



**Mémoire présenté pour l'obtention du Diplôme Universitaire  
d'actuariat de l'ISFA  
et l'admission à l'Institut des Actuaires**

Par : Lucas Blanc

Titre : Maîtrise du risque Tempête-Grêle au sein de Groupama Rhône-Alpes Auvergne

Confidentialité : ☒ Non    ☐ Oui    (Durée : ☐ 1 an    ☐ 2 ans)

---

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

*Membres présents du jury de l'Institut  
des Actuaires :*

*Entreprise :*

Nom : Groupama Rhône Alpes Auvergne

Signature :

*Membres présents du Jury du Master  
Actuariat de l'ISFA :*

*Directeur de Mémoire en entreprise :*

Nom : Rémi GREGOIRE

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

---

Les risques naturels deviennent ces dernières années un sujet important au vue de l'augmentation de leur fréquence d'apparition et de leur sévérité. En France, où le climat est devenu de plus en plus imprévisible, les événements climatiques naturels représentent une menace croissante pour les compagnies d'assurance. Parmi les événements climatiques les plus coûteux, le risque de tempête occupe une place importante, se classant juste après les inondations et les sécheresses en termes de charge sinistre. Pour les compagnies d'assurance, la gestion de ce risque est essentielle pour préserver leur stabilité financière et garantir leur capacité à indemniser les assurés en cas de sinistre majeur.

Ce mémoire s'intéresse à la manière dont le risque de tempête est traité au sein de Groupama Rhône-Alpes Auvergne (GRAA), l'une des caisses régionales du groupe Groupama, un acteur clé de l'assurance en France. La gestion de ce risque repose sur des méthodes de prévision des charges sinistres, aussi bien déterministes que stochastiques, qui permettent d'anticiper les impacts financiers potentiels des tempêtes. Il existe également des dispositifs de réassurance et de prévention mis en place pour renforcer la solvabilité de la compagnie lorsque des aléas climatiques à fortes intensités surviennent.

L'objectif de ce travail est double : d'une part, proposer une méthode d'estimation de la charge sinistres suite au recensement d'une tempête majeure, en tenant compte des spécificités du portefeuille d'assurés de GRAA et des données climatiques propres à cette région. D'autre part, analyser les différents moyens mis en place, en amont de la survenance de la tempête, pour limiter sa sinistralité. Cette démarche vise à offrir des recommandations adaptées pour optimiser la résilience de la compagnie face aux risques climatiques futurs.

---

*Mots-clés : Risques climatiques ; Tempête ; Grêle, Modélisation ; Solvabilité ; Météorologique*

## Abstract

---

Natural risks have become an important topic in recent years due to the increasing frequency and severity of their occurrence. In France, where the climate has become increasingly unpredictable, natural weather events represent a growing threat to insurance companies. Among the most costly climatic events, storm risk holds a significant position, ranking just after floods and droughts in terms of claims burden. For insurance companies, managing this risk is essential to maintaining their financial stability and ensuring their ability to compensate policyholders in the event of a major loss.

This thesis focuses on how storm risk is managed within Groupama Rhône-Alpes Auvergne (GRAA), one of the regional entities of the Groupama group, a key player in the French insurance market. The management of this risk relies on both deterministic and stochastic methods for forecasting claims costs, which help anticipate the potential financial impacts of storms. In addition, reinsurance and prevention mechanisms are implemented to strengthen the company's solvency in the face of severe weather events.

The objective of this work is twofold: first, to propose a method for estimating the claims burden following the occurrence of a major storm, taking into account the specific characteristics of GRAA's policyholder portfolio and the regional climate data. Second, to analyze the various preventive measures put in place prior to a storm to mitigate its potential claims impact. This approach aims to provide tailored recommendations to enhance the company's resilience to future climate-related risks.

---

*Keywords : Climate risks; Storm ; Hail ; Insurance modeling ; Solvency ; Meteorological*

# Note de Synthèse

*Mots-clés : Risques climatiques ; Tempête ; Grêle ; Modélisation ; Solvabilité ; Météorologique*

## Introduction

Le changement climatique constitue un défi majeur pour la société, impactant divers secteurs, dont l'assurance. La fréquence et l'intensité croissantes des événements climatiques extrêmes, comme les tempêtes, les inondations et les sécheresses, engendrent des coûts financiers importants. En France, ces catastrophes ont généré 50 milliards d'euros d'indemnisations en 40 ans, touchant en moyenne 5 700 communes par an.

Parmi ces phénomènes, les tempêtes et la grêle sont particulièrement préoccupants. Une tempête est considérée comme une catastrophe naturelle si les vents dépassent 145 km/h, sinon elle est couverte par la garantie Tempête-Grêle-Neige (TGN). La grêle, quant à elle, provoque des dommages significatifs aux cultures, bâtiments et véhicules.

Le mémoire se concentre sur ces risques climatiques et leur prise en charge assurantielle au niveau de la caisse régionale Groupama Rhône-Alpes Auvergne (GRAA), acteur majeur du secteur, qui a enregistré 550 millions d'euros de charges liées aux tempêtes et à la grêle entre 2018 et 2023.

## Objectifs

Ce rapport vise à analyser la gestion du risque Tempête-Grêle dans une compagnie d'assurance comme GRAA, et notamment les tempêtes de grandes envergures. Il explore le contexte et les enjeux de ces risques climatiques, les méthodes d'évaluation des sinistres actuels et propose une nouvelle approche de modélisation pour mieux estimer le coût d'une tempête de grande envergure, en fonction des caractéristiques climatiques et des profils des assurés. Enfin, il examine les stratégies mises en place pour limiter leur impact financier pour la caisse régionale, notamment à travers la réassurance et la prévention.

## Contexte

Au niveau de GRAA, la charge sinistre liée aux tempêtes est la plus importante parmi celles des risques climatiques sur les dernières années. De plus, ces phénomènes s'intensifient de plus en plus et leur implication dans les résultats de la caisse devient de plus en plus importante. C'est pourquoi

la gestion de ce risque représente aujourd'hui un enjeu majeur pour Groupama Rhône-Alpes Auvergne.

Actuellement, lorsqu'un évènement de vents violents ou de grêles intervient, ce qui signifie soit qu'un évènement national a été déclaré soit que le seuil de montant de sinistres déclarés en 72h (qui est de 310k€ en 2025) a été dépassé, les cadences de règlements des tempêtes historiques sont utilisées pour estimer la charge totale de la tempête. Cependant cette méthode ne prends pas en compte les caractéristiques propres aux tempêtes et aux assurés exposés.

Pour pallier à ces problèmes, une autre méthode pour modéliser la charge sinistres de ces tempêtes appelées "majeures" sera développée dans cette étude, la méthode "Fréquence x Coûts Moyens".

## Construction de la base de données

La construction de la base de données est essentielle pour modéliser la sinistralité lors de la survenance d'une tempête. Elle repose sur le regroupement de plusieurs bases distinctes, sachant que la modélisation se fera uniquement sur les contrats habitation :

### — La base météorologique

Les données météorologiques proviennent de deux sources : le site [www.meteo.data.gouv.fr](http://www.meteo.data.gouv.fr) pour les données liées au vent et aux quantités de précipitation et le centre d'étude météorologique Kéraunos, dont GRAA a obtenu une base contenant des informations sur la taille des grêlons tombés dans toute la France. Elles incluent des relevés quotidiens des stations situées dans la zone de Groupama Rhône-Alpes Auvergne (GRAA) depuis 2004.

### — La base portefeuille

Elle regroupe tous les contrats d'assurance habitation de GRAA entre 2012 et 2025. Les variables principales incluent le numéro de contrat, l'année de souscription, le numéro insee, la nature de l'habitation, la nature de l'assuré, la période de construction du bâtiment, l'usage de l'habitation, le nombre de pièces, et le niveau de garantie.

### — La base sinistre

La base sinistres recense tous les sinistres sur bâtiments des segments de revue pour les survenances de 2012 à 2025. Les variables utiles sont le numéro d'avis de sinistre, la date de survenance, le code postal, le coût du sinistre, et la description du sinistre.

Afin de distinguer les tempêtes dont la cause principale est le vent et celles provoquées par la grêle. Deux modèles différents seront construits pour s'adapter au mieux aux caractéristiques de ces deux aléas.

Ainsi, une tempête de grêle survenue en juin 2022 et une tempête de vent en décembre 2019 ont été identifiées comme tempêtes de références pour l'entraînement et le test des modèles. Les tempêtes ont été sélectionnées en fonction de leur charge sinistre et du nombre d'assurés exposés. Puisque les bases d'assurés présents dans les numéros insee comportant au moins un sinistre lié à ces tempêtes présentaient des pourcentages d'assurés sinistrés trop faibles (inférieurs à 3%), il a été décidé de considérer que les assurés exposés aux tempêtes représentaient ceux présents dans les numéros insee comportant au moins 5% d'assurés sinistrés.

Phénomène	Date	Notation	Nombre d'assurés exposés	Nombre de sinistres	Montant de sinistres
Grêle	3 juin 2022	G1	20 212	3055	44,6M€
Vent	21-23 décembre 2019	V1	4409	812	1,6M€

TABLE 1 – Présentation des tempêtes retenues

## Modélisation de la fréquence de sinistres

Pour prédire la probabilité de sinistre de chaque assuré, plusieurs variables ont été jugées comme explicatives de la sinistralité ou non d'un assuré :

- Vitesse du vent : Vitesse maximale de vent enregistrée le jour de survenance du sinistre dans la station météo la plus proche (ajustée avec la distance avec le lieu du sinistre).
- Quantité de précipitations : Précipitations tombées en 24h le jour de survenance du sinistre dans la station météo la plus proche (ajustée avec la distance avec le lieu du sinistre).
- Taille des grêlons : Influence directe sur la fréquence des sinistres liés à la grêle.
- Nature de l'assuré et de l'habitation : les propriétaires de maisons sont plus touchés que les locataires ou les propriétaires d'appartements.
- Usage de l'habitation : Les résidences principales révèlent des taux de sinistralité plus élevés que les résidences secondaires
- Nombre de pièces de l'habitation : Plus la taille de l'habitation est grande, plus la probabilité de sinistre est élevée.
- Période de construction de l'habitation : Les constructions antérieures à 1950 présentent une plus grande sinistralité.
- Formule : Au plus il y a de garanties dans le contrat de l'assuré, au plus celui-ci a de chances de déclarer un sinistre.

Trois types de modèles probabilistes entraînés ont été testés :

- Arbre de décision : Simplicité d'interprétation, mais risque de sur-apprentissage.
- Forêts aléatoires : Combinaison de plusieurs arbres de décision pour améliorer la robustesse et la précision des prédictions.
- Modèles linéaires généralisés (GLM) : L'estimation des probabilités de sinistre se fait à l'aide d'une régression logistique.

Le choix du modèle le plus adapté pour chaque tempête s'est fait en comparant trois métriques : l'Accuracy, le F1-Score et l'AUC. Ces métriques sont calculées en appliquant le modèle entraîné avec 80% de chaque base de données sur les 20% restant, appelées bases de test.

### Pour la tempête V1 :

Métriques	Arbre de décision	Forêts aléatoires	MLG
Accuracy	0,95	0,96	0,82
F1-Score	0,86	0,89	0,53
AUC	0,89	0,95	0,72

TABLE 2 – Métriques des modèles pour la tempête V1

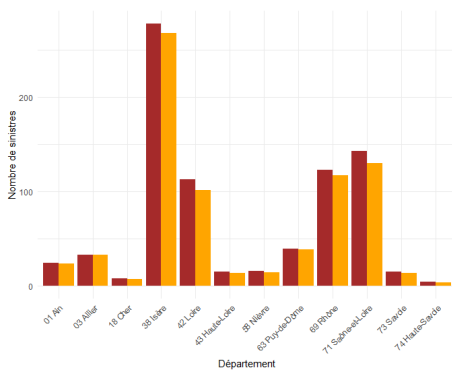
Pour la tempête G1 :

Métriques	Arbre de décision	Forêts aléatoires	MLG
Accuracy	0,86	0,87	0,84
F1-Score	0,54	0,51	0,53
AUC	0,83	0,70	0,70

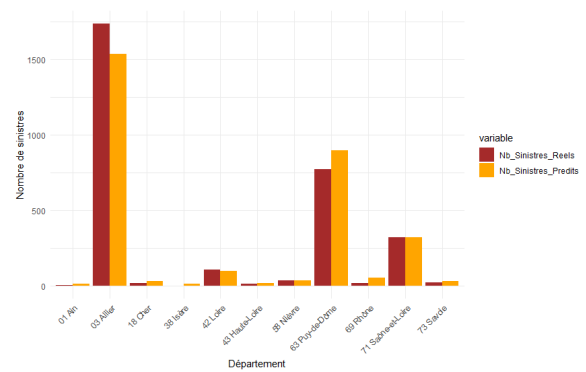
TABLE 3 – Métriques des modèles pour la tempête G1

En comparant les résultats de chaque modèle pour les deux tempêtes, il en ressort que le modèle le plus performant pour la base d'assurés exposés à la tempête V1 est la forêt aléatoire tandis ce que celui pour les assurés exposés à la tempête G1 est l'arbre de décision.

Pour tester la cohérence des modèles retenus, ils ont été appliqués aux bases entières des tempêtes V1 et G1. Ainsi, la comparaison entre les nombres de sinistres réels et prédits par départements est exprimée sur ces graphes :



(a) Nombres de sinistres réels et prédits pour la tempête V1



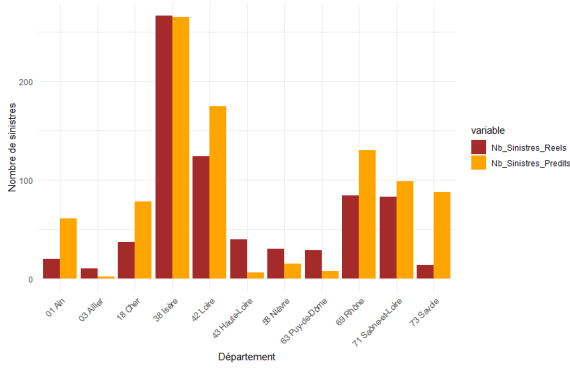
(b) Nombres de sinistres réels et prédits pour la tempête G1

Ces résultats montrent une bonne estimation globale du nombre de sinistres sur chaque départements, variable qui n'intervient pas dans les modèles pour pouvoir les appliquer à d'autres tempêtes. Les modèles ont d'ailleurs été testés sur deux autres tempêtes, ce qui a donné lieu aux prédictions suivantes :

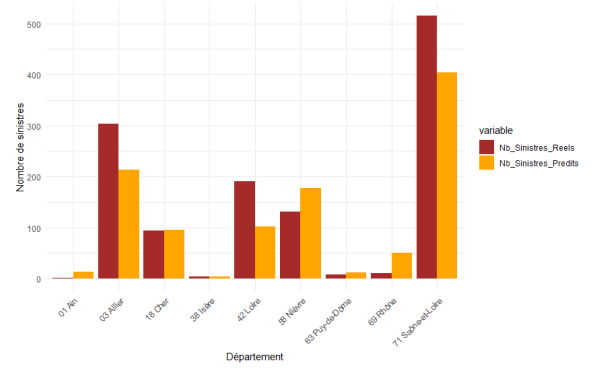
Tempête	Nombre Sinistres Réel	Nombre Sinistres Prédit
G2	1259	1072
V2	757	914

TABLE 4 – Présentation des résultats des modèles sur les tempêtes V2 et G2

Et les résultats ont sans surprise révélés des écarts plus importants mais qui restent malgré tout satisfaisants.



(a) Nombres de sinistres réels et prédits pour la tempête V2



(b) Nombres de sinistres réels et prédits pour la tempête G2

## Estimation des coûts moyens par assuré

Le but de cette partie est d'associer à chaque assuré un coût moyen de sinistre en se basant sur ses caractéristiques et les conditions météo auxquelles il a été exposé.

Les analyses sur l'influence de chaque sur les montants de sinistres ont montrées que chacune des tempêtes V1 et G1, trois variables sont explicatives du coût d'un sinistre. Le coût moyen associé à un assuré  $i$  exposé à la tempête V1 est donc calculé comme suit :

$$\mathbb{E}[\text{Coût}_i | X_i] = \text{Mean}(\text{CTP\_ajust} \mid \text{Inter\_vent} = \text{Inter\_vent}_i \cap \text{Inter\_precip} = \text{Inter\_precip}_i \cap \text{nb\_pieces} = \text{nb\_pieces}_i)$$

Tandis ce que pour un assuré  $j$  exposé à la tempête G1, la formule devient :

$$\mathbb{E}[\text{Coût}_j | X_j] = \text{Mean}(\text{CTP\_ajust} \mid \text{Classe\_grêlons} = \text{Classe\_grêlons}_j \cap \text{Inter\_precip} = \text{Inter\_precip}_j \cap \text{nb\_pieces} = \text{nb\_pieces}_j)$$

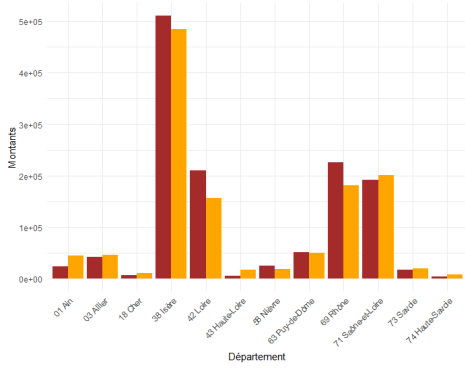
## Estimation de la charge totale de sinistre

Les deux parties précédentes ont permis de calculer pour chaque assuré sa charge de sinistre estimée, et ainsi en sommant les charges pour chacun des assurés, une estimation de la charge totale des tempêtes en habitation est obtenue :

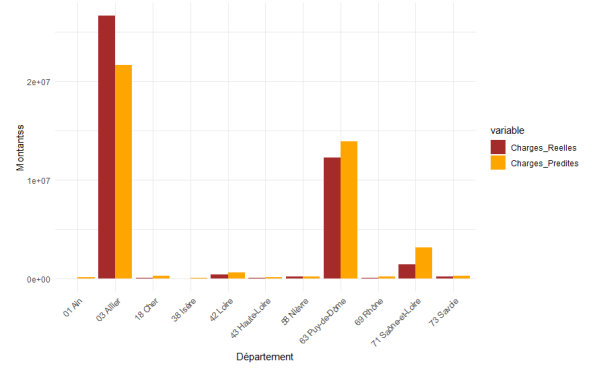
$$\text{Charge totale prédite} = \sum_{i=1}^n P(\text{Presence}_i = 1) \times \mathbb{E}[\text{Coût}_i | X_i]$$

