.

Pour calculer cette augmentation progressive de 20% sur l'horizon 2030-2050, un taux de croissance annuel constant est appliqué. La croissance composée suit la formule suivante :

$$V_{20} = V_0 * (1 + r)^{20}$$

En fixant $V_{20} = V_0 * 1,20$, le taux de croissance annuel r est déterminé par l'équation :

$$1,20 = (1 + r)^{20}$$

$$\Leftrightarrow r = (1,20)^{\frac{1}{20}} - 1 \approx 0,92\%$$

Ainsi, pour atteindre une augmentation totale de 20% sur 20 ans, une hausse annuelle de 0,92% sera appliquée à la prime pure grêle, calculée chaque année sur la base de la prime de l'année précédente.

Ajustement du zonier face au changement climatique

Comme mentionné précédemment, le changement climatique influe également sur la répartition géographique des événements extrêmes. Il est donc essentiel de surveiller régulièrement le zonier afin d'intégrer les nouveaux épisodes de grêle ainsi que les sinistralités émergentes et d'actualiser continuellement les zones à risque.

Dans cette optique, Covéa a modélisé l'évolution de la fréquence spatiale des orages de grêle entre la période historique et les projections pour 2050, comme présenté dans "Changement climatique & assurance : Quelles conséquences sur la sinistralité à horizon 2050 ?" [20]. Les résultats de cette analyse sont illustrés dans les cartes ci-dessous :

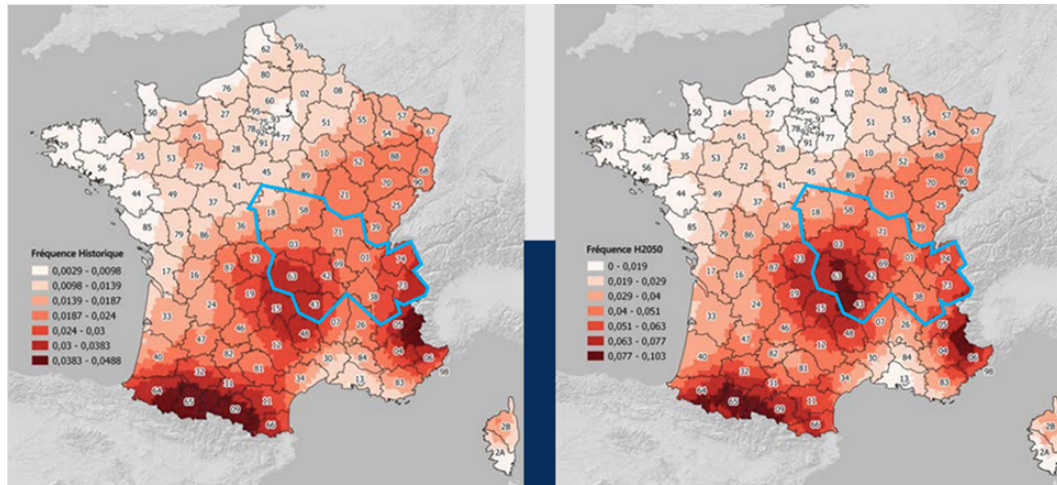


FIGURE 50 – Évolution de la fréquence spatiale des orages de grêle (en nombre d'occurrence par an) sur la période historique et à horizon 2050 [20].

Ces cartes montrent qu'il ne faut pas s'attendre à des changements spatiaux majeurs dans les régions les plus exposées, bien qu'un déplacement vers le nord-est des zones de forte exposition soit perceptible. Par ailleurs toutes les régions devraient connaître une augmentation notable de la fréquence des orages de grêle. En observant la région d'étude, l'Auvergne-Rhône-Alpes qui est entourée en bleu sur les deux cartes, on note un léger glissement vers l'est des départements les plus exposés, avec un risque accru pour les départements 1, 3, 18, 42, 43, 58, 63, 69, et 71, tandis que les départements 38, 73, et 74 pourraient être relativement moins exposés par rapport aux autres qu'historiquement.

Ainsi, bien que la carte des zones les plus exposées demeure relativement stable, ces observations indiquent que des ajustements du zonier seront indispensables pour l'adapter aux réalités climatiques, à l'évolution du portefeuille d'assurés et à l'augmentation des risques dans les nouvelles zones d'exposition.