Les graphiques suivants montrent les charges réelles et prédites pour les tempêtes V1 et G1 :





(a) Charges de sinistres réelles et prédites pour la tempête V1

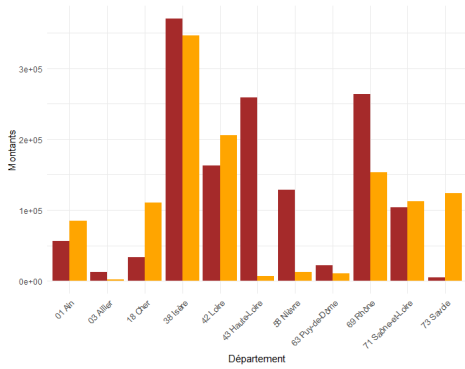


(b) Charges de sinistres réelles et prédites pour la tempête G1

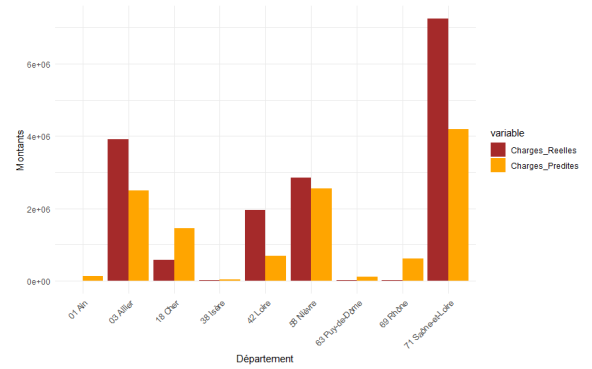
Ces applications montrent des résultats cohérents pour les tempêtes sur lesquelles ont été entraînés les modèles, malgré une légère sous-estimation du montants des sinistres. Voyons ce qu'il en est en appliquant ces modèles aux tempêtes V2 et G2 :

Tempête	Charge Sinistres Réelle	Charge Sinistres Prédite
G2	16,6M€	12,1M€
V2	1,4M€	1,2M€

TABLE 5 – Présentation des résultats des modèles de charge sur les tempêtes V2 et G2



(a) Charges de sinistres réelles et prédites pour la tempête V2

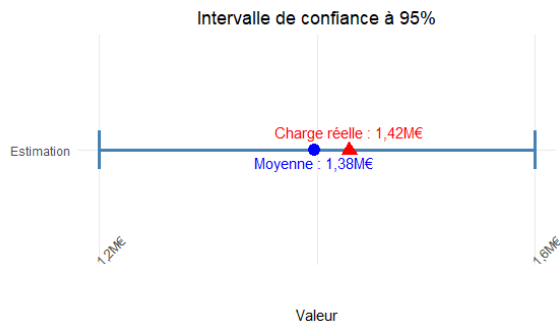


(b) Charges de sinistres réelles et prédites pour la tempête G2

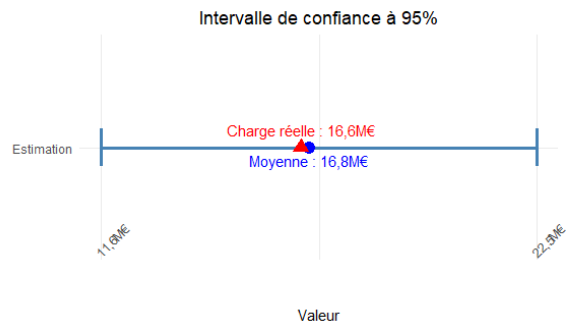
On obtient des valeurs prédites encore une fois inférieures aux valeurs réelles mais les estimations respectent tout de même bien les ordres de grandeurs des valeurs réelles.

## Intervalles de confiance

Afin de déterminer les incertitudes liées aux modèles développés, une méthode bootstrap a été appliquée à chacun d'entre eux. Voici les intervalles de confiances à 95% obtenus :



(a) Intervalle de confiance à 95% associé au modèle de la tempête V2



(b) Intervalle de confiance à 95% associé au modèle de la tempête G2

Ces graphiques montrent que le recalibrage des modèles pour un nombre élevé d'échantillons bootstrap permet d'obtenir des valeurs moyennes de charges de sinistres estimées plus proche de la réalité que l'application simple des modèles aux bases d'assurés.

Cependant la grande amplitude de l'intervalle de confiance associé au modèle prédisant le coût de la tempête G2 montre que cette méthode présente toujours certaines limites qui font augmenter les incertitudes sur les prédictions de coût.

## Limites de la méthode

Les écarts et les incertitudes élevées que l'on observe sur les résultats précédents peuvent venir de plusieurs éléments :

- L'extraction des données météorologiques : Les données relatives à chaque assuré correspondant à celles de la station la plus proche du lieu de l'habitation ajustées avec la distance entre les deux, l'incertitude par rapport à ces données peuvent être non-négligeables.
- Historique de tempêtes majeures faible : Le faible nombre de tempêtes majeures dans les bases de données de GRAA constitue également un facteur d'augmentation des incertitudes liées aux modèles.
- Sélection du meilleur modèle : Le choix du modèle le plus performant basé uniquement sur la comparaison des métriques a pu laisser place à un sur-ajustement des modèles aux tempêtes d'entraînement, notamment pour la tempête G1 dont le modèle d'arbre de décision a été retenu alors que le risque de sur-ajustement sur ce type de modèle est supérieur à celui des forêts aléatoires ou des régressions logistiques.
- Nombre de critères pour estimer les coûts moyens : En effet pour chacun des modèles, seuls 3 critères se sont révélés impactant par rapport au coût d'un sinistre. La recherche d'autres critères pourrait être une solution pour affiner les groupes d'assurés et ainsi réduire l'incertitude liée à ces estimations.

## Réassurance

La réassurance chez GRAA est structurée principalement autour de protections en excédent de sinistre par événement et par risque (réassurance non-proportionnelle) combinées à une réassurance en QUOTE-PART (réassurance proportionnelle) à l'aide d'un pool de rétrocession qui rassemble une partie des sinistres subis par chacune des caisses régionales et qui les ré-allouent à chaque entité en fonction de ses cotisations acquises.

L'application de ce système de réassurance montre une forte influence sur la réduction des coûts des

sinistres à la charge de GRAA pour les événements de tempête et en particulier les plus importants. Concernant les primes de réassurances, Groupama utilise des outils de marché pour estimer la probabilité de survenance d'une tempête ainsi que sa gravité. Ces outils incluent des courbes de charges sinistres et de périodes de retour développées par des sociétés comme RMS et ABS Consulting. La construction de ces courbes implique une approche multidisciplinaire intégrant les modules aléa, enjeu, vulnérabilité et résultats.

Les courbes de charge de sinistre et de période de retour permettent d'estimer les pertes économiques potentielles et de planifier des mesures d'atténuation et de réponse appropriées.

## Prévention

Des dispositifs sont mis en place auprès des assurés pour atténuer les probabilités de survenances des sinistres ainsi que leur coût lorsqu'un événement de tempête ou de grêle a lieu :

- Des alertes météo par SMS : Envoyés aux assurés concernés lorsque des intempéries sont prévues dans la journée.
- Des programmes d'incitation de et de réduction de primes : GRAA possède des partenariats avec des entreprises proposant des moyens de se protéger des tempêtes comme HB Collection qui propose des bâches anti-grêle à mettre sur son véhicule. Les assurés sinistrés en s'étant munis de ces bâches se voient leur franchise offerte.
- Des programmes de sensibilisation : Communication et prévention envers les assurés pour qu'ils puissent appliquer les bons gestes en cas de survenance de tempête.

## Conclusion

Ce mémoire met en évidence l'impact croissant des tempêtes et de la grêle sur le secteur de l'assurance et la nécessité d'outils performants pour anticiper ces risques. Ainsi, une méthode d'estimation de la charge sinistres lors de la survenance d'une tempête de grande ampleur basée sur les caractéristiques du portefeuille d'assuré et les circonstances météorologiques a été développée, les tests réalisés ont montré des résultats significatifs mais certaines limites restent à être travaillées pour permettre de réduire les incertitudes liées à ceux-ci. Par ailleurs, l'importance de la réassurance et des actions de prévention a été soulignée comme levier clé pour limiter les pertes et assurer la solvabilité de la caisse régionale. Enfin, les modèles élaborés pourraient être étendus à d'autres types de contrats, ouvrant ainsi des perspectives d'optimisation dans la gestion des risques climatiques.

# Synthesis Note

*Keywords : Climate risks ; Storm ; Hail ; Modeling ; Solvency ; Meteorology*

## Introduction

Climate change represents a major challenge for society, impacting various sectors, including insurance. The increasing frequency and intensity of extreme weather events such as storms, floods, and droughts result in significant financial costs. In France, these disasters have led to €50 billion in compensations over 40 years, affecting an average of 5,700 municipalities per year.

Among these phenomena, storms and hail are particularly concerning. A storm is considered a natural disaster if wind speeds exceed 145 km/h; otherwise, it is covered by the Storm-Hail-Snow (SHS) insurance guarantee. Hail, on the other hand, causes significant damage to crops, buildings, and vehicles.

This thesis focuses on these climate risks and their insurance coverage at the regional fund Groupama Rhône-Alpes Auvergne (GRAA), a major player in the sector, which recorded €550 million in storm and hail-related expenses between 2018 and 2023.

## Objectives

This report aims to analyze the management of Storm-Hail risk within an insurance company such as GRAA, particularly focusing on large-scale storms. It explores the context and challenges of these climate-related risks, current claims assessment methods, and proposes a new modeling approach to better estimate the cost of a major storm based on climatic characteristics and policyholder profiles. Finally, it examines the strategies implemented to mitigate their financial impact on the regional fund, particularly through reinsurance and prevention.

## Context

At GRAA, storm-related claims expenses have been the highest among climate risks in recent years. Furthermore, these phenomena are becoming more intense, making their impact on the company's financial results increasingly significant. As a result, managing this risk is now a major concern for Groupama Rhône-Alpes Auvergne.

Currently, when a windstorm or hailstorm event occurs—either because a national event has been declared or the threshold of reported claims within 72 hours (set at €310,000 in 2025) has been exceeded—the settlement patterns of past storms are used to estimate the total cost of the storm. However, this method does not take into account the specific characteristics of the storms or the profiles of the exposed policyholders.

To address these issues, a different method for modeling storm-related claims expenses will be developed in this study : the "Frequency x Average Cost" method.

## Database Construction

Building a database is essential for modeling claims when a storm occurs. It involves merging several distinct datasets, with modeling focusing exclusively on home insurance policies :

### — Meteorological database

Meteorological data come from two sources : the website [www.meteo.data.gouv.fr](http://www.meteo.data.gouv.fr) for wind and precipitation data and the meteorological research center Kéraunos, from which GRAA obtained a database containing information on hailstone sizes across France. This data includes daily records from weather stations in the Groupama Rhône-Alpes Auvergne (GRAA) area since 2004.

### — Policy database

This database includes all GRAA home insurance contracts from 2012 to 2025. Key variables include policy number, subscription year, INSEE code, type of dwelling, type of policyholder, construction period of the building, dwelling usage, number of rooms, and coverage level.

### — Claims database

The claims database records all building-related claims for the selected review segments between 2012 and 2025. Relevant variables include claim notice number, occurrence date, postal code, claim cost, and claim description.

To distinguish between storms caused primarily by wind and those caused by hail, two separate models will be built to best adapt to the characteristics of each hazard.

Thus, a hailstorm in June 2022 and a windstorm in December 2019 have been identified as reference storms for model training and testing. The storms were selected based on their claims expenses and the number of exposed policyholders. Since the proportion of policyholders with claims in the INSEE areas affected by these storms was too low (below 3%), it was decided to consider that exposed policyholders were those located in INSEE areas where at least 5% of policyholders had claims.

Phenomenon	Date	Notation	Number of Exposed Policyholders	Number of Claims	Claims Amount (€)
Hail	June 3, 2022	G1	20,212	3,055	44.6M€
Wind	December 21-23, 2019	V1	4,409	812	1.6M€

TABLE 6 – Presentation of selected storms

## Claims Frequency Modeling

To predict the probability of a claim for each policyholder, several variables were considered as explanatory for claims occurrence :

- Wind speed : Maximum wind speed recorded on the claim date at the nearest weather station (adjusted for distance).
- Precipitation amount : 24-hour precipitation recorded on the claim date at the nearest weather station (adjusted for distance).
- Hailstone size : Direct influence on the frequency of hail-related claims.
- Type of policyholder and dwelling : Homeowners are more affected than tenants or apartment owners.
- Dwelling usage : Primary residences have higher claim rates than secondary homes.
- Number of rooms : The larger the dwelling, the higher the probability of a claim.
- Construction period : Homes built before 1950 exhibit higher claim frequencies.
- Coverage plan : More comprehensive insurance plans increase the likelihood of a claim being reported.

Three probabilistic models were tested :

- Decision Tree : Easy to interpret but prone to overfitting.
- Random Forest : Combines multiple decision trees to improve robustness and prediction accuracy.
- Generalized Linear Models (GLM) : Uses logistic regression to estimate claim probabilities.

The best model for each storm was chosen based on three metrics : Accuracy, F1-Score, and AUC. These metrics were calculated by training the model on 80% of each dataset and testing it on the remaining 20%.

### For Storm V1 :

Metrics	Decision Tree	Random Forest	MLG
Accuracy	0.95	0.96	0.82
F1-Score	0.86	0.89	0.53
AUC	0.89	0.95	0.72

TABLE 7 – Model metrics for Storm V1

### For Storm G1 :

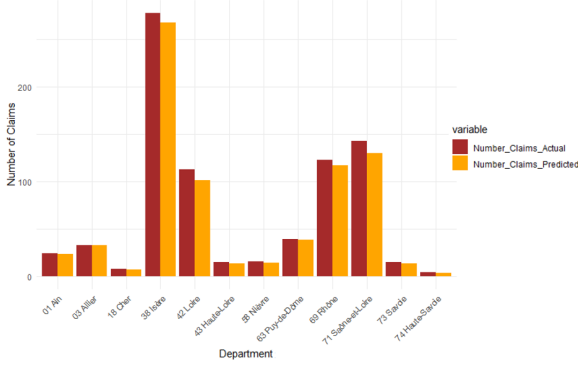
Metrics	Decision Tree	Random Forest	MLG
Accuracy	0.86	0.87	0.84
F1-Score	0.54	0.51	0.53
AUC	0.83	0.70	0.70

TABLE 8 – Model metrics for Storm G1

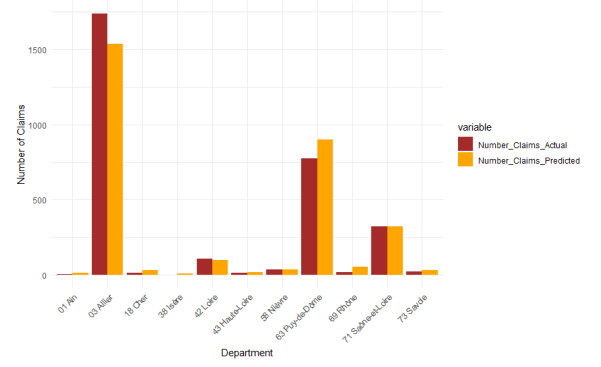
Comparing the results of each model for the two storms, it appears that the most efficient model for the insured individuals exposed to Storm V1 is the Random Forest, whereas for those exposed to

Storm G1, the Decision Tree performs best.

To test the consistency of the selected models, they were applied to the entire datasets of Storms V1 and G1. Thus, the comparison between the actual and predicted number of claims per department is represented in these graphs :



(a) Actual and predicted number of claims for Storm V1



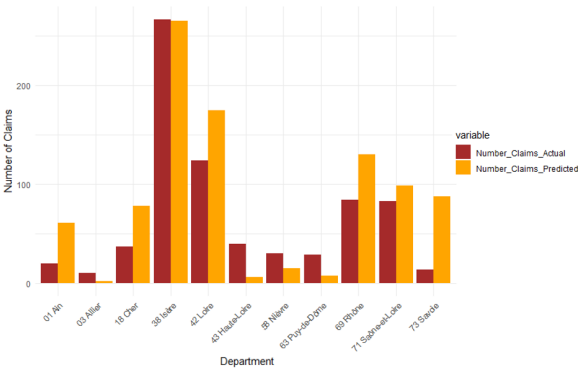
(b) Actual and predicted number of claims for Storm G1

These results show a good overall estimation of the number of claims per department, a variable that is not included in the models to allow them to be applied to other storms.

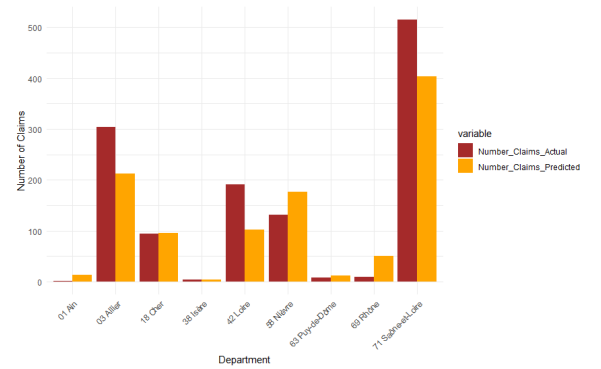
The models were also tested on two other storms, yielding the following predictions :

Storm	Actual Number of Claims	Predicted Number of Claims
G2	1259	1072
V2	757	914

TABLE 9 – Presentation of model results for Storms V2 and G2



(a) Actual and predicted number of claims for Storm V2



(b) Actual and predicted number of claims for Storm G2

As expected, the results revealed larger discrepancies, but they remain satisfactory nonetheless.

## Average Cost Estimation

The objective of this section is to associate each insured individual with an average claim cost based on their characteristics and the weather conditions to which they were exposed.

Analyses on the influence of different factors on claim amounts have shown that for each of the storms V1 and G1, three variables explain the cost of a claim. The average cost associated with an insured individual  $i$  exposed to Storm V1 is therefore calculated as follows :

$$\mathbb{E}[\text{Cost}_i|X_i] = \text{Mean}(\text{CTP\_adjusted} \mid \text{Wind\_Interval} = \text{Wind\_Interval}_i \cap \text{Precip\_Interval} = \text{Precip\_Interval}_i \cap \text{num\_rooms} = \text{num\_rooms}_i)$$

Whereas for an insured individual  $j$  exposed to Storm G1, the formula becomes :

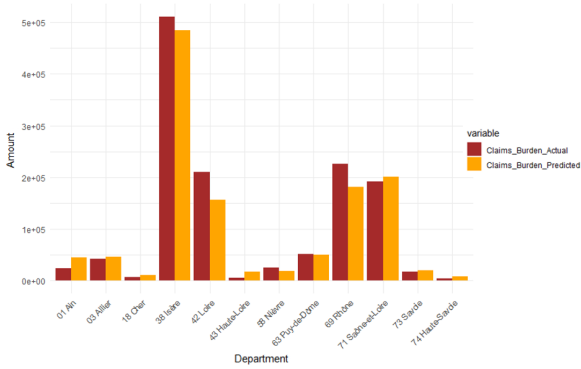
$$\mathbb{E}[\text{Cost}_j|X_j] = \text{Mean}(\text{CTP\_adjusted} \mid \text{Hailstone\_Class} = \text{Hailstone\_Class}_j \cap \text{Precip\_Interval} = \text{Precip\_Interval}_j \cap \text{num\_rooms} = \text{num\_rooms}_j)$$

## Estimation of the Total Claim Burden

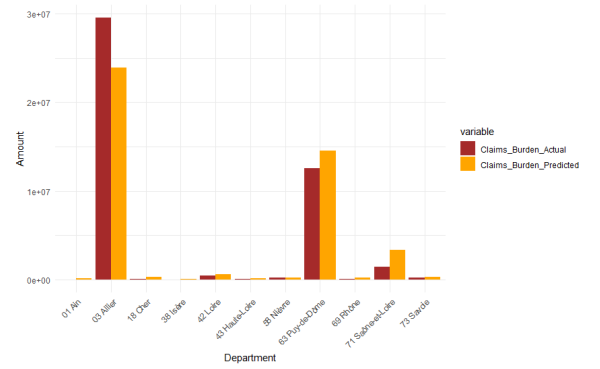
The previous sections allowed for the calculation of the estimated claim burden for each insured individual. By summing these burdens for all insured individuals, an estimate of the total claim burden for the storms on housing is obtained :

$$\text{Predicted total burden} = \sum_{i=1}^n P(\text{Presence}_i = 1) \times \mathbb{E}[\text{Cost}_i|X_i]$$

The following graphs show the actual and predicted claim burdens for Storms V1 and G1 :



(a) Actual and predicted claim burdens for Storm V1



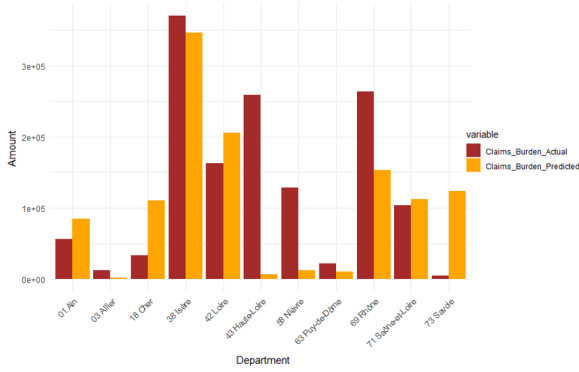
(b) Actual and predicted claim burdens for Storm G1

These applications show consistent results for the storms on which the models were trained, despite a slight underestimation of claim amounts. Let's now examine the results when applying these models to Storms V2 and G2 :

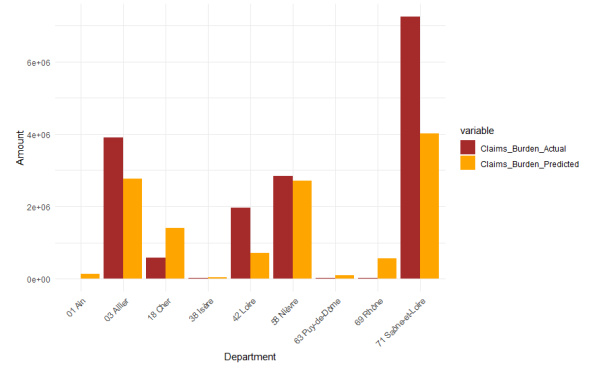


Storm	Actual Claim Burden	Predicted Claim Burden
G2	16,6M€	12,1M€
V2	1,4M€	1,2M€

TABLE 10 – Presentation of model results for claim burden on Storms V2 and G2



(a) Actual and predicted claim burdens for Storm V2

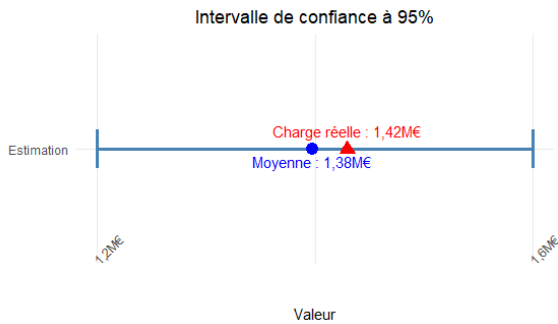


(b) Actual and predicted claim burdens for Storm G2

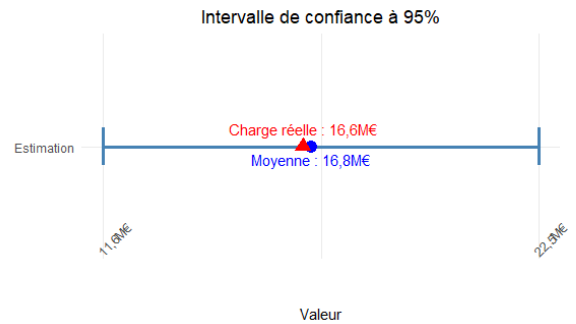
Once again, the predicted values are lower than the actual values, but the estimates still maintain the correct order of magnitude compared to the actual values.

## Confidence intervals

To determine the uncertainties associated with the developed models, a bootstrap method was applied to each of them. The following 95% confidence intervals were obtained :



(a) 95% confidence interval associated with the V2 storm model



(b) 95% confidence interval associated with the G2 storm model

These graphs show that recalibrating the models with a large number of bootstrap samples yields average estimated claims costs that are closer to reality than simply applying the models to the policyholder databases.

However, the wide confidence interval associated with the model predicting the cost of storm G2

indicates that this method still has certain limitations, which increase the uncertainty of the cost predictions.

## Limitations of the Method

The discrepancies observed in the previous results may be due to several factors :

- Extraction of meteorological data : The data related to each insured individual corresponds to that of the nearest weather station to the insured's residence, adjusted for the distance between the two. The uncertainty regarding this data may be significant.
- Limited history of major storms : The small number of major storms in GRAA's databases is also a factor that increases the uncertainty associated with the models.
- Selection of the best model : The choice of the most effective model based solely on metric comparisons may have led to overfitting to the training storms. This is particularly relevant for Storm G1, where the decision tree model was selected, despite the fact that this type of model is more prone to overfitting compared to random forests or logistic regressions.
- Number of criteria used to estimate average costs : Indeed, for each model, only three criteria were found to have a significant impact on claim costs. Exploring additional criteria could be a way to refine policyholder groupings and thereby reduce the uncertainty of these estimates.

## Reinsurance

Reinsurance at GRAA is primarily structured around excess-of-loss protections per event and per risk (non-proportional reinsurance), combined with QUOTA-SHARE reinsurance (proportional reinsurance). This is achieved through a retrocession pool that consolidates a portion of the claims incurred by each regional fund and reallocates them to each entity based on its earned premiums. The application of this reinsurance system significantly reduces the cost of claims borne by GRAA for storm events, particularly for the most severe ones.

Regarding reinsurance premiums, Groupama utilizes market tools to estimate the probability of storm occurrence and its severity. These tools include loss load curves and return periods developed by companies such as RMS and ABS Consulting. The construction of these curves involves a multidisciplinary approach integrating hazard, exposure, vulnerability, and results modules.

Loss load and return period curves allow for the estimation of potential economic losses and the planning of appropriate mitigation and response measures.

## Prevention

Several measures have been implemented for policyholders to mitigate both the probability of claims and their associated costs when a storm or hail event occurs :

- Weather alerts via SMS : Sent to affected policyholders when severe weather is forecasted for the day.
- Incentive and premium reduction programs : GRAA has partnerships with companies offering protective solutions against storms. For example, HB Collection provides anti-hail covers for vehicles. Policyholders who use these covers and suffer damage are exempt from paying their deductible.
- Awareness programs : Communication and prevention initiatives are in place to educate policyholders on the best actions to take in case of a storm.

## Conclusion

This thesis highlights the growing impact of storms and hail on the insurance sector and the need for effective tools to anticipate these risks. To that end, a method for estimating claim costs following the occurrence of a storm—based on the characteristics of the policyholder portfolio and meteorological conditions—has been developed. Furthermore, the importance of reinsurance and preventive actions has been emphasized as key levers for limiting losses. Finally, the models developed could be extended to other types of insurance contracts, thus opening up new opportunities for optimization in the management of climate-related risks.

# Remerciements

Je voudrais en premier lieu remercier mon tuteur d’alternance Rémi GREGOIRE, qui est également manager de l’équipe actuariat, non seulement pour m’avoir permis d’intégrer cette magnifique équipe et d’avoir passé une année très enrichissante par rapport à mes connaissances sur le milieu de l’actuariat, mais aussi pour m’avoir parfaitement accompagné tout au long de la création de ce mémoire, sans oublier les autres membres de l’équipe actuariat, Aymen TAMNAT et Yohan NERAUDAU, respectivement actuaire junior et stagiaire au moment de mon alternance.

Je souhaiterais également remercier le directeur financier de Groupama Rhône-Alpes Auvergne, Régis LONGIN, qui a été de très bons conseils et qui a pris le temps de répondre à toutes mes interrogations.

Un grand merci à Simon TENTILLIER et Ségolène TATIN, membres de l’équipe ETUDES ET PILOTAGE TECHNIQUE, qui m’ont aidé lors de l’extraction et la création de mes bases de données, ainsi qu’à leur manager, Cyril GERLIER.

Un merci aussi à Rémi CHAPELON, membre de l’équipe ACTUARIAT ET ETUDES TECHNIQUES de la caisse nationale de Groupama, qui m’a apporté son expertise concernant les modélisations des charges sinistre.

Pour finir, un merci tout particulier à l’assistante de la direction financière, Roseline MATHE, pour son soutien tout au long de cette année et son assistance lors de chacune de mes démarches.

# Table des matières

<b>Note de Synthèse</b>	<b>3</b>
<b>Synthesis Note</b>	<b>11</b>
<b>Remerciements</b>	<b>19</b>
<b>Table des matières</b>	<b>20</b>
<b>Introduction Générale</b>	<b>23</b>
<b>1 Contexte</b>	<b>25</b>
1.1 Contexte et Importance des risques climatiques . . . . .	25
1.1.1 Présentation générale des catastrophes naturelles . . . . .	25
1.1.2 Evolution de la sinistralité . . . . .	26
1.1.3 Contexte juridique sur la garantie Cat Nat . . . . .	28
1.2 Contexte et importance des risques Tempêtes, Grêle et Neige . . . . .	29
1.2.1 Définitions . . . . .	29
1.2.2 Evolution de la sinistralité TGN . . . . .	31
1.2.3 Futur de ce risque . . . . .	32
1.3 Gestion actuelle du risque Tempête-Grêle chez GRAA . . . . .	33
1.3.1 Groupama et GRAA . . . . .	33
1.3.2 Risque Tempête-Grêle . . . . .	35
1.3.3 Méthode d'estimation actuelle du coût d'une tempête . . . . .	37
1.3.4 Autre méthode proposée : Fréquence x Coût Moyen . . . . .	38
<b>2 Modélisation de la charge sinistre d'une tempête majeure sur le portefeuille de GRAA</b>	<b>40</b>

## TABLE DES MATIÈRES

2.1	Construction de la base de données . . . . .	40
2.1.1	Données météorologiques . . . . .	40
2.1.2	Portefeuille d'assurés . . . . .	42
2.1.3	Base sinistres . . . . .	45
2.1.4	Base finale . . . . .	49
2.2	Modélisation de la fréquence de sinistres . . . . .	50
2.2.1	Recherche de critères discriminants . . . . .	50
2.2.2	Métriques . . . . .	59
2.2.3	Arbre de décision . . . . .	61
2.2.4	Forêts aléatoires . . . . .	66
2.2.5	Modèles linéaires généralisés . . . . .	70
2.2.6	Analyse des résultats . . . . .	74
2.3	Estimation de la charge sinistre totale des tempêtes . . . . .	78
2.3.1	Estimation du cout moyen des sinistres . . . . .	78
2.3.2	Obtention des estimations de charges sinistres . . . . .	81
2.4	Incertitudes liées aux modèles : Méthode Bootstrap . . . . .	84
2.4.1	Principe de la méthode . . . . .	84
2.4.2	Application aux modèles obtenus . . . . .	85
2.4.3	Obtention d'intervalles de confiance . . . . .	86
2.4.4	Limites de la méthode . . . . .	87
<b>3</b>	<b>Outils de protection contre les tempêtes majeures</b>	<b>90</b>
3.1	Outils de Réassurance . . . . .	90
3.1.1	Différents traités de réassurance . . . . .	90
3.1.2	Formules de réassurance à GRAA . . . . .	91
3.1.3	Application aux charges de sinistres des tempêtes modélisées . . . . .	94
3.2	Outils de prévention . . . . .	96
3.2.1	Les alertes météorologiques . . . . .	96
3.2.2	Programme d'incitation et de réduction de primes . . . . .	97
3.2.3	Education et sensibilisation . . . . .	98
	<b>Conclusion Générale</b>	<b>99</b>

## TABLE DES MATIÈRES

<b>Bibliographie</b>	<b>101</b>
----------------------	------------

# Introduction Générale

Le changement climatique est aujourd'hui l'un des plus grands défis auxquels notre société doit faire face. La multiplication des événements climatiques extrêmes tels que les tempêtes, les inondations et les sécheresses, ainsi que l'augmentation de leur intensité, ont des répercussions considérables sur divers secteurs, notamment l'assurance. En effet, ces événements perturbent non seulement les écosystèmes naturels et les infrastructures humaines, mais engendrent également des coûts financiers très importants pour les compagnies d'assurance et les assurés.

Sur les 40 dernières années, les catastrophes naturelles ont représenté 50 milliards d'euros d'indemnités versées en France, avec en moyenne 5700 communes concernées chaque année, selon une étude du Ministère de la Transition Écologique. Ces indemnités couvrent une variété de dommages allant de la destruction partielle ou totale de biens immobiliers, aux pertes agricoles et aux perturbations des activités économiques. L'impact financier de ces phénomènes est ainsi considérable et nécessite une prise en charge efficace par le secteur assurantiel.

Une tempête peut être catégorisée, d'un point de vue assurantiel, comme une catastrophe naturelle si les rafales de vent enregistrées atteignent 145 km/h. Dans le cas contraire, l'indemnisation se fera sous la garantie Tempête-Grêle-Neige (TGN). Cette distinction est essentielle pour la gestion des indemnités et la mise en place des stratégies de prévention et de protection des biens assurés.

En plus des tempêtes, la grêle constitue également un risque climatique majeur couvert par cette garantie. Les orages de grêle peuvent provoquer d'importants dommages, notamment sur les véhicules, les bâtiments et les cultures agricoles. Les grêlons, pouvant atteindre plusieurs centimètres de diamètre, engendrent des coûts considérables pour les compagnies d'assurance. En effet, en fonction de leur intensité et de la surface touchée, ces épisodes peuvent représenter plusieurs millions d'euros d'indemnités.

Le sujet de ce mémoire portera sur les événements assurés par cette garantie TGN, en mettant en lumière les différents enjeux et méthodes employées pour limiter leur impact.

Le Groupe Groupama est un des acteurs majeurs de l'assurance en France puisqu'il est notamment le 1er assureur dans le secteur agricole et le 3ème assureur dans le domaine de l'habitation. Ces deux secteurs étant fortement exposés aux risques climatiques, les bases de données et les historiques sinistres à disposition constituent un excellent point de départ pour notre étude.

De plus, la région Auvergne-Rhône-Alpes est particulièrement vulnérable face aux risques de tempêtes et de grêle à fortes intensités. En effet, ces événements représentent chez Groupama Rhône-Alpes Auvergne (GRAA) le plus gros risque climatique en termes de montants indemnisés ces dernières années. Ainsi, entre 2018 et 2023, les tempêtes et la grêle ont engendré 550 millions d'euros de charges, soit un montant bien supérieur à celui des sécheresses (150 millions d'euros) et des inondations. Ces chiffres illustrent l'importance de la prise en compte de ces risques dans la stratégie de gestion et de prévention des sinistres climatiques.



## TABLE DES MATIÈRES

L'enjeu pour GRAA réside dans la gestion des deux types de tempêtes : les mineures, plus fréquentes mais moins coûteuses, et les majeures, peu fréquentes mais à sinistralités élevés. L'estimation de la charge de sinistres de ces dernières est un exercice crucial dans le cadre du provisionnement mais cependant difficile par manque d'historique.

L'objectif de ce mémoire sera donc de présenter une méthode d'estimation de la charge sinistres d'une tempête majeure suite à sa survenance, ainsi que les mécanismes mis en place permettant de limiter leur impact.

Pour cela trois sections distinctes seront détaillées :

Le chapitre 1 sera consacré au contexte et à l'importance des risques climatiques, et en particulier du risque Tempête-Grêle dans le monde, en France et pour GRAA. Il abordera également la méthode utilisée actuellement pour estimer la charge des sinistres d'une tempête une fois celle-ci survenue.

Ensuite, le chapitre 2 aura pour objectif de construire, à partir des bases historiques de sinistres, des portefeuilles d'assurés et des données météorologiques récoltées sur le territoire de GRAA, une nouvelle méthode de modélisation permettant d'estimer le coût d'une tempête une fois qu'elle ai été définie comme majeure. Cette modélisation intégrera les différentes variables climatiques et socio-économiques influençant l'impact financier de ces phénomènes.

Enfin, le chapitre 3 présentera les différents outils permettant à GRAA de se protéger contre les tempêtes à fortes intensités et de limiter leur sinistralité. Il analysera les dispositifs de réassurance ainsi que les actions de prévention mises en place auprès des assurés.

# Chapitre 1

## Contexte et Impact du risque Tempêtes dans le monde et pour GRAA

### 1.1 Contexte et Importance des risques climatiques

En assurance, les sinistres peuvent se décomposer en trois catégories : les sinistres de fréquence (Vol, bris de glasse, dégâts des eaux, accidents de la route mineurs), les sinistres d'intensité (Responsabilité civile corporelle, professionnelle, familiale) et les sinistres cycliques ou pluriannuels (événements naturels). Les sinistres de fréquence et d'intensité sont plus faciles à maîtriser pour les assureurs du fait de leurs caractères répétitifs et historiques. Ce qui est moins le cas des phénomènes climatiques naturels qui relèvent plus de l'aléatoire. C'est pourquoi ces risques représentent un enjeu important dans le monde de l'assurance.