En conclusion, avec la grille tarifaire proposée dans la section précédente et l'augmentation des coûts liés au dérèglement climatique, un désengagement dans les zones les plus à risque est à craindre. Ainsi, des changements réglementaires susceptibles de compenser ce déséquilibre en incitant les assureurs à ne pas se désengager des zones risquées seront examinés dans la sous-partie suivante. Cette étude a également permis, d'un point de vue politique de souscription, d'anticiper les évolutions tarifaires climatiques plutôt que de les subir. En prenant en compte les projections climatiques, elle offre une vision long-termiste sur des aspects encore peu étudiés aujourd'hui, permettant ainsi aux assureurs de s'adapter de manière proactive aux risques croissants.

3.3.2 Politique de souscription avec prise en compte de changements réglementaires

Face à l'intensification des risques climatiques et à la problématique de l'inassurabilité, le rapport de mission Langrenay, intitulé "Adapter le système assurantiel français face à l'évolution des risques climatiques" [19], a été remis aux ministres Bruno Le Maire et Christophe Béchu le 2 avril 2024. Ce rapport, déjà mentionné dans la première partie, émet des recommandations visant à adapter le système assurantiel français aux aléas croissants, soulignant ainsi des enjeux sociétaux majeurs. Les propositions formulées visent à renforcer la résilience du marché de l'assurance, à mieux protéger les assurés et à garantir un accès équitable aux couvertures.

Certaines recommandations, particulièrement pertinentes pour l'assurance des dommages causés par la grêle, pourraient transformer les pratiques de souscription, de tarification et de compensation dans ce domaine. Un accent est mis sur la limitation de la discrimination géographique, un enjeu central pour créer un modèle de souscription plus juste et durable. L'intégration de ces mesures vise à diminuer les effets de sélection géographique associés à l'utilisation d'un zonier, tout en promouvant une approche assurantielle inclusive et responsable. Par ailleurs, la grille tarifaire proposée initialement présentait des limites en termes d'équité pour les zones les plus à risque, risquant de décourager les populations concernées à souscrire une assurance. Il apparaît donc essentiel d'ajuster ces mécanismes afin de limiter la discrimination géographique et de garantir une accessibilité équitable à l'assurance.

Discrimination géographique

Le rapport met en avant des mesures pour contrer ce phénomène de retrait de certains assureurs des zones à forte exposition, qui se manifeste par des pratiques tarifaires dissuasives et des politiques d'exclusion. Il recommande ainsi plusieurs actions pour limiter cette discrimination géographique et encourager les assurés à adopter des mesures de prévention :

- **Cartographie harmonisée des zones exposées aux risques climatiques** : Une cartographie nationale des zones à risques, harmonisée et partagée entre l'État, la Caisse Centrale de Réassurance (CCR), et les acteurs de l'assurance, serait créée. Basée sur les Plans de Prévention des Risques Naturels (PPRN), cette cartographie faciliterait l'identification des zones à risques, assurerait une présence continue des assureurs dans ces zones, et permettrait d'implémenter des mesures de prévention ciblées en fonction de l'exposition des assurés.
- **Mise en place d'un bonus-malus fiscal entre zones** : ce système permettrait d'accorder des bonus fiscaux aux assureurs fortement présents dans les zones exposées aux risques climatiques, tout en imposant des malus fiscaux à ceux qui désertent ces zones. Cette mesure inciterait à équilibrer les primes proposées, en augmentant celles des assurés dans les zones vertes tout en réduisant celles des assurés dans les zones rouges.
- **Renforcement des mécanismes de nudging** : il s'agirait d'un système de bonus-malus sur la prime d'assurance pour les assurés prenant des mesures de prévention contre les risques climatiques. Cela renforcerait indirectement le dispositif en réduisant l'exposition de certains cas à des risques majeurs. Par exemple, dans le cadre de l'assurance contre le risque de grêle, un assuré disposant de volets anti-grêle et d'une toiture résistante aux intempéries pourrait bénéficier d'une réduction significative de sa prime, notamment en zone rouge.
- **Mécanisme de modulation des surprimes Cat Nat entre zones** : Le rapport suggère également de moduler la contribution additionnelle sur les surprimes Cat Nat pour réduire les écarts de coût entre zones à faible et forte exposition. Plutôt qu'un taux de prélèvement unique de 12%, le rapport propose des **taux modulés**, tout en maintenant un prélèvement moyen de 12% pour l'ensemble du territoire. Par exemple :
 - 0% voire taux négatif pour les zones rouges à forte exposition,
 - 5% pour les zones orange à risque modéré,
 - Jusqu'à 30% voire 50% pour les zones vertes faiblement exposées.