#### 1.1.1 Présentation générale des catastrophes naturelles

##### Définition

Selon l'INSEE,

”Une catastrophe naturelle est caractérisée par l'intensité anormale d'un agent naturel (inondation, coulée de boue, tremblement de terre, avalanche, sécheresse...) lorsque les mesures habituelles à prendre pour prévenir ces dommages n'ont pu empêcher leur survenance ou n'ont pu être prises. Un arrêté interministériel constate l'état de catastrophe naturelle. Il permet l'indemnisation des dommages directement causés aux biens assurés, en vertu de la loi du 13 juillet 1982, relative à l'indemnisation des victimes de catastrophes naturelles.”

Ce type d'évènement doit donc être causé par la nature, mais l'impact de l'homme peut également jouer sur la probabilité de survenance d'une catastrophe naturelle (réchauffement climatique, mauvaise gestion des ressources) et sur sa sévérité (urbanisation, densité de population).

Les catastrophes naturelles peuvent donc prendre différentes formes, les principales sont :

- **Inondation / Tsunami** : Le plus gros tsunami recensé jusqu'à présent est celui de 2004 dans l'océan indien causant la mort de plus de 200 000 personnes..
- **Eruptions volcaniques** : La dernière en date fût celle de Reykjanes en Islande s'étendant de décembre 2023 jusqu'en mars 2024 et causant l'évacuation de milliers de personnes et l'émission de gaz volcaniques.

## CHAPITRE 1. CONTEXTE

- **Tempêtes / Tornadoe** : On en retrouve principalement dans les régions tropicales comme en Floride en 2022 où l'ouragan Ian a causé des destructions massives.
- **Tremblement de terre** : Les dégâts matériels et immatériels peuvent être causés directement par le séisme ou par le potentiel tsunami provoqué par celui-ci. On notera récemment le séisme de Noto au Japon en 2024 qui a été accompagné d'un tsunami ayant fait 299 morts et 1300 blessés.
- **Sécheresse** : En 2024, aux Etats-Unis, environ 43% du pays a subi des conditions de sécheresse, affectant plus de 130 millions de personnes et 229 millions d'acres de cultures.
- **Feu de forêt** : Ceux en Australie en 2019-2020 plus connu sous le nom de "Black Summer" ont brûlé plus de 18 millions d'hectares de forêts, détruits des milliers de maisons et tué ou déplacé près de 3 milliards d'animaux.

### 1.1.2 Evolution de la sinistralité

#### Dans le monde

Au niveau mondial, voici comment se répartissent ces différents événements :

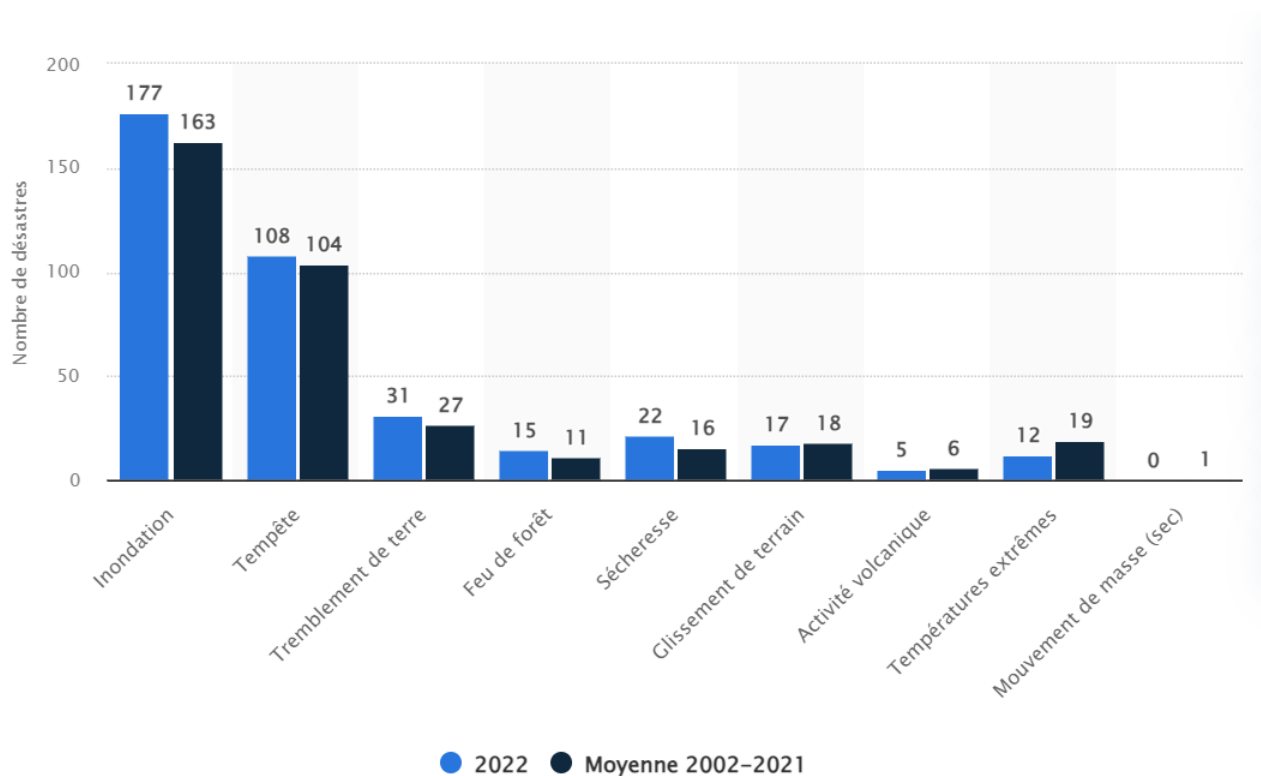


FIGURE 1.1 – Nombre de Cat Nat pour chaque type d'événement (Source : Statista)

On peut voir sur le graphique ci-dessus que la plupart des catastrophes naturelles qui surviennent dans le monde sont des inondations et des tempêtes. On observe également que la fréquence d'événements climatiques est globalement supérieure sur l'année 2022 que sur la moyenne des 20 dernières années.

Cette analyse est cohérente avec le fait que la sinistralité due aux catastrophes naturelles est croissante sur ces dernières années, comme le montre l'Organisation Météorologique Mondiale dans son

## CHAPITRE 1. CONTEXTE

”Atlas de la mortalité et des pertes économiques dues à des phénomènes météorologiques, climatiques et hydrologiques extrêmes (1970-2019)” :

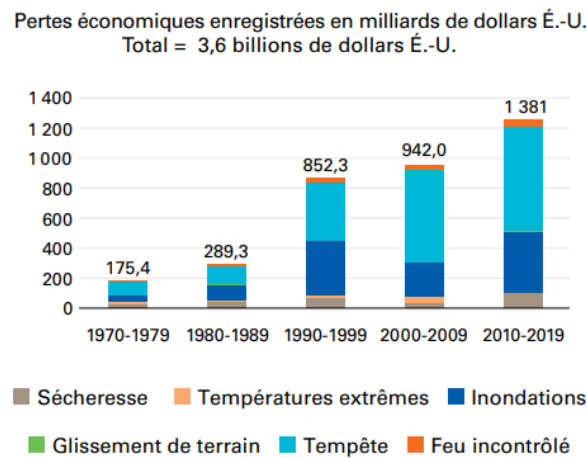


FIGURE 1.2 – Coût des catastrophes naturelles dans le monde par décennie (Source : Organisation Météorologique Mondiale)

### En France

Au niveau national, la Caisse Centrale de Réassurance (CCR) a produit un rapport sur les catastrophes naturelles en France de 1982 à 2021 qui indique que la répartition par périls de sinistres est différente de ce qu’on a vu au niveau mondial :

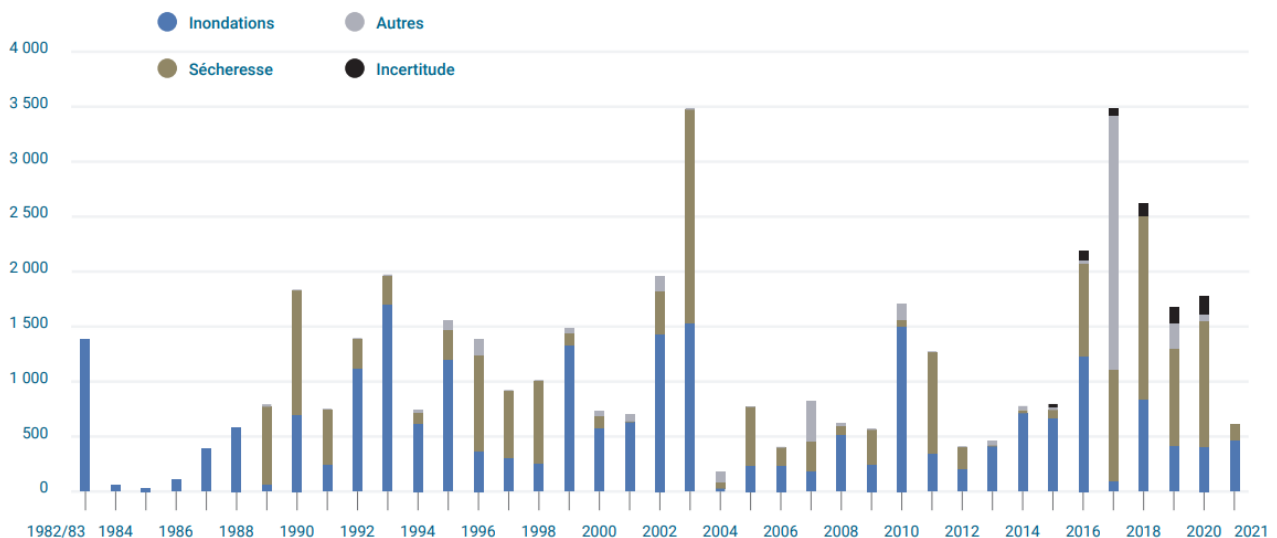


FIGURE 1.3 – Répartition des sinistres Cat Nat par périls de sinistres 1982-2021 (Source : Caisse Centrale de Réassurance)

En effet une grande partie des indemnités par rapport aux catastrophes naturelles en France sont liées à des inondations ou de la sécheresse. Et lorsque l’on regarde le coût de cette sinistralité par année de survenance :

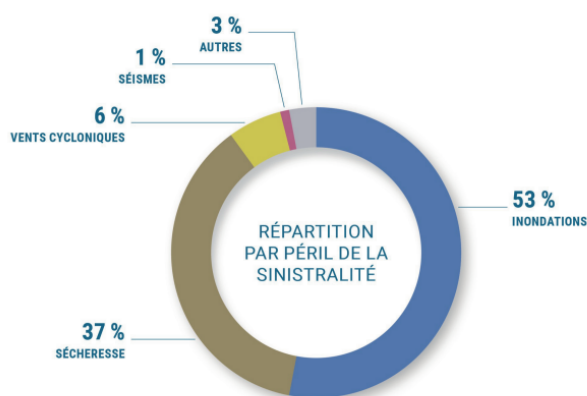


FIGURE 1.4 – La sinistralité catastrophes naturelles non-auto de 1982 à 2021 (en millions d’euros 2021) (Source : Caisse Centrale de Réassurance)

On remarque que les tempêtes ne représentent que 6% de cette charge de sinistre. La figure 1.3 valide aussi une nouvelle fois l’idée que les sinistres liés aux catastrophes naturelles sont supérieures en fréquence et en sévérité ces dernières années par rapport au début du 21ème siècle. D’où l’importance de s’assurer contre ces événements et pour les assureurs de maîtriser au mieux ce risque.

### 1.1.3 Contexte juridique sur la garantie Cat Nat

Après avoir montré la nécessité de plus en plus grande de gérer au mieux ce risque de catastrophes naturelles, voyons le dispositif mis en place par les sociétés d’assurance pour assurer les particuliers et les entreprises : la garantie Cat Nat.

#### Régime légal d’indemnisation

Le régime légal des catastrophes naturelles est encadré par la loi de 13 juillet 1982, cette loi crée pour la première fois un régime d’indemnisation obligatoire qui vise à protéger les assurés contre les effets des catastrophes naturelles, qui jusqu’alors n’étaient que très peu assurés. Selon cette loi, ces effets correspondent aux ”dommages matériels directs non assurables ayant eu pour cause déterminante l’intensité anormale d’un agent naturel lorsque les mesures habituelles à prendre pour prévenir ces dommages n’ont pu empêcher leur survenance ou n’ont pu être prises”.

Le processus d’indemnisation se fait en plusieurs étapes :

#### — Parution d’un arrêté au Journal Officiel

Lorsque l’événement climatique exceptionnel survient, le Maire a 24 mois pour déposer un dossier de demande de reconnaissance de catastrophe naturelle auprès du préfet. Celui-ci va déposer un dossier après de la Commission Interministérielle qui déterminera quelles communes sont concernées par la reconnaissance ou non de l’état de catastrophe naturelle. La garantie Cat Nat ne s’applique que si un arrêté paru au Journal Officiel constate l’état de catastrophe naturelle dans la zone où se trouve les biens de l’assuré et si les dommages subis sont des conséquences directes de la catastrophe naturelle en question.

## CHAPITRE 1. CONTEXTE

### — Demande d'indemnisation de la part de l'assuré

Une fois l'arrêté paru, l'assuré dispose de 30 jours pour déclarer ses sinistres et transmettre à son assureur une estimation des pertes.

### — Indemnisation

L'assureur a alors un mois pour saisir un expert, si nécessaire, qui produira un rapport qui permettra à l'assureur de faire une proposition d'indemnisation. Après réception de l'accord par rapport à la proposition de la part de l'assuré, l'indemnisation se fera dans un délai de 21 jours. Si la compagnie d'assurance préfère faire intervenir des entreprises pour la réparation des dommages subits, elle dispose d'un mois pour missionner cette entreprise.

## Exclusions de garantie

La garantie Cat Nat est une garantie obligatoire pour tous les contrats d'assurance de dommages mise à part pour les biens qui ont été implantés dans des zones inconstructibles après la publication d'un plan de prévention des risques naturels. Il y a également quelques cas que la loi exclu de la garantie Cat Nat :

- les biens et les activités exercés dans les régions et collectivités d'Outre-Mer (Polynésie française, Nouvelle-Calédonie)
- les biens qui ne sont pas couverts par les contrats d'assurance de dommages (tombes) ou les contrats multirisques habitation (clôtures, terrains, jardins)
- les véhicules pour lesquels seule la garantie responsabilité civile obligatoire a été souscrite

La loi prévoit également des franchises maximum qui peuvent être prévues dans les contrats d'assurance dommages. Celle-ci est de 380€ pour les biens à usage d'habitation et non professionnel sauf dans les cas de sinistres sécheresse ou réhydratation des sols.

Les principaux périls couverts par ce régime sont :

- l'inondation et raz-de-marée
- les séismes et mouvement de terrains
- la sécheresse
- l'avalanche

A noter que les périls qui nous intéressent dans cette étude (tempêtes, grêles et neige) ne sont pas pris en charges par la garantie Cat Nat.

## 1.2 Contexte et importance des risques Tempêtes, Grêle et Neige

### 1.2.1 Définitions

#### — Tempêtes

Une tempête est un phénomène météorologique caractérisé par des vents violents qui peuvent s'accompagner de fortes pluies. Elle peut être caractérisée par les valeurs de rafales de vent maximum enregistrées, la durée de l'évènement et la surface sur laquelle s'est étendue la tempête. Les tempêtes naissent à partir du creusement d'une dépression atmosphérique près du sol, l'évolution de cette dépression donne lieu à des vents plus ou moins rapides que l'on mesure grâce à l'échelle de Beaufort.



FIGURE 1.5 – Echelle de Beaufort (Source : MétéoFrance)

Dans le domaine des assurances et de la garantie tempête, cette dénomination est plutôt utilisée lorsque les rafales enregistrées sont supérieures à 100km/h. La dénomination "vent violent" est elle utilisée pour des rafales moyennes d'au moins 89 km/h (10 sur l'échelle de Beaufort).

— **Grêle**

La grêle est une précipitation solide sous forme de grains de glace qui peut endommager sévèrement les cultures, les toitures, les véhicules, et autres biens. La grêle se forme lors de fortes ascendances d'air dans les orages, où les gouttes d'eau sont projetées en altitude et gèlent avant de retomber.

— **Neige**

Les chutes de neige, particulièrement lorsqu'elles sont abondantes, peuvent causer des dommages significatifs. Les accumulations de neige peuvent entraîner l'effondrement de toitures, des dommages aux infrastructures de transport et des coupures de courant. La neige fondue peut également causer des inondations.

### Différence d'indemnisation avec les Cat Nat

Depuis 1990, pour les assureurs, la tempête est distinguée des catastrophes naturelles sauf s'il s'agit d'un évènement cyclonique pour lesquels les vents ont atteint :

- 145 km/h en moyenne à 10 m du sol
- 215 km/h en rafales.

Dans ce cas, c'est la garantie "catastrophe naturelle" qui rentre en compte et l'assuré doit attendre la parution d'un arrêté au Journal Officiel validant l'état de catastrophe naturelle pour sa commune avant de pouvoir demander une indemnisation.

Dans le cas où la tempête ne respecte pas ces conditions, c'est la garantie TGN (Tempêtes, Grêles et Neige) qui va couvrir les effets de cet évènement. Cette garantie obligatoire est incluse dans les contrats d'assurance habitation et prend en charge les dommages liés directement à ces évènements, mais aussi les dégâts causés par la pluie ou les infiltrations d'eau.

Les délais d'indemnisation pour les sinistres que couvrent cette garantie sont différents de ceux de la garantie Cat Nat. En effet l'assuré doit déclarer le sinistre dans un délai de 5 jours après avoir pris connaissance des dommages. Ensuite l'assureur doit transmettre une proposition d'indemnisation après avoir eu recours ou non à une expertise et verser l'indemnité entre 10 et 30 jours après réception de l'accord de l'assuré en fonction du contrat.