Dans le cadre des travaux de ce mémoire, le dernier point présenté est étudié et appliqué à la garantie grêle. Bien que la grêle ne fasse pas partie du régime Cat Nat, cette situation peut soulever des interrogations. En effet, comme montré précédemment, la grêle peut représenter une part significative de la sinistralité liée aux phénomènes naturels. Ainsi, une évolution du cadre réglementaire pourrait être envisagée, avec l'hypothèse d'intégrer la grêle au régime Cat Nat pour mieux adapter le système assurantiel à l'intensification de ce risque climatique.

Par ailleurs, cette étude est pertinente car elle illustre comment les recommandations du rapport Langrenay peuvent être mises en pratique, notamment pour limiter la discrimination géographique dans la tarification. Elle s'inscrit également dans un contexte marqué par l'augmentation de la cotisation au régime Cat Nat, passée de 12% à 20% au 1er janvier 2025 [10], témoignant des défis financiers croissants pour l'assurance en France. En réponse, et en s'appuyant sur les spécificités du portefeuille client ainsi que sur le zonier élaboré, les ajustements suivants sont proposés :

- 2% pour les zones rouges
- 10% pour les zones orange
- Environ 100% pour les zones vertes faiblement exposées.

Ces taux ont été choisis après plusieurs tests pour garantir une cohérence avec l'étude et répondre aux critères suivants :

- Maintien d'un prélèvement moyen de 20% sur l'ensemble du territoire,
- Calcul basé sur les primes pures médianes (pour un logement de 5 pièces)
- Répartition équilibrée des assurés : 50% dans les communes vertes, 35% dans les communes orange et 15% dans les communes rouges.

Comparatif de la tarification avec surprime Cat Nat modulée

Le tableau suivant illustre l'impact de cette surprime modulée sur la prime pure grêle, en comparaison avec une surprime uniforme de 20% appliquée à toutes les zones. La grille tarifaire prend en compte le nombre de pièces et la zone de risque :

Zones de risque / Nombre de pièces	Bas		Moyen		Haut	
	Prime pure grêle avec surprime Cat Nat uniforme de 20%	Prime pure grêle avec surprime Cat Nat modulée	Prime pure grêle avec surprime Cat Nat uniforme de 20%	Prime pure grêle avec surprime Cat Nat modulée	Prime pure grêle avec surprime Cat Nat uniforme de 20%	Prime pure grêle avec surprime Cat Nat modulée
1	5,14 €	8,56 €	33,32 €	30,55 €	169,68 €	144,23 €
2	6,52 €	10,86 €	40,04 €	36,71 €	196,60 €	167,11 €
3	7,90 €	13,16 €	46,76 €	42,87 €	223,51 €	189,99 €
4	9,28 €	15,46 €	53,48 €	49,03 €	250,43 €	212,86 €
5	10,66 €	17,76 €	60,20 €	55,19 €	277,34 €	235,74 €
6	12,04 €	20,06 €	66,92 €	61,35 €	304,26 €	258,62 €
7	13,42 €	22,36 €	73,64 €	67,51 €	331,18 €	281,50 €
8	14,80 €	24,66 €	80,36 €	73,67 €	358,09 €	304,38 €
9	16,18 €	26,96 €	87,08 €	79,83 €	385,01 €	327,26 €
10	17,56 €	29,26 €	93,80 €	85,99 €	411,92 €	350,14 €
11	18,94 €	31,56 €	100,52 €	92,15 €	438,84 €	373,01 €
12	20,32 €	33,86 €	107,24 €	98,31 €	465,76 €	395,89 €
13	21,70 €	36,16 €	113,96 €	104,47 €	492,67 €	418,77 €
14	23,08 €	38,46 €	120,68 €	110,63 €	519,59 €	441,65 €
15	24,46 €	40,76 €	127,40 €	116,79 €	546,50 €	464,53 €

FIGURE 51 – Grille tarifaire avec surprime Cat Nat uniforme de 20% ou modulée

Ce tableau montre donc une augmentation de 40% entre la prime Cat Nat uniforme et celle modulée pour la zone verte, une diminution de 9% pour la zone orange, et enfin une diminution de 18% dans la zone rouge.