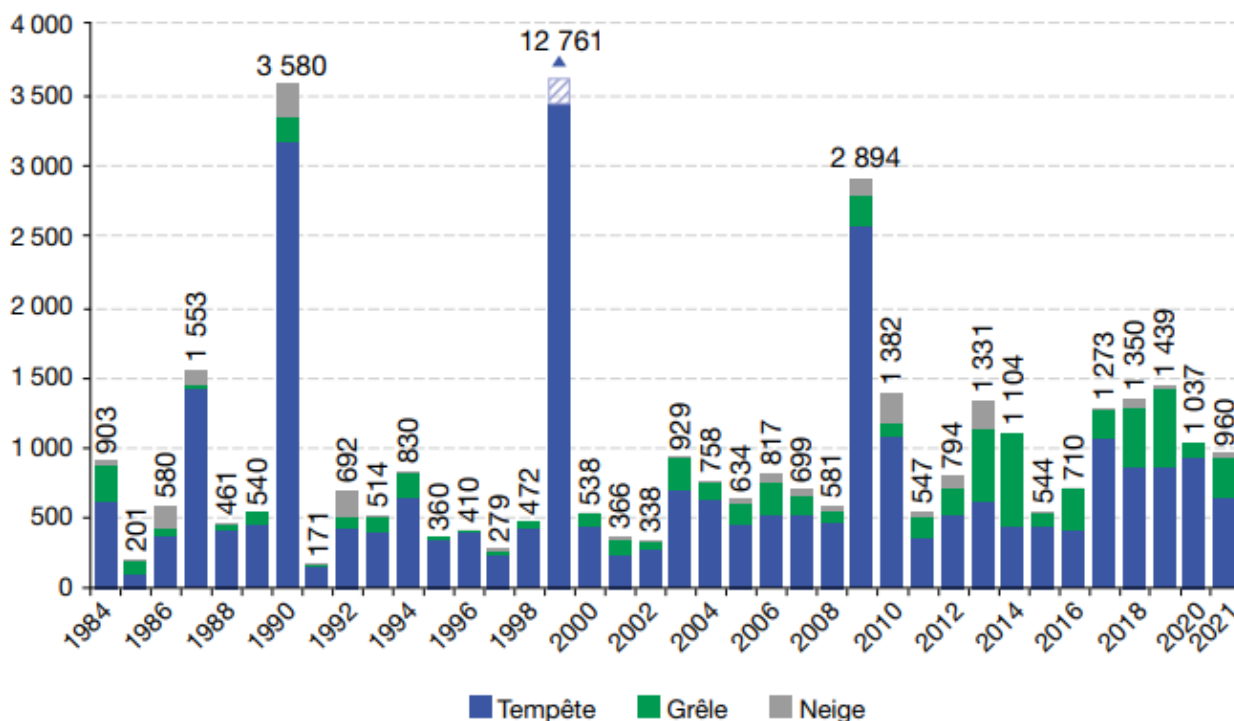
#### 1.2.2 Evolution de la sinistralité TGN

Le risque TGN ne fait pas exception aux risques climatiques. Pour preuve, voici un graphique de la sinistralité liée à ces évènements au cours du temps. A noter que ces chiffres prennent uniquement en compte la garantie TGN sur bâtiment (hors assurances automobile et récoltes) en France Métropolitaine.



## INDEMNISATIONS VERSÉES AU TITRE DES TEMPÊTE-GRÊLE-NEIGE (TGN) DE 1984 À 2021

En M€<sub>2021</sub> (réévalués par l'indice FFB\*)



\* Indice de la Fédération française du bâtiment.

FIGURE 1.6 – Sinistralité liée aux Tempête-Grêle-Neige (Source : France Assureurs)

Si l'on enlève les années exceptionnelles qui ont vu arriver des tempêtes majeures sur le territoire français comme les tempêtes Daria (1990), Lothar et Martin (1999) et Klaus (2009), on observe une augmentation de la sinistralité assurantielle liée aux Tempête-Grêle-Neige (TGN) depuis 1984. En effet, la charge annuelle moyenne des dégâts causés aux bâtiments par les TGN est passée de 706 M€ sur la période 1984-1989, à 830 M€ entre 1990 et 2015, pour atteindre 1 128 M€ depuis 2016. De plus, sur les trente dernières années, parmi les différents aléas naturels donnant lieu à des indemnités chaque année (inondation, sécheresse, etc.), les TGN sont ceux dont le montant annuel est le plus élevé en France.

### 1.2.3 Futur de ce risque

En 2021, France Assureurs a mené une étude décrivant à horizon 2050 l'évolution des différents périls climatiques dont la sécheresse, l'inondation et les tempêtes. Cette étude indique que contrairement à la sécheresse et à l'inondation, le risque tempête en France sera peu impacté par les facteurs socio-économiques et climatiques sur les 30 prochaines années. En effet, les projections distinguent par de tendances significatives par rapport au changement climatique. Le seul facteur faisant augmenter la prévision de charges sinistres en France serait l'effet richesse, qui ferait passer la charge sinistre en France de 31 Milliard d'euros sur la période 1989-2019 de à 46 Milliard sur la période 2020-2050.

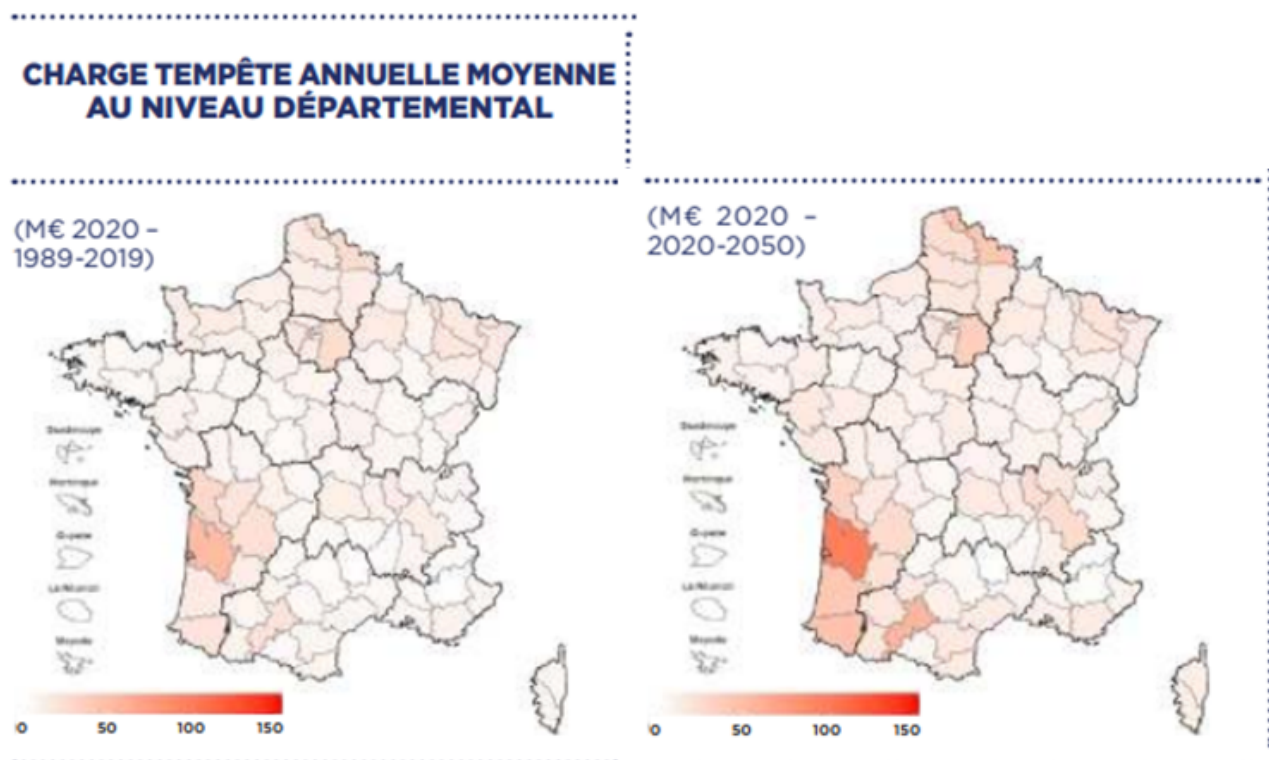


FIGURE 1.7 – Evolution de la charge Tempête annuelle moyenne en France (Source : France Assureurs))

## 1.3 Gestion actuelle du risque Tempête-Grêle chez GRAA

### 1.3.1 Groupama et GRAA

#### Le Groupe Groupama

9ème assureur généraliste en France en 2023 selon L'Arguse de l'Assurance, Groupama est un groupe mutualiste spécialisé dans les assurances agricoles, santé individuelle, habitation et auto. Il est présent dans toute la France grâce à ses 2600 agences et ses 31 000 collaborateurs au service de 6,5 millions de clients et sociétaires (Chiffres 2023). Ce groupe a également une forte portée internationale puisqu'il est présent dans 9 pays étrangers dont l'Italie, la Croatie ou encore la Grèce avec un total de 5,5 millions de clients et sociétaires, mais aussi en Nouvelle-Calédonie via Groupama Gan Pacifique.

Au niveau organisationnelle, Groupama s'appuie sur l'interaction entre 3 catégories d'entités :

- **Les caisses locales** : elles constituent la base de l'organisation mutualiste de Groupama et permettent d'établir une véritable proximité avec les assurés. Les Caisses Locales se réassurent auprès des Caisses Régionales selon un mécanisme de réassurance spécifique par lequel Groupama Rhône Alpes Auvergne se substitue aux Caisses Locales de sa circonscription pour l'exécution de leurs engagements d'assurance à l'égard des sociétaires. Le réseau Groupama compte 2400 Caisses Locales, dont 359 chez Groupama Rhône-Alpes Auvergne.
- **Les caisses régionales** : ce sont des entreprises de réassurance sous le contrôle de l'organe central Groupama Assurances Mutuelles auprès duquel elles se réassurent, qui sont responsables de leur gestion, de leur politique tarifaire et de produits. Le réseau Groupama compte 11 Caisses Régionales d'Assurance et de Réassurance Mutuelles Agricoles (dont 9 métropolitaines et 2

d'outre-mer) et 2 caisses spécialisées.

- **Groupama Assurances Mutuelles** : l'organe central du Groupe est une caisse nationale de réassurance mutuelle agricole, forme de société d'assurance mutuelle qui pratique la réassurance et assure le pilotage opérationnel du Groupe et de ses filiales. Groupama Assurances Mutuelles est le réassureur des Caisses Régionales et l'organe central du réseau Groupama.

### Les caisses régionales Groupama

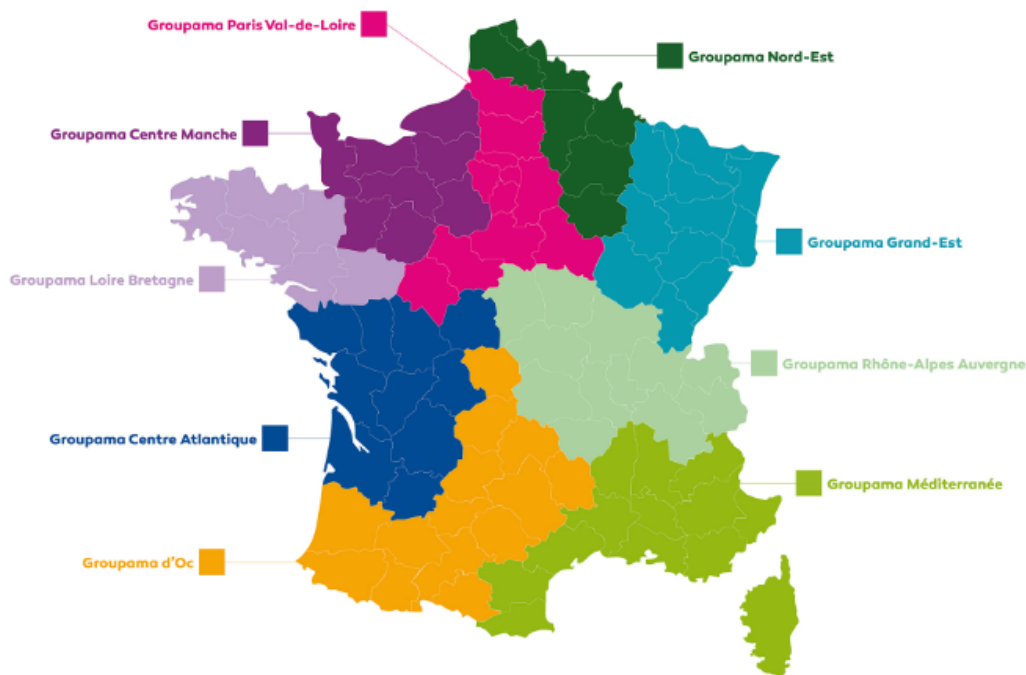


FIGURE 1.8 – Schéma des différentes caisses régionales de Groupama

#### L'entité Groupama Rhône-Alpes Auvergne

Groupama Rhône-Alpes Auvergne (GRAA) est donc l'une des caisses régionales du groupe Groupama. Elle est notamment celle qui a réalisée le plus gros chiffre d'affaire en 2022 et 2023.

Petite subtilité, les 12 départements sur lesquels interviennent GRAA ne sont pas les mêmes que ceux qui composent la région Auvergne Rhône-Alpes. En effet, les départements de l'ardèche (07), du Cantal (15) et de la Drôme (26) sont sous la responsabilité de Groupama Méditerranée (GMED) alors que les départements du Cher (18), de la Nièvre (58) et de la Saône-et-Loire (71), qui appartiennent d'un point de vue géographique à la région Bourgogne Franche-Comté, sont reliés à GRAA. Ce changement de tracé de territoire provient du fait que les départements ayant les mêmes profils de risque ont été rassemblés sous les mêmes entités.



FIGURE 1.9 – Départements de la région AURA

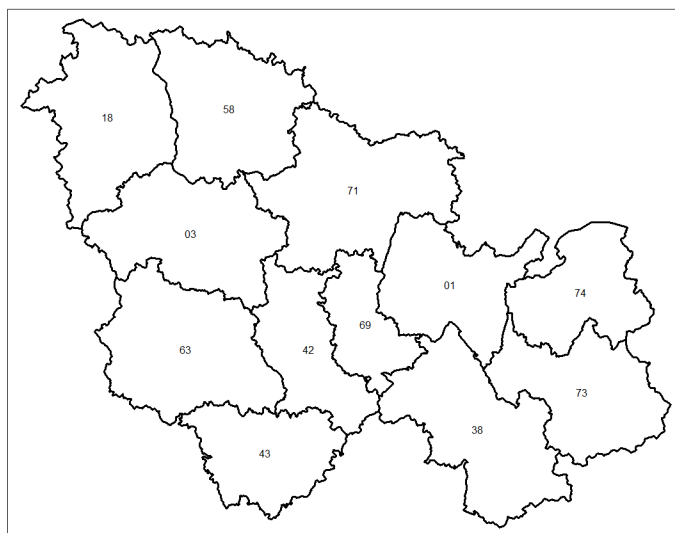


FIGURE 1.10 – Départements de l'entité GRAA

La Caisse régionale est une entreprise qui opère dans tous les métiers de l'assurance avec, notamment, un réseau commercial de salariés et des services de gestion. Elle définit et met en œuvre la politique générale pour son territoire, elle joue également un rôle de réassureur pour les caisses locales. Les administrateurs régionaux jouent un rôle prépondérant puisqu'ils se prononcent sur les orientations générales de la caisse régionale. La gestion opérationnelle est déléguée par le Conseil d'Administration au Directeur Général de la caisse régionale, assisté par un Directeur Général Délégué, tous deux assurant la direction effective de Groupama Rhône-Alpes Auvergne au titre de Dirigeants Effectifs.

### 1.3.2 Risque Tempête-Grêle

#### Impact sur le portefeuille

D'un point de vue actuariel, les différents types de risque sont regroupés selon la norme Solvabilité 2 en différentes Lines of Business (LoB). Ensuite chaque LoB est divisée en plusieurs segments de revue. Voici la liste des divers LoB ainsi que certains segments que l'on peut y retrouver :

## CHAPITRE 1. CONTEXTE

- 01\_MEDICAL : Santé
- 02\_INCOME : ap\_hors\_hors\_rente (assurance de personne hors rentes), gav (Garantie Accident de la Vie)
- 04\_MTPL : auto\_rc (Responsabilité civile automobile)
- 05\_MOTOR\_OTHER : auto\_dom (Dommage automobile)
- 06\_MAT : aviation, marine\_transport
- 07\_FIRE : cat\_ccr, dab (Dommage Aux Biens), tempete\_neige, serres, incendie, recoltes, mrc
- 08\_GTPL : rc (Responsabilité Civile), construction\_rcd (Responsabilité civile décennale en construction)
- 10\_LEGAL : pj (Protection Juridique)
- 11\_ASSISTANCE
- 12\_DIVERS
- 29\_DEPENDANCE

Ainsi, la garantie TGN est classée dans la LoB FIRE et dans les segments tempete\_neige et serres. Les travaux de ce mémoire porteront donc sur ces 2 segments.

Au niveau des cotisations, l'addition de ces segments de revue représente en fonction des années entre 3,5 et 5% du montant de cotisation totale généré par les segments modélisés. A titre de comparaison, le segment cat\_ccr, qui regroupe les risques de catastrophes naturelles réassurées par la Caisse Centrale de Réassurance, représente 3% du portefeuille globale des segments modélisés.

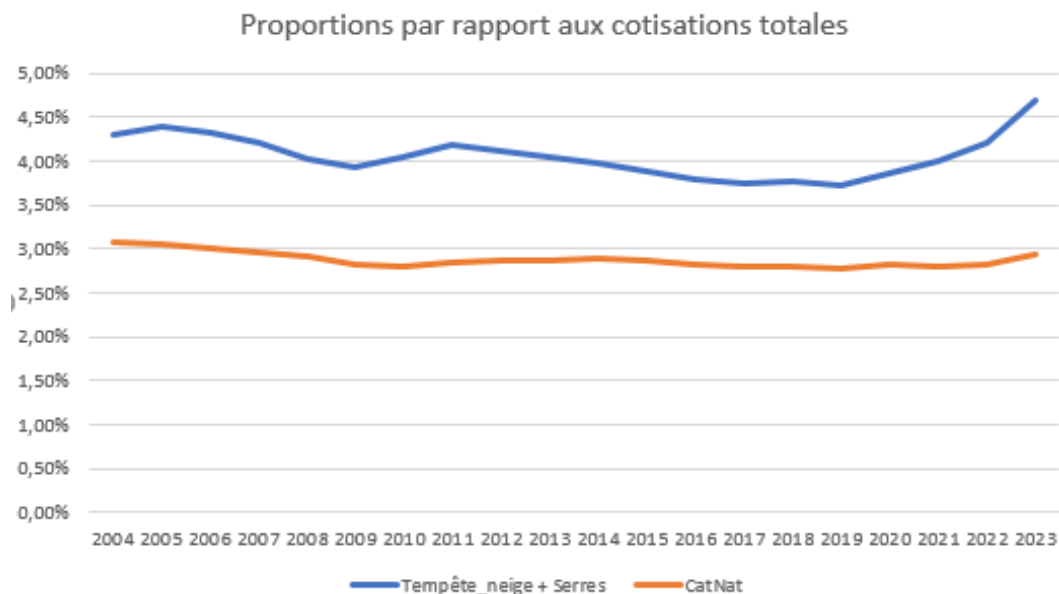


FIGURE 1.11 – Proportion des segments Tempêtes neige + Serres et Cat-CCR par rapport aux cotisations

On remarque sur ce graphique que l'exposition du portefeuille par rapport à ces risques reste stable sur les 20 dernières années avec seulement une légère augmentation de la proportion de cotisations destinées aux segments tempête\_neige et serres depuis 2022.

Au niveau de la rentabilité, le risque tempête grêle neige est très aléatoire, comme le montre ce graphique représentant le S/C comptable (c'est-à-dire la somme des sinistres divisées par la somme des cotisations) pour chaque année de survenance :

## CHAPITRE 1. CONTEXTE

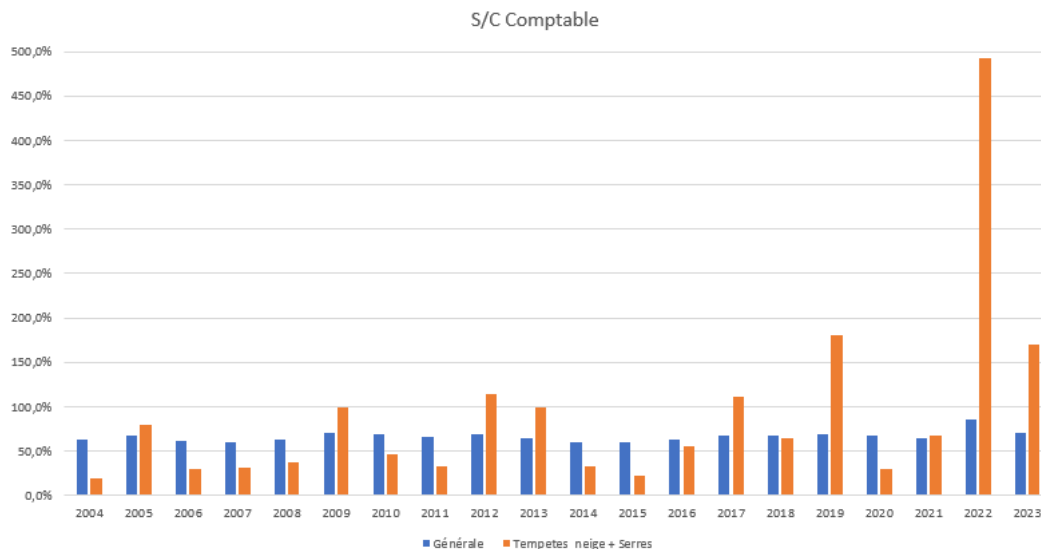


FIGURE 1.12 – S/C Comptables général et des segments Tempêtes Neige + Serres

En effet, on observe des valeurs de S/C extrêmement différents en fonction des années avec notamment une sinistralité sur la survenance 2022 exceptionnellement élevée, principalement à cause des orages de grêles survenus durant l'été qui ont fait d'énormes dégâts sur une grande partie du territoire de GRAA.

Cette analyse appuie une nouvelle fois l'importance de bien gérer ce risque difficilement prévisible qui, malgré la faible proportion qu'elle a dans le portefeuille de GRAA, peut impacter le S/C total puisque l'on voit que le S/C de la survenance 2022 est au-dessus de ceux des autres années.

### 1.3.3 Méthode d'estimation actuelle du coût d'une tempête

#### Distinction tempêtes mineures/majeures

Actuellement, GRAA s'appuie sur deux modèles différents pour évaluer la sinistralité d'une tempête, un pour les tempêtes jugées mineures et un pour les tempêtes jugées majeures.

Une tempête est considérée comme majeure lorsqu'au moins l'une des deux conditions suivantes est remplie :

Condition 1 : La caisse nationale de Groupama (GMA) déclare un « événement national » en lien avec un épisode climatique intense (ex. : vents violents, pluies intenses, chutes de grêles) sur une période courte, généralement de 1 à 3 jours. Cette déclaration se fonde sur le recensement d'au minimum 3M€ (seuil de l'exercice 2025) de sinistres tempête ou grêle sur ces 72h.

Condition 2 : Le montant cumulé des sinistres déclarés, sur le territoire de Groupama Rhône-Alpes Auvergne, sur une période de 72h dépasse un seuil de sinistres qui s'élève pour l'année 2025 à 310k€. Dépassé ce seuil, un événement Tempête régional est alors déclaré.

Lorsque l'un de ces critères est rempli, l'événement est formellement reconnu comme une tempête majeure. Cela permet de traiter cet épisode de manière distincte dans les modèles d'estimation de charge sinistre et d'enclencher les mécanismes de couverture financière associés.

### Tempêtes majeures

Ainsi, lorsque qu'une tempête jugée majeure est déclarée, leur coût est calculé à partir des premières estimations réalisées par des experts sur le terrain combinées aux cadences de règlement des grosses tempêtes historiques.

En effet, à chaque mois d'inventaire, le taux d'avancement de la charge sinistre sur la même période que celle séparant la date de survenance de la tempête modélisée et la date inventaire est calculée pour les tempêtes historiques et appliquée à la tempête dont l'on souhaite estimer la sinistralité. Ce processus est appliqué pour chaque métier (Agricole, Habitation, Entreprises, Collectivités ...).

Le procédé est résumé dans le schéma ci-dessous :

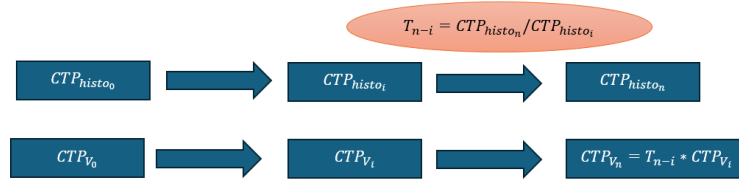


FIGURE 1.13 – Schéma du processus actuel d'estimation de la charge sinistre d'une tempête

- $n$  représente ici le nombre de mois nécessaires pour que tous les sinistres liés à la tempête soient clos. La charge sinistre ne pourra donc plus évoluer au-delà.
- $CTP_{histo_i}$  et  $CTP_{V_i}$  représentent la charge sinistre observée  $i$  mois après la survenance respectivement des tempêtes historiques et de la tempête à modéliser.

Ce type de modèle permet d'obtenir une prédiction de la charge sinistre globale de la tempête à partir des données historiques mais ne prends pas en compte deux éléments importants :

- Le portefeuille d'assurés listant ceux qui ont été exposés à la tempête qui a évolué entre la date de survenance des tempêtes historiques et celle des tempêtes récentes.
- Les caractéristiques météorologiques propres à la tempête survenue qui peuvent entraîner un allongement ou une réduction de la durée de prise en charge des sinistres comparée à celle des tempêtes historiques.

C'est pourquoi une méthode alternative va être développée dans la suite de l'étude, dans le but de proposer une amélioration du processus d'estimation de la charge sinistre d'une tempête majeure.

#### 1.3.4 Autre méthode proposée : Fréquence x Coût Moyen

Pour pallier aux problèmes énoncés précédemment, une autre méthode d'estimation de la charge sinistre d'une tempête majeure après que celle-ci soit survenue sera développée, ce processus consiste à associer à chaque assuré deux valeurs :

- Sa probabilité d'être sinistré suite à la survenance de la tempête.
- S'il déclare un sinistre, le coût moyen de celui-ci.

Pour cela, un modèle de fréquence de sinistre est d'abord construit à partir des variables propres à chaque assuré impactant la probabilité qu'il déclare ou non un sinistre. Ensuite, le coût moyen d'un sinistre pour chaque catégorie d'assuré est calculé en s'appuyant sur les variables qui ont une influence sur le coût des sinistres.

## CHAPITRE 1. CONTEXTE

Ainsi, cette méthode permettra d'obtenir une base de données comme suit :

Numéro d'assuré	Probabilité de déclaration d'un sinistre	Coût Moyen d'un sinistre	Charge prédite
$A_1$	$P_1$	$C_1$	$P_1 * C_1$
$A_2$	$P_2$	$C_2$	$P_2 * C_2$
...	...	...	...
$A_{n-1}$	$P_{n-1}$	$C_{n-1}$	$P_{n-1} * C_{n-1}$
$A_n$	$P_n$	$C_n$	$P_n * C_n$

TABLE 1.1 – Présentation de la méthode Frequence x Coût Moyen

L'estimation globale du coût de de la tempête s'obtiendra en ajoutant les charges prédites pour chaque assuré de la table.



## Chapitre 2

# Modélisation de la charge sinistre d'une tempête majeure sur le portefeuille de GRAA

### 2.1 Construction de la base de données

La première étape pour réaliser notre modélisation de la sinistralité lors de la survenance d'une tempête, et sans doute la plus importante, est la constitution de notre base de donnée. Cette base est le regroupement de la base météorologique, qui a été présentée dans la partie sur le module aléa, des bases portefeuille et sinistres, qui constitueront le module enjeu présenté lors de la partie I.

#### 2.1.1 Données météorologiques

Les données météorologiques sont évidemment des éléments indispensables à analyser lorsque l'on veut prédire la charge sinistre d'une tempête. Le but de ce module est donc de récupérer dans des bases d'historiques de données météo, les variables qui pourraient expliquer l'augmentation de la probabilité de sinistralité et la sévérité lors de la survenance d'une tempête.

##### Base de vent

Les données utilisés pour cette base ont été extraites du site [www.météo.data.gouv.fr](http://www.météo.data.gouv.fr), qui propose des relevés météorologiques quotidiennes en libre accès provenant de divers stations réparties dans toutes la France. Celles qui nous intéressent pour ce mémoire sont évidemment les stations situées sur le territoire de GRAA.

De chaque stations a donc été extrait pour chaque jour depuis 2004 :

- Site\_meteo : le nom de la commune où se situe le site
- Longitude : La longitude de la localisation du site
- Latitude : La latitude de la localisation du site
- Altitude : L'altitude de la localisation du site
- Date : Date de relevé des données (en AAAAMMJJ)
- CODE\_POSTAL : le code postal dans lequel se trouve le site

## CHAPITRE 2. MODÉLISATION DE LA CHARGE SINISTRE D'UNE TEMPÊTE MAJEURE SUR LE PORTEFEUILLE DE GRAA

- Qté\_Précip : Quantité de précipitations tombés en 24h (en mm)
- Temp\_min : Température minimale sous abri (en °C)
- Temp\_max : Température maximale sous abri (en °C)
- Vent\_moy : Moyenne quotidienne de la vitesse du vent moyenne sur 10 minutes (en m/s) à 10m de la surface du sol
- Vent\_max : Maximum quotidien de la vitesse maximale horaire du vent instantanée (en m/s) à 10m de la surface du sol

Au vue des nombreuses cellules vides que comportaient certaines stations, un travail de sélection des stations comportant l'ensemble des données nécessaires pour notre étude a été réalisé, voici la localisation de chacune des stations météo retenues :

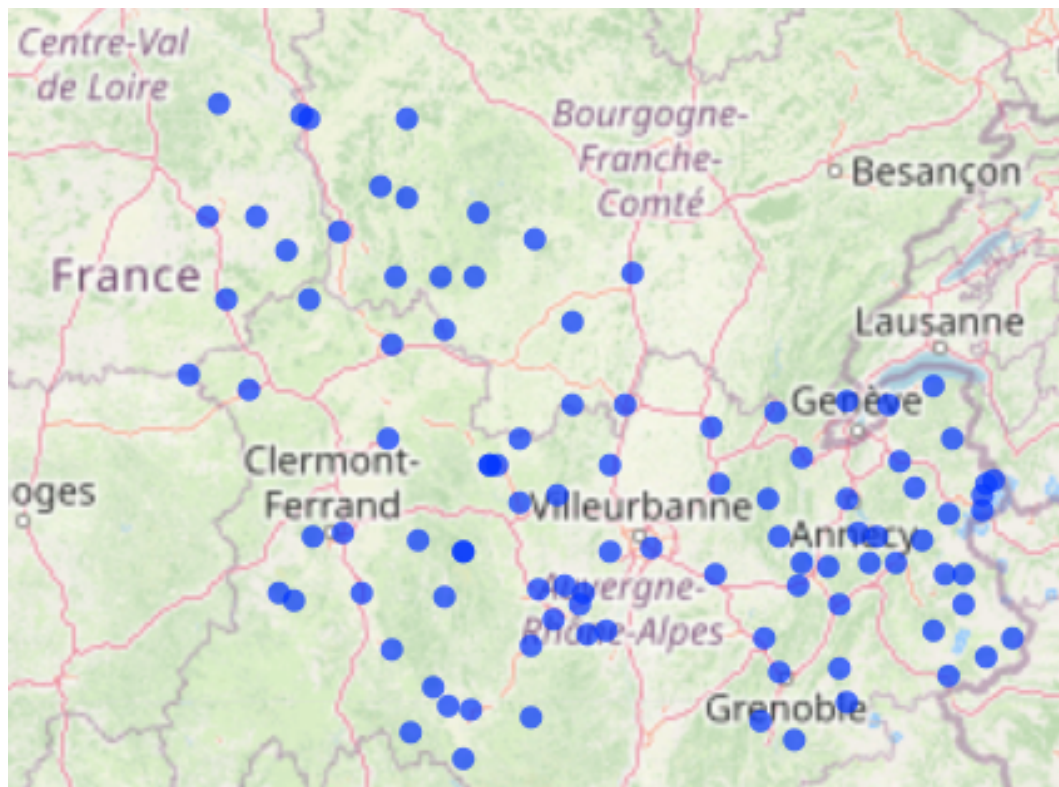


FIGURE 2.1 – Positions des différents sites météorologiques

Le seul autre changement fait sur la base originelle est la modification de l'unité de mesure du vent en km/h (en multipliant par 3,6 les valeurs en m/s).

### Base de grêles

Etant donné qu'aucune variable ne se rapporte à la grêle dans les données que propose [www.météo.data.gouv.fr](http://www.météo.data.gouv.fr), une autre source sera utilisée pour obtenir ces données.

Il n'existe aucune base de données en libre service qui offre des données suffisamment qualitatives pour être interprétées, la solution est donc de se tourner vers les bases de données payantes que proposent certains observatoires. Et il se trouve que le groupe Groupama possède des bases de données qu'elle achète à l'observatoire français Keraunos dans le cadre de leurs études sur les risques climatiques. KERAUNOS, également connu sous le nom d'Observatoire Français des Tornades et des Orages Violents, est le premier bureau d'études en France spécialisé dans la prévision et la gestion des risques liés aux phénomènes orageux. Il est une référence dans le domaine de la prévision, du suivi et de

## CHAPITRE 2. MODÉLISATION DE LA CHARGE SINISTRE D'UNE TEMPÊTE MAJEURE SUR LE PORTEFEUILLE DE GRAA

l'étude de la grêle, de la foudre et des pluies diluviennes.

L'une de ces bases sera particulièrement utile puisqu'elle recense divers informations sur la taille des grêlons tombés quotidiennement par commune sur l'ensemble du territoire français depuis 2006. Voici donc la liste des données mis à disposition :

- Date : Date de relevé des données (en AAAAMMJJ)
- Commune : D'où ont été relevées les données
- Code INSEE
- Département
- Valeur basse estimée : Estimation basse de la taille du diamètre des grêlons (en cm)
- Valeur haute estimée : Estimation haute de la taille du diamètre des grêlons (en cm)
- Classe : Répartition en 6 classes (A0,A1,A2,A3,A4,A5), le chiffre correspondant à l'arrondi à l'entier inférieur de la Valeur haute estimée
- Gros\_grelons : Variable binaire qui prend la valeur 0 si la Valeur haute estimée est inférieure ou égale à 2cm et 1 sinon.

Les 2 valeurs que nous retiendrons pour analyser les tombés de grêles seront les variables Classe et Gros\_grelons.

### 2.1.2 Portefeuille d'assurés

La première base au niveau des données de GRAA qui est extraite est la base portefeuille. On cherche ici à connaître quels sont les différents types d'assurés de GRAA et notamment les critères qui feront qu'il seront plus ou moins touchés lorsqu'une tempête interviendra. La base mise à disposition par GRAA dans le cadre de cette étude concerne les assurés possédant un contrat habitation, nos modélisations futures se feront donc pour ce type d'assurés.

#### Base initiale

Pour obtenir notre base portefeuille, une première table est extraite. Elle regroupe tous les contrats de GRAA assurant les habitations entre 2012 et 2025. Voici les principales variables qui la compose :

- Num\_Contrat : Le numéro de contrat correspondant à l'assuré
- Annee\_Sous : L'année de souscription du contrat
- CODE\_POSTAL : le code postal dans lequel se trouve l'habitation assuré
- CODE\_INSEE : Code associé à la commune dans laquelle est située l'habitation assurée
- Cotisation : La cotisation payée durant l'année par l'assuré pour l'assurance de l'habitation.
- Nature\_Hab : La nature de l'habitation (maison, appartement, chalet, mobil-home, ...)
- Nature\_assure : La nature de l'assuré (propriétaire, locataire, colocataire, usufruitier, ...)
- periode\_const.bat : la période de construction du bâtiment assuré. Elle comporte 5 catégories (avant 1915, de 1916 à 1950, de 1951 à 1975, de 1976 à 2000, après 2000)
- usage\_hab : l'usage de l'habitation (Principale ou secondaire)
- nb\_pieces : Le nombre de pièces que comporte l'habitation
- formule : Le niveau de garantie choisi par l'assuré (Essentiel, Confort 1, Confort 2), la formule Confort 2 étant celle offrant la couverture la plus complète.
- ZONE : L'urbanisation de la ville où se trouve l'habitation (NON URBAIN, PERI-URBAIN, URBAIN)

## CHAPITRE 2. MODÉLISATION DE LA CHARGE SINISTRE D'UNE TEMPÊTE MAJEURE SUR LE PORTEFEUILLE DE GRAA

### Analyses et modifications apportées

#### — Variable `Nature_Hab`

Pour des raisons de simplification, les différentes natures d'habitations ont été regroupées en 3 groupes :

- MAISON = Maison + Monobloc + Dépendance fermée + Dépendance ouverte + Demeure historique
- APPARTEMENT = Appartement
- AUTRE = Autre + Mobil-Home

Vérifions si cette variable est utilisable avec sa répartition dans le portefeuille :

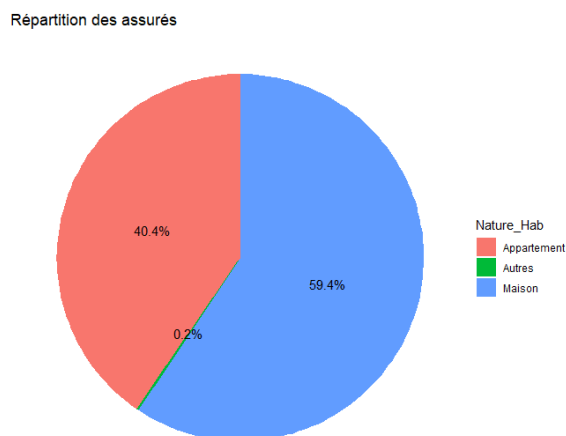


FIGURE 2.2 – Répartition des différentes nature d'habitation

En retirant les habitations de la catégorie "Autre", on obtient une répartition équilibrée de cette variable donc ces catégories sont conservées.