Ces simulations montrent que la modulation des surprimes Cat Nat permet de maintenir une couverture viable dans les zones à risque tout en limitant l'impact financier pour les assurés en zones fortement exposées. Les ajustements proposés contribuent ainsi à un système assurantiel plus équitable et adapté aux nouvelles contraintes climatiques. Cependant, cette mesure reste insuffisante, car on observe encore un écart significatif, de l'ordre de 1 à 15, entre les zones vertes et les zones rouges. C'est pourquoi d'autres recommandations issues du rapport Langreny doivent être prises en compte, ainsi que des choix stratégiques des assureurs, tels que d'éventuels rééquilibrages avec d'autres types d'assurances.

En intégrant les recommandations du rapport Langreny, cette étude propose une approche évolutive qui permet d'adapter le système assurantiel aux défis climatiques. Le processus de modulation des surprimes Cat Nat et les ajustements proposés pour la tarification de la grêle illustrent l'importance de l'adaptation de la politique de souscription dans un contexte de dérèglement climatique. En révisant la grille tarifaire et en introduisant des mécanismes de prévention, cette étude montre que l'on peut assurer l'accessibilité de l'assurance tout en maintenant une couverture adéquate dans les zones à risque élevé. Ces ajustements offrent ainsi une solution pour réduire la discrimination géographique et améliorer la justice sociale au sein du marché de l'assurance.

Limites des hypothèses retenues

Cette étude repose sur plusieurs hypothèses fondamentales qui nécessitent une analyse critique et prudente. Tout d'abord, les projections climatiques utilisées, bien qu'appuyées sur des scénarios scientifiques fiables tels que le RCP8.5, comportent des incertitudes inhérentes aux modèles climatiques, en particulier en ce qui concerne l'évolution à long terme de la fréquence et de l'intensité des épisodes de grêle. Ensuite, la modulation des surprimes Cat Nat et l'ajustement des grilles tarifaires reposent sur des répartitions géographiques et des paramètres économiques actuels, qui pourraient être modifiés par des évolutions imprévues, notamment si l'on considère l'élaboration d'un zonier multirisques climatiques à une échelle nationale.

Enfin, les recommandations issues du rapport Langrenoy, bien qu'elles offrent des pistes prometteuses, n'ont pas encore été mises en œuvre, ce qui pourrait limiter l'impact des ajustements tarifaires proposés dans cette étude. Ces incertitudes soulignent la nécessité d'une approche adaptable et évolutive pour répondre aux défis posés par les changements climatiques et leurs répercussions sur le secteur assurantiel.

Conclusion

Les événements de grêle historiques de 2022 ont mis en lumière l'urgence de mieux comprendre, anticiper et gérer ces phénomènes, dont l'intensité croissante est accentuée par le dérèglement climatique. Ce mémoire s'inscrit dans cette problématique en proposant un zonier du risque grêle pour l'assurance habitation, spécifiquement appliqué à la région Auvergne-Rhône-Alpes. Cet outil vise à répondre à deux ambitions fondamentales : optimiser l'anticipation du risque en offrant une tarification adaptée et favoriser des politiques assurantielles proactives en matière de prévention. Cette étude innove en intégrant la théorie de la crédibilité dans un cadre localisé, encore peu exploré dans ce domaine, et apporte ainsi une contribution originale à la gestion des risques climatiques croissants.

La méthodologie repose sur la modélisation des deux composantes clés de la tarification du risque assurantiel : la fréquence des sinistres et leur coût moyen. Trois approches ont été explorées pour construire ce zonier. La méthode la plus convaincante, et celle qui a été retenue, est celle utilisant la théorie de la crédibilité, à travers le modèle de Bühlmann-Straub. En effet, cette approche permet une adaptation locale tout en tirant parti des données globales, optimisant ainsi la segmentation des risques et l'estimation des primes pures. Ce zonier constitue également une base stratégique pour identifier les zones vulnérables, orienter les actions de prévention et renforcer les infrastructures face aux risques climatiques croissants.

Cependant, le travail réalisé comporte certaines limites et axes d'améliorations liés à la complexité du risque grêle, dont la prévision dépend de nombreux facteurs météorologiques et géographiques. Ces principales limites et axes d'améliorations sont :

- **Limites des données disponibles :**

- **Profondeur historique faible :** Les données utilisées couvrent seulement cinq années (2019-2023), ce qui limite l'analyse de phénomènes rares ou extrêmes. L'élargissement de l'horizon temporel améliorerait la robustesse des projections.
- **Dépendance à une seule base de données assurantielle :** Les résultats se basent sur la base de données de sinistralité d'un seul assureur. Intégrer d'autres bases de données diversifierait et renforcerait les analyses.
- **Fiabilité des données externes :** La base météorologique Keraunos, bien que précieuse, présente des biais liés à la densité de population, influençant le signalement des événements. En effet, comme elle est en partie complétée par les observations de la population, des communes densément peuplées seront plus susceptibles d'avoir des chutes de grêles recensées que des communes avec une densité de population plus faible. La confrontation à d'autres bases de données météorologiques pourrait améliorer la fiabilité.
- **Manque de données détaillées des biens assurés :** Des informations détaillées, telles que la surface des habitations, le type et l'inclinaison des toits, enrichiraient significativement les modèles.