#### — Variable `Nature_assure`

Les différentes natures d'assurés ont également été regroupés en 3 groupes :

- Propriétaire = Propriétaire + Copropriétaire
- Locataire = Locataire + Colocataire
- Bailleur = Bailleur + Usufruitier

Vérifions si cette variable est utilisable avec la proportion de chacune de ces valeurs dans le portefeuille :

## CHAPITRE 2. MODÉLISATION DE LA CHARGE SINISTRE D'UNE TEMPÊTE MAJEURE SUR LE PORTEFEUILLE DE GRAA

Répartition des assurés

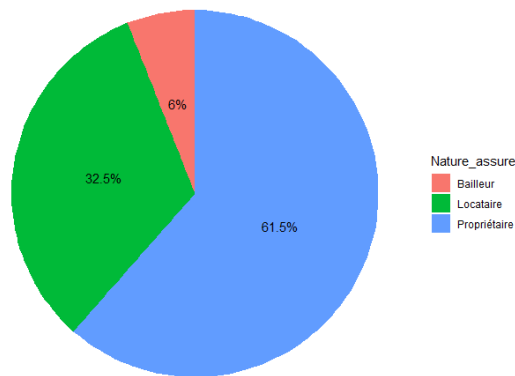


FIGURE 2.3 – Répartition des différentes natures d'assuré

On obtient une répartition plus déséquilibrée, en particulier à cause de la forte présence de l'élément "Propriétaire". Pour pallier à ce problème, une nouvelle variable regroupant les deux précédentes va être formée.

- **Variable assure\_habitation** Cette variable créée consiste en la combinaison des variables Nature\_assure et Nature\_Hab en 4 catégories :
  - Propriétaire Maison
  - Propriétaire Appartement
  - Locataire
  - Bailleur

Vérifions sa répartition dans le portefeuille :

Répartition des assurés

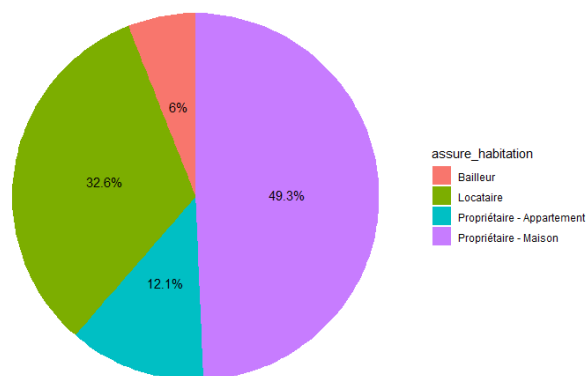


FIGURE 2.4 – Répartition des différents assurés et habitations

Malgré la proportion de la valeur "Bailleur" qui reste légère, cette nouvelle variable est utilisable et convient à la description de la nature des assurés et de leur habitation.

## CHAPITRE 2. MODÉLISATION DE LA CHARGE SINISTRE D'UNE TEMPÊTE MAJEURE SUR LE PORTEFEUILLE DE GRAA

### 2.1.3 Base sinistres

Afin de pouvoir entraîner et appliquer nos modèles de fréquence et coût moyen, une phase de recherche de tempêtes de vent et de grêle sur lesquelles nous allons pouvoir nous appuyer doit être entreprise. Pour cela une première base générale sur sera d'abord formée, une extraction des plus grosses tempêtes sera ensuite effectuée sur cette base.

#### Base initiale

La base à partir de laquelle la base sinistres va être construite recense tous les sinistres sur bâtiments des segments de revue qui nous intéressent sur un historique de 14 ans, c'est-à-dire sur les survenances de 2012 à 2025. En voici les variables qui nous seront utiles :

- ADS : Numéro de l'Avis De Sinistre, qui regroupe plusieurs sinistres survenus sur le même intervalle de temps (maximum 2 jours) et qui concernent des sinistres du même segment de revue (auto, habitation, incendie, catnat). Ces sinistres sont donc considérés faisant parti du même événement climatique.
- IDSin : Numéro d'identifiant sinistre
- Num\_Contrat : Le numéro de contrat de l'assuré sinistré
- Annee\_Surv : L'année de survenance du sinistre
- Date\_Surv : Date de survenance du sinistre sous la forme AAAAMMMJJ
- CODE\_POSTAL : le code postal dans lequel se trouve le bâtiment assuré
- CODE\_INSEE : Code associé à la commune dans laquelle est située le sinistre
- CTP : Coût du sinistre à la date d'inventaire
- Descr\_sinistre : Description sur la nature du sinistre (exemples : "TEMPETE GRELE SUR TOITURE BATIMENT AGRICOLE HANGAR", "VENT ENDOMMAGE CREMAILLERE DE LA SERRE")
- Metier : Une variable catégorielle qui la topologie du bâtiment assuré qui comporte plusieurs valeurs différentes (Habitation, Associations, Collectivité, Entreprises ...)

#### Analyses et Modifications apportées

Certains critères de cette base ont été modifiés :

##### — Variable Département

Un premier travail d'analyse de la variable département a été réalisé. En effet, certains sinistres se situent dans des départements extérieurs à GRAA, cela peut être dû par exemple à une résidence secondaire appartenant à un assuré domicilié dans un territoire de GRAA ou encore un assuré du portefeuille courtage qui lui s'étend sur toute la France. Les sinistres de cette catégories ont été retirés de la base de données, ce qui l'a très peu impactée puisqu'ils représentaient moins de 1% du portefeuille et des sinistres.

##### — Variable CTP

Certains sinistres ont un montant de 0 car les avis de sinistres étant créés 1 ou 2 jours après la survenance, le sinistre a pu être annulé ou non indemnisé par la suite, le montant de CTP pour ce sinistre est alors remis à 0. Ce type de sinistre a donc également été retiré de la base

## CHAPITRE 2. MODÉLISATION DE LA CHARGE SINISTRE D'UNE TEMPÊTE MAJEURE SUR LE PORTEFEUILLE DE GRAA

de données pour éviter que le nombre de sinistres ne soit impacté.

### — Variable Métier

Puisque la base portefeuille nous avons construite se réfère aux contrats habitations, un filtre sur cette modalité a été nécessaire.

### — Variable Survenance

L'étude menée se base sur le montant total des sinistres en fonction du vent. La question qui s'est posée a donc été de déterminer à partir de quelle année de survenance on peut considérer les sinistres comme clos, c'est-à-dire que le montant à date d'inventaire est le montant définitif. Pour cela, l'inventaire mensuel du montant de charges pour chaque ADS présent dans la base a été extrait. Voici par exemple l'évolution de la charge sinistre de l'ADS numéro "A692022010A" caractérisant un événement de la plus grosse tempête recensée dans notre historique qui est survenu en juin 2022 :

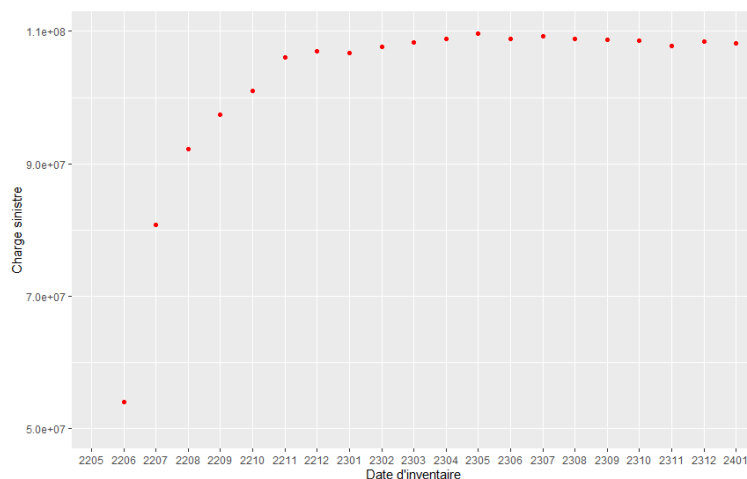
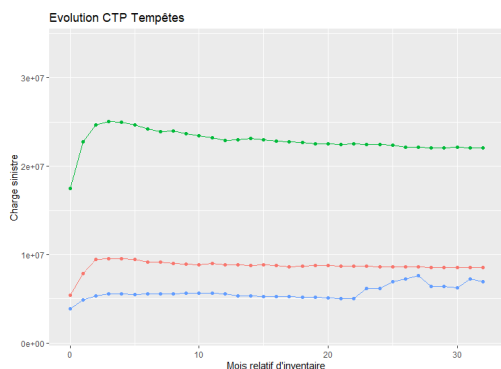


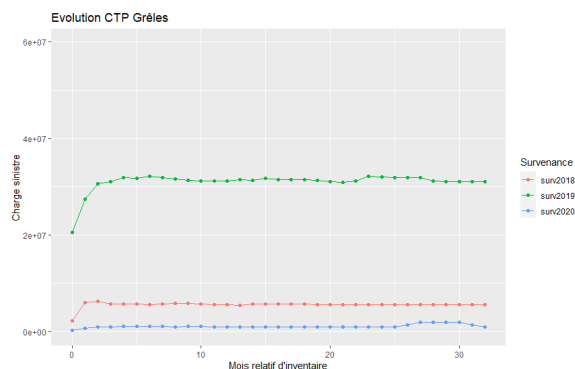
FIGURE 2.5 – Evolution de la charge sinistre pour un ADS

On observe sur ce graphe que le montant de charge sinistre pour cet ADS a doublé sur les 6 premiers mois et qu'il a stagné depuis. Pour vérifier cette idée en prenant en compte l'ensemble des ADS de la base sinistre, regardons l'évolution de la charge sinistre en fonction du mois relatif d'inventaire, c'est-à-dire le nombre de mois après lequel les ADS sont survenus, pour 3 années de survenance :

## CHAPITRE 2. MODÉLISATION DE LA CHARGE SINISTRE D'UNE TEMPÊTE MAJEURE SUR LE PORTEFEUILLE DE GRAA



(a) Evolution de la charge sinistre selon le mois d'inventaire relatif pour les tempêtes



(b) Evolution de la charge sinistre selon le mois d'inventaire relatif pour les grêles

On remarque que ce soit pour le risque Tempête ou le risque Grêle, on peut définir une période au alentours de 10 mois pour laquelle la charge sinistre sera quasiment définitive.

Le choix a donc été de retirer de la base sinistres les événements survenus après mai 2024 qui peuvent, à la date d'inventaire de 2025, ne pas être clos.

Certains critères ont été ajoutés à cette base de données :

### — Type\_Sinistre

Un problème a été rencontré concernant la distinction entre les tempêtes dues au vent et celles dues à la grêle. En effet, la variable Descr\_sinistre pouvait être un élément sur lequel s'appuyer pour faire cette différenciation mais la description n'étant pas formatée, puisqu'elle est rentrée à la main par le gestionnaire sinistre, cette donnée n'était pas exploitable. La distinction entre les différents phénomènes climatiques s'est donc fait en fonction de la saisonnalité. Voici un exemple des deux variables relatives au vent dans notre base météo pour la station située à Ambérieu dans l'Ain (01) sur la période 2019-2022 :

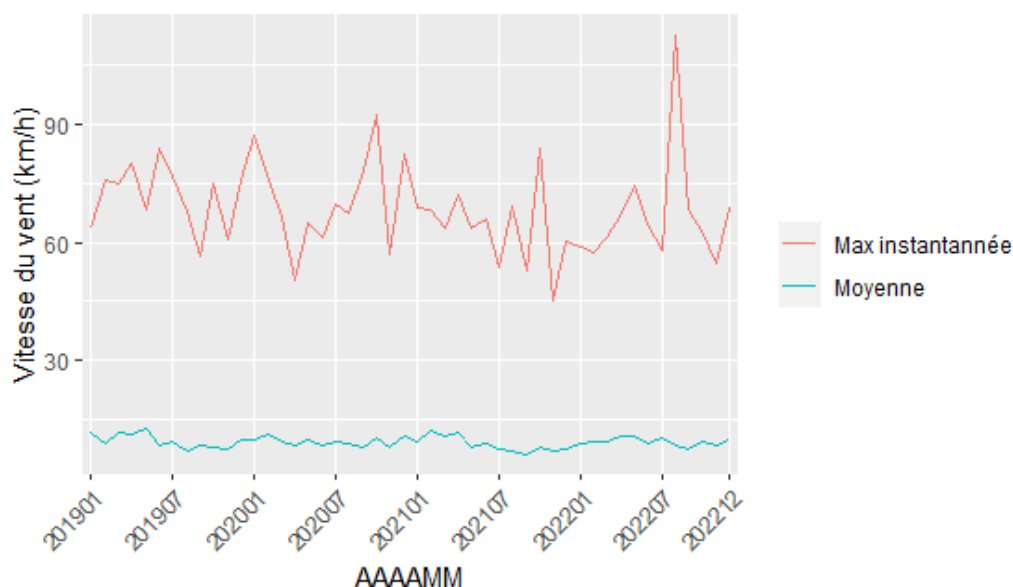


FIGURE 2.7 – Historique de vent de la station AMBERIEU

Par souci de simplification du graphique, les données quotidiennes ont été remplacées par



## CHAPITRE 2. MODÉLISATION DE LA CHARGE SINISTRE D'UNE TEMPÊTE MAJEURE SUR LE PORTEFEUILLE DE GRAA

les données mensuelles (en faisant la moyenne sur le mois des moyennes quotidiennes extraites pour cette station et le maximum sur le mois des maximum quotidiens).

On remarque que les pics de vitesses maximales de vent (correspondant donc à des tempêtes de vent) sont principalement situés entre la fin d'année et le début de l'année. On peut donc en déduire que le reste de l'année correspond aux sinistres pour lesquels le vent a moins d'importance, les grêles.

Voyons si l'on observe ce même phénomène pour la grêle à l'aide de notre base récoltée précédemment :

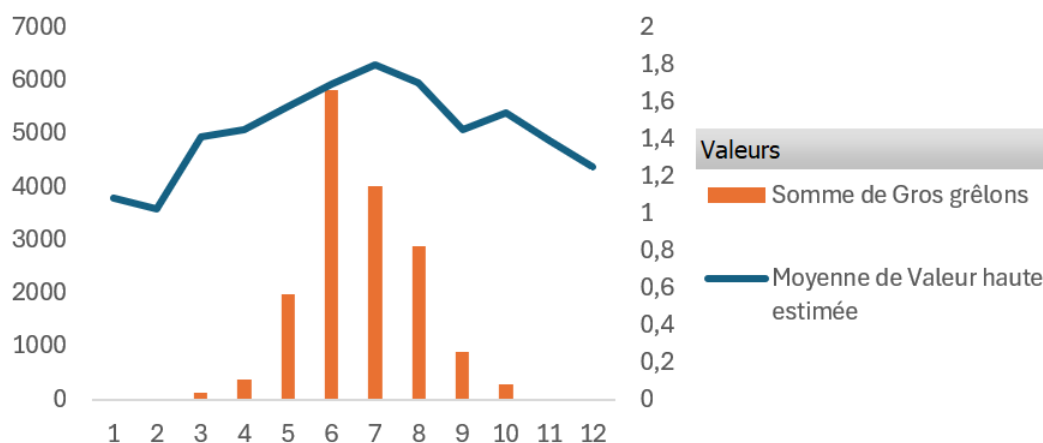


FIGURE 2.8 – Saisonnalités des tombées de grêles

On observe bien une nette augmentation de la taille moyenne des grêlons durant la période juin-juillet-août et les grêlons recensés entre novembre et février ne sont pour la plupart pas définis pour "Gros grêlons", qui sont pour rappel les grêlons dont la Valeur haute estimée est supérieure ou égale à 2,5cm.

Le choix de simplification fait au final est donc le suivant : pour les sinistres survenus entre les mois d'octobre et d'avril, le type de sinistre considéré sera la VENT, et pour le reste de l'année la variable prendra la valeur GRELE.

### Sélection des tempêtes de référence

L'historique de sinistres de GRAA comporte 2 grosses tempêtes de grêles survenues les 3 et 21 juin 2022 (avec des charges sinistres de respectivement 46 et 18M€ en habitation). La plus importante en nombre et en montant de sinistres étant celle du 3 juin, elle servira donc à entraîner et tester nos modèles de fréquence et de coût moyen. Par mesure de simplification, elle sera notée tempête V1 dans la suite de l'étude.

Concernant les tempêtes de vent, GRAA n'a pas enregistré de sinistralités d'ampleur similaire à celles des tempêtes de grêle. Le processus sera donc reproduit avec une tempête que l'on notera V1 de charges sinistres de 1,8M€ en habitation.

## CHAPITRE 2. MODÉLISATION DE LA CHARGE SINISTRE D'UNE TEMPÊTE MAJEURE SUR LE PORTEFEUILLE DE GRAA

### 2.1.4 Base finale

La base finale consiste donc en un rassemblement des divers informations obtenues dans les différentes bases construites.

#### Association base sinistre et base météorologique

En partant de la base sinistres, les données météo peuvent désormais être ajoutées pour apporter des précisions quant aux circonstances de ces sinistres.

Le problème rencontré est qu'il n'y avait pas de station météo dans chacune des communes dans lesquelles un sinistre est survenu. Il a donc fallu associer à chaque commune sinistrée la station météo la plus proche.

Pour cela, le logiciel R a été utilisé est notamment la fonction "distHaversine" que l'on peut obtenir en installant le package "geosphere". Cette fonction renvoie, à partir de 2 longitudes et latitudes, la distance entre ces 2 lieux. Il a donc suffi d'obtenir les coordonnées GPS des communes sinistrées (puisque celles des sites météo étaient déjà connues) pour calculer la distance d'une commune avec chaque site météorologique et de retenir la plus faible, et ce pour chaque commune sinistrée.

Ainsi, on obtient une base qui donne pour tous les sinistres de notre base :

- Les variables données dans la base sinistre
- Les données météorologiques recensées au site météo le plus proche de la commune sinistrée le jour de survenance du sinistre

#### Association avec la base portefeuille

Les bases sinistres-météo et portefeuille sont désormais prêtes à être intégrées à un modèle de prédiction de la sinistralité lors de la survenance d'une tempête ou d'un événement de grêle sur habitation. Leur regroupement est effectué à l'aide du numéro de contrat, présent à la fois dans la base portefeuille et la base sinistres.