- **Potentielle perte d'information liée aux retraitements :**

- Le choix de ne conserver que les sinistres TGN en lien avec un événement de grêle Keraunos a pu entraîner des erreurs d'attribution (exclusion de sinistres grêle ou inclusion de sinistres tempête). En effet, ce retraitement peut ajouter des sinistres qui sont en réalité de la tempête (tempête et grêle dans la même période de 5 jours), ou à l'inverse, des sinistres liés à la grêle peuvent être supprimés si Keraunos n'a pas recensé d'événements de grêle ce jour-là dans cette commune.

- **Hypothèses méthodologiques restrictives :**

- **Répartition uniforme des assurés au sein des communes :** La connaissance de la localisation GPS des assurés permettrait d'améliorer le modèle.
- **Sélection de la taille maximale des grêlons pour évaluer l'impact :** La connaissance de la localisation GPS des chutes de grêle et leur taille, améliorerait les résultats.

• **Limites des scénarios de sensibilités :**

- Les projections climatiques sont celles proposées par Covéa à horizon 2050, mais cette charge peut être bien différente, dans un sens comme dans l'autre, dans les faits.
- Les projections réglementaires proviennent pour la plupart du rapport Langrenay. Cependant il ne s'agit ici que d'hypothèses d'évolutions réglementaires. De plus, concernant la proposition d'évolution de la surprime CatNat, les calculs ont été effectués en s'appuyant sur notre classification actuelle du risque (50% des communes en risque faible, 35% en risque modéré et 15% en risque élevé). Toutefois, il est probable que cette classification soit différente dans le cadre de cette mesure.

Au-delà de son rôle technique, le zonier ambitionne d'aider la politique de souscription des assureurs en donnant un levier pour une gestion plus juste, transparente et socialement responsable des risques. Il permet une segmentation précise des zones de risques, garantissant une équité tarifaire pour les assurés et alignant les primes sur les expositions réelles. Cette démarche contribue à rétablir un équilibre financier pour l'assureur tout en renforçant la confiance des assurés dans un contexte de volatilité croissante des aléas climatiques. Le zonier met en lumière le rôle central de l'assureur en tant qu'acteur social et économique, non seulement dans la gestion des sinistres, mais aussi dans la prévention des risques.

En identifiant les zones à forte vulnérabilité, le zonier ouvre la voie à des actions ciblées telles que des incitations à l'installation de dispositifs de protection (toitures renforcées, stores résistants) accompagnées d'incitations financières (réductions de primes pour les assurés adoptant des mesures de prévention). La mise en place de « contrats durables », intégrant un cofinancement des équipements de prévention, pourrait également renforcer la résilience des assurés. Il répond ainsi à une double mission : protéger les assurés tout en contribuant à la résilience des territoires face aux aléas climatiques.

Cependant, la gestion des risques climatiques ne peut reposer uniquement sur les assureurs. Ainsi, en favorisant une tarification différenciée et adaptée, le zonier soutient une politique d'engagement tripartite, impliquant l'assureur, les pouvoirs publics et les citoyens afin de renforcer les infrastructures et améliorer la résilience des territoires.

En conclusion, ce mémoire met en avant la nécessité d'une gestion proactive et collective des risques climatiques. La méthodologie employée et les résultats obtenus soulignent la valeur ajoutée d'une approche combinant segmentation fine des risques et exploitation de la théorie de la crédibilité. Si des améliorations sont nécessaires, notamment en termes de données et d'hypothèses, ce travail constitue une base solide pour des stratégies assurantielles novatrices. Il met également en lumière un enjeu clé : le rôle des assureurs évolue vers une responsabilisation accrue des assurés, tout en appelant à une action collective face aux défis climatiques. Face à l'évolution rapide des aléas naturels, seule une coopération entre assureurs, pouvoirs publics et citoyens pourra garantir une gestion efficace et équitable des risques à venir.

Références

- [1] Anelfa : <https://www.anelfa.asso.fr/L-Echelle.html>
- [2] Anelfa : <https://www.anelfa.asso.fr/Climatologie-de-la-grele.html>
- [3] Encyclopédie de l'environnement : <https://www.encyclopedia-environnement.org/zoom/formation-de-grele/>
- [4] Europe 1 : <https://www.europe1.fr/societe/Comment-se-forment-les-grelons-662458>
- [5] Keraunos : <https://www.keraunos.org/climatologie/les-orages-en-france/grele-climatologie-france-records-saisonnalite-risque-degats>
- [6] Météo Paris : <https://www.meteo-paris.com/actualites/la-grele-en-france-et-dans-le-monde>
- [7] MRN: Tempête de grêle ELA, 6 au 10 juin 2014
- [8] Wikipédia: Grêle
- [9] Présentation du seuil de vigilance MRN grêle, MRN
- [10] Publication des arrêtés renforçant les moyens d'action du régime d'indemnisation des catastrophes naturelles et du fonds de garantie des victimes, Ministère de l'économie, des finances et de la souveraineté industrielle et numérique
- [11] Mémoire d'actuariat de Candice Poujade, "Application de la théorie de la crédibilité dans le contexte de l'assurance des arrêts de travail", 2019-2020, <https://www.institutdesactuaire.com/docs/mem/d22346b7c7ce91f74b2ed46295acfed3.pdf>
- [12] Cours actuariat de l'assurance non-vie, ENSAE, Nicolas Baradel, https://www.nicolasbaradel.fr/enseignement/ressources/cours_assurance_dommage.pdf
- [13] Cours GLM, INSA Toulouse, Juliette Chevallier, <https://juliette-chevallier.pages.math.cnrs.fr/>
- [14] Cours "Théorie de la crédibilité", Pierre Théron, ISFA, 2017-2018, <https://www.therond.fr/wp-content/uploads/cours/Credibilite.pdf>
- [15] Cours "Théorie de la crédibilité", Olivier Wintenberger, slides Olivier Lopez, <https://www.wintenberger.fr/cours/CompProba/Credibilite.pdf>
- [16] Cours "Principes de base de l'assurance dommage", Adrien Suru, Dauphine, 2017-2018
- [17] L'assurance des événements naturels en 2022, Fédération française de l'Assurance
- [18] Conséquences du changement climatique sur le coût des catastrophes naturelles en France à l'horizon 2050, CCR
- [19] Thierry Langrenoy, Gonéri Le Cozannet, Myriam Merad, « Adapter le système assurantiel français face à l'évolution des risques climatique », avril 2024. <https://www.vie-publique.fr/rapport/293621-adapter-le-systeme-assurantiel-face-levolution-des-risques-climatiques>
- [20] Livre blanc Covéa : « Changement climatique et assurance : Quelles conséquences sur la sinistralité à horizon 2050 ? », Janvier 2022, 202202_Livre_Blanc_Covéa_Risques_Climatiques.pdf
- [21] Lettre d'information de la Mission Risques Naturels: "Aléa grêle: point sur les connaissances", https://www.mrn.asso.fr/wp-content/uploads/2020/07/lettre-n34_vf.pdf
- [22] Scientists say large hail will become more common in warming climate, 2024, Northern Illinois University, <https://newsroom.niu.edu/scientists-say-large-hail-will-become-more-common-in-warming-climate/>
- [23] Forecasting large hail and lightning using additive logistic regression models and the ECMWF forecasts, Francesco Battaglioli et al, 2023, nhess-23-3651-2023.pdf
- [24] Gensini, V.A., Ashley, W.S., Michaelis, A.C. et al. Hailstone size dichotomy in a warming climate. npj Clim Atmos Sci 7, 185 (2024). <https://doi.org/10.1038/s41612-024-00728-9>
- [25] Cook, R. Dennis. "Detection of Influential Observation in Linear Regression". Technometrics, vol. 19, no 1, février 1977, p. 15 18. DOI.org (Crossref), <https://doi.org/10.1080/00401706.1977.10489493>.
- [26] Berthet, Claude, Jean Dessens et José Luis Sánchez (2011). "Regional and yearly variations of hail frequency and intensity in France". In : Atmospheric Research 100.4, p. 391-400.
- [27] Dessens, J, C Berthet et JL Sanchez (2015). "Change in hailstone size distributions with an increase in the melting level height". In : Atmospheric Research 158, p. 245-253.
- [28] Bühlmann, H. and Gisler, A., (2005), A Course in Credibility Theory and its Applications.