#### Dernière modification : Détermination du lieu des tempêtes

Un dernier problème reste à être détaillé : quels sont les assurés qui ont été exposés aux tempêtes ? En effet 3 choix différents pouvaient être faits :

1. Considérer que tous les détenteurs d'un contrat habitation à GRAA au moment de la survenance de la tempête ont été exposés.
2. Considérer que seules les communes dont au moins un assuré a été sinistré ont été exposées.
3. Considérer qu'une commune a été exposée à partir du moment où un certain pourcentage de ses assurés ont subis un sinistre.

La première proposition n'a pas été retenue car de nombreux codes postaux, voire de départements, n'ont pas reçu de dégâts liés à ces tempêtes, les statistiques sur le pourcentage de sinistrés provoqués par les tempêtes auraient donc été faussés.

La deuxième proposition permet de mieux localiser les zones d'apparition des tempêtes. Cependant,

## CHAPITRE 2. MODÉLISATION DE LA CHARGE SINISTRE D'UNE TEMPÊTE MAJEURE SUR LE PORTEFEUILLE DE GRAA

en appliquant cette méthode, le nombre d'assurés exposés devenait beaucoup trop important comparé au nombre de sinistrés. En effet on obtenait par exemple 150 000 assurés exposés pour la tempête G1 dont 3600 sinistrés, ce qui représente un taux de 2,4%, un ratio bien trop faible pour pouvoir produire un modèle de fréquence de sinistres avec des résultats satisfaisants. Ceci vient du fait que certains numéro insee de communes comportaient tous leurs assurés dans la base alors qu'un ou deux étaient sinistrés.

La troisième proposition permet de pallier cette disproportion en instaurant un taux de sinistralité à partir duquel on peut considérer que la commune a réellement subi la tempête. Il ne fallait en revanche pas appliquer un taux trop important pour garder un maximum de données dans notre base. En choisissant un seuil de 5% de sinistrés, les proportions de sinistrés pour les tempêtes G1 et V1 sont montés à respectivement 15,1% 18,6%, alors que leur charge de sinistres totale n'ont diminuées que de 3,4% et 6,9%.

En résumé, voici les informations sur les tempêtes retenues et qui formeront nos 2 bases de données nécessaires pour la suite de notre étude :

Phénomène	Date	Notation	Nombre d'assurés exposés	Nombre de sinistres	Montant de sinistres
Grêle	3 juin 2022	G1	20 212	3055	44,6M€
Vent	21-23 décembre 2019	V1	4409	812	1,6M€

TABLE 2.1 – Présentation des tempêtes retenues

## 2.2 Modélisation de la fréquence de sinistres

Le premier modèle qui sera développé aura pour but de déterminer le nombre de sinistres provoqué par une tempête de vent ou de grêle. Pour cela, on cherche pour chaque assuré de notre base de données sa probabilité d'avoir un sinistre une fois que les données liées à la tempête sont connues. Trois différents types de modèles probabilistes seront donc testés : le modèle de régression linéaire, le modèle d'arbre de décision ainsi que le modèle random forest.

### 2.2.1 Recherche de critères discriminants

#### La force et la vitesse du vent

Lorsque l'on veut étudier la probabilité de survenance ou la sévérité d'une tempête, la première donnée que l'on pense à regarder est évidemment la force du vent, voici comment elle est calculée :

$$F = \frac{1}{2} \times p \times V^2 \times A$$

où :

- $F$  est la force exercée par le vent (en Newtons,  $N$ ).
- $p$  est la densité de l'air (en kilogrammes par mètre cube,  $\text{kg/m}^3$ ).
- $V$  est la vitesse du vent (en mètres par seconde,  $\text{m/s}$ ).
- $A$  est la surface de la zone sur laquelle le vent agit (en mètres carrés,  $\text{m}^2$ ).

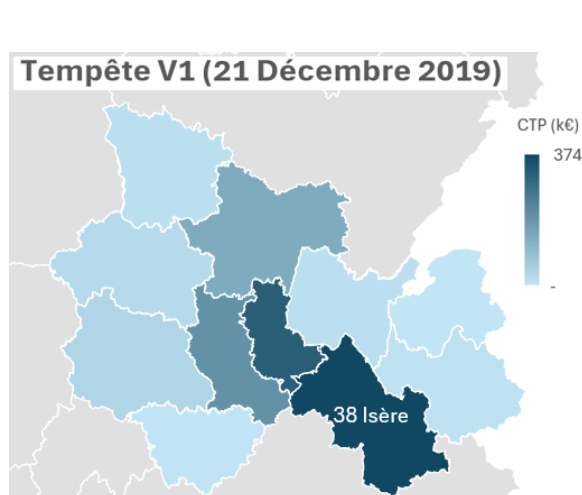
Dans notre cas, la densité pourrait être récupérée mais la surface de la zone sur laquelle le vent agit est en revanche beaucoup plus compliquée à déterminer. C'est pourquoi la variable relative au vent qui est choisie est la vitesse.

## CHAPITRE 2. MODÉLISATION DE LA CHARGE SINISTRE D'UNE TEMPÊTE MAJEURE SUR LE PORTEFEUILLE DE GRAA

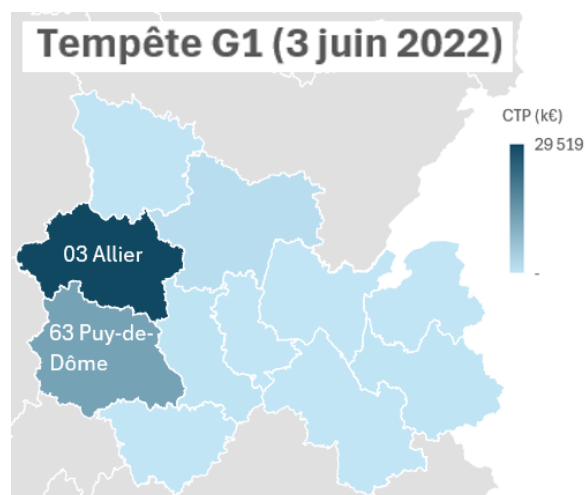
Dans la base de données extraite, deux indicateurs sont en rapport à la vitesse du vent : la moyenne quotidienne de la vitesse du vent moyenne sur 10 minutes et le maximum quotidien de la vitesse de vent instantanée.

Comme expliqué dans la première partie, les relevés utilisés par les société d'assurance pour déterminer l'importance d'une tempête est la vitesse des rafales de vent, qui correspond à la moyenne de la vitesse de vent sur une période de 2 minutes avant l'enregistrement. Parmi les variables en notre possession, la vitesse maximale instantannée est celle qui se rapproche le plus de cette définition, cette donnée sera donc désormais utilisée pour caractériser l'intensité du vent.

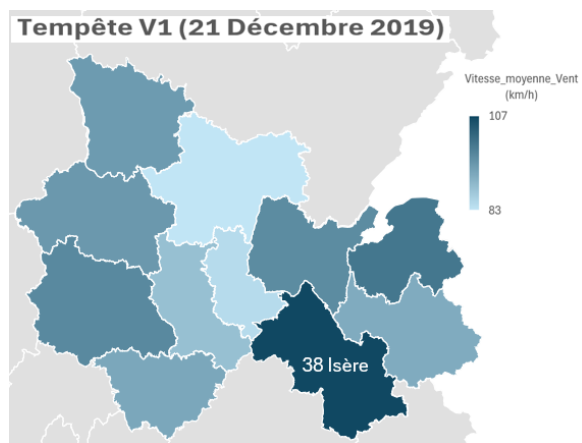
Vérifions si cette variable a en effet un impact sur la sinistralité au niveau de nos bases d'assurés :



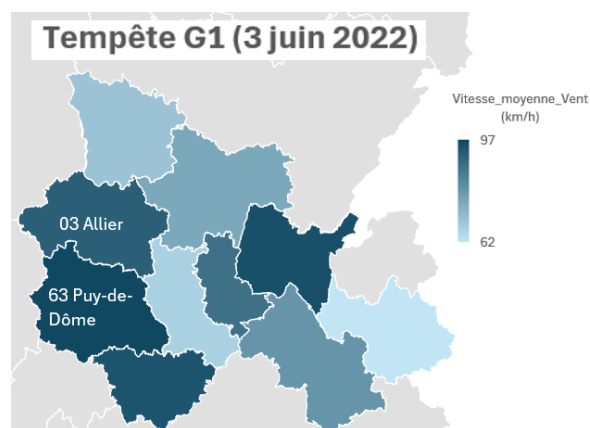
(a) Charge de sinistres par département pour la tempête V1



(b) Charge de sinistres par département pour la tempête G1



(a) Moyenne de la variable Vitesse\_vent par département pour la tempête V1



(b) Moyenne de la variable Vitesse\_vent par département pour la tempête G1

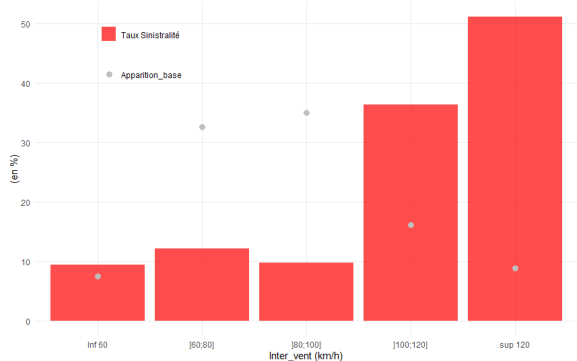
Les graphiques ci-dessus montrent que pour notre tempête V1 que l'on a considéré principalement lié au vent, cette variable a en effet un impact puisque le département le plus touché en terme de montant de sinistres (Isère) est également celui dont les rafales de vent les plus importantes en moyenne ont été enregistrées.

Cette observation s'applique dans une moindre mesure à la tempête G1 pour laquelle les départements de l'Allier et du Puy-de-Dôme, dans lesquels elle s'est concentrée, font partis des départements avec

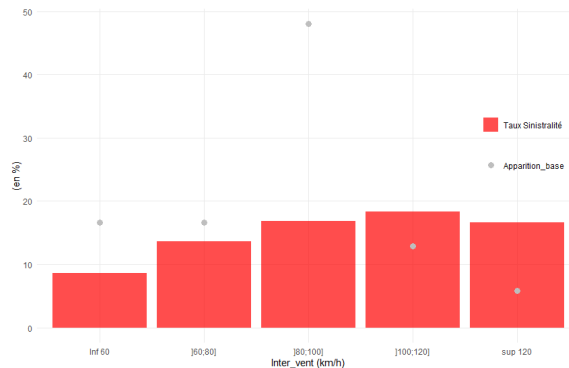
## CHAPITRE 2. MODÉLISATION DE LA CHARGE SINISTRE D'UNE TEMPÊTE MAJEURE SUR LE PORTEFEUILLE DE GRAA

la plus haute vitesse moyenne de vent mais d'autres départements ont subis les mêmes conditions de vent et ont pourtant moins été impactés par la tempête.

Vérifions désormais si l'incidence du vent sur le nombre de sinistres dans nos bases d'assurés :



(a) Proportion de nombre de sinistres par vitesse de vent pour la Tempête V1



(b) Proportion de nombre de sinistres par vitesse de vent pour la Tempête G1

Par mesure de simplification de la présentation des graphiques, une variable nommée *Inter\_vent* a été créée et qui classe les valeurs de *Vitesse\_vent* en 5 intervalles. A noter que les barres en rouge sur ce graphique représentent les taux, parmi les assurés exposés à des vitesses de vent dans le même intervalle, des assurés étant sinistrés. Quant aux points gris, ils représentent pour chaque intervalle sa proportion dans la base de données. Par exemple pour la tempête V1, 35% des assurés ont été exposés à des rafales entre 80 et 100 km/h et parmi ces assurés là, 10% ont été sinistrés.

On remarque aisément sur le graphique de gauche, celui sur la tempête de vent, que le pourcentage de sinistrés au niveau des habitations soumis à des rafales élevées (notamment au dessus de 100 km/h) est plus importante que pour le reste des bâtiments. En revanche, cette observation s'applique beaucoup moins pour la tempête de grêle, pour laquelle on observe tout de même un faible taux de sinistralité pour les vitesses de vent inférieures à 60 km/h (6%).

Il en ressort de ces observations que la variable *Vitesse\_vent* est bien explicative du nombre de sinistres provoqué par nos tempêtes, surtout au niveau de la tempête V1.

### La quantité de précipitations

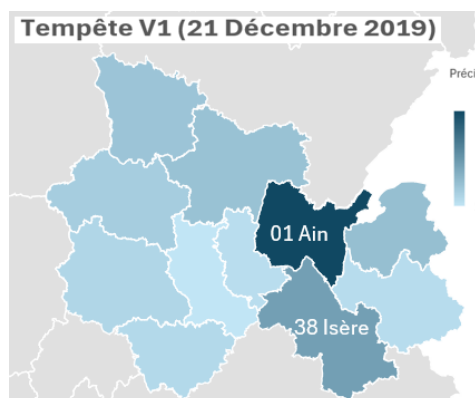
Une autre variable en lien avec les conditions météorologiques va être analysée, il s'agit des précipitations tombées en 24h les jours de survenance des tempêtes

Utilisons les mêmes axes d'analyses que pour la vitesse de vent :

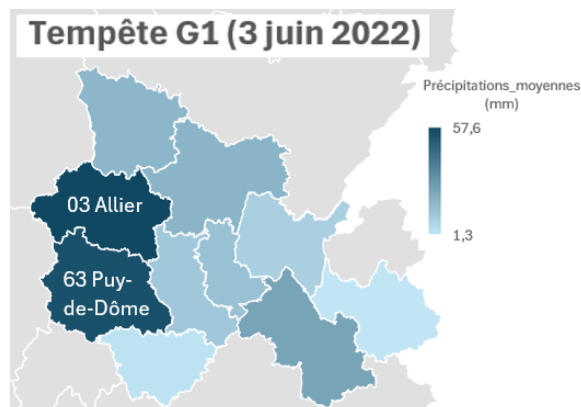
On observe tout d'abord des quantités de précipitations supérieures pour la tempêtes G1, comparé à celles de la tempête V1, ce qui valide une nouvelle le choix fait de distinguer les tempêtes principalement liées au vent et celles principalement liées à la grêle.

De plus, sur la carte décrivant la tempête V1, on retrouve bien une quantité élevée de précipitations dans le département de l'Isère, qui est celui le plus touché, mais on observe une quantité plus importante de précipitations tombée dans l'Ain, un département relevant une charge sinistres pourtant bien inférieure à la moyenne.

## CHAPITRE 2. MODÉLISATION DE LA CHARGE SINISTRE D'UNE TEMPÊTE MAJEURE SUR LE PORTEFEUILLE DE GRAA

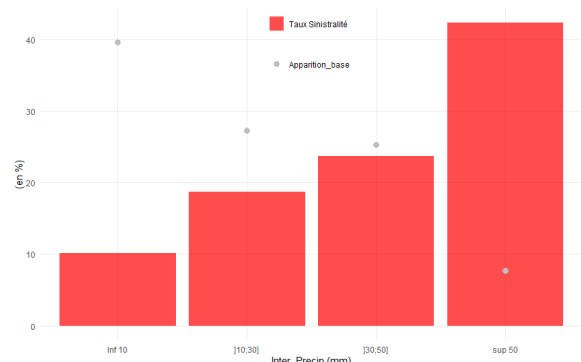


(a) Moyenne de la variable Précipitations par département pour la tempête V1

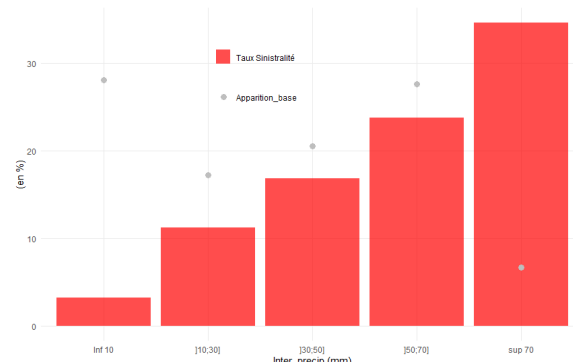


(b) Moyenne de la variable Précipitations par département pour la tempête G1

L'effet de cette variable se montre en revanche bien plus explicite au niveau de la tempête G1. En effet les 2 départements subissant le plus grand nombre de dégâts sont également ceux dont les quantités de précipitations relevées sont les plus grandes.



(a) Proportion de nombre de sinistres par quantité de précipitations pour la Tempête V1



(b) Proportion de nombre de sinistres par quantité de précipitations pour la Tempête G1

De la manière que la variable *Inter\_vent*, la variable *Inter\_precip* a été ajoutée pour présenter les graphiques ci dessus.

On peut clairement remarquer ici une augmentation des proportions de sinistrés lorsque des quantités élevées de précipitations sont relevées.

A l'image de la variable *Vitesse\_vent*, il a été montré que la variable *Précipitations* était explicative du nombre de sinistres, mais cette fois-ci particulièrement pour la tempête G1, en cohérence avec le fait qu'elle est été considérée comme une tempête de grêle et la tempête V1 comme une tempête de vent.

### La taille des grêlons

Quant à nos données de grêles, nous pouvons remarquer tout d'abord sur le graphique ci-dessous que la quantité de gros grêlons enregistrée au niveau du mois de juin 2022 (mois de survenance des tempêtes de grêles retenues pour entraîner et tester nos modèles) est bien anormalement élevé comparé à la quantité de gros grêlons enregistrée durant les mois de juin des autres années.

## CHAPITRE 2. MODÉLISATION DE LA CHARGE SINISTRE D'UNE TEMPÊTE MAJEURE SUR LE PORTEFEUILLE DE GRAA

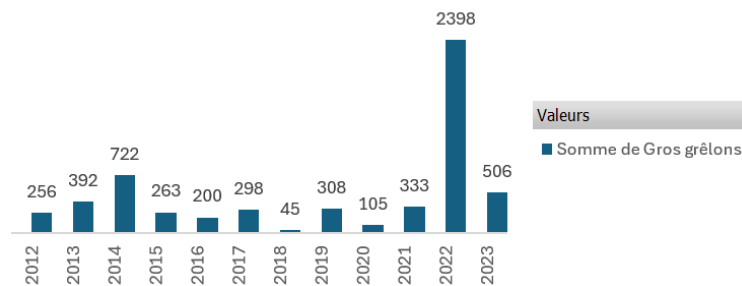


FIGURE 2.14 – Quantités de Gros grêlons tombées durant les mois de juin de chaque année

Sur la carte ci-dessous, qui montre les pourcentages de grêlons jugés comme "Gros grêlons" par département le jour de survenance de la tempête G1, on observe une proportion très importante de Gros grêlons dans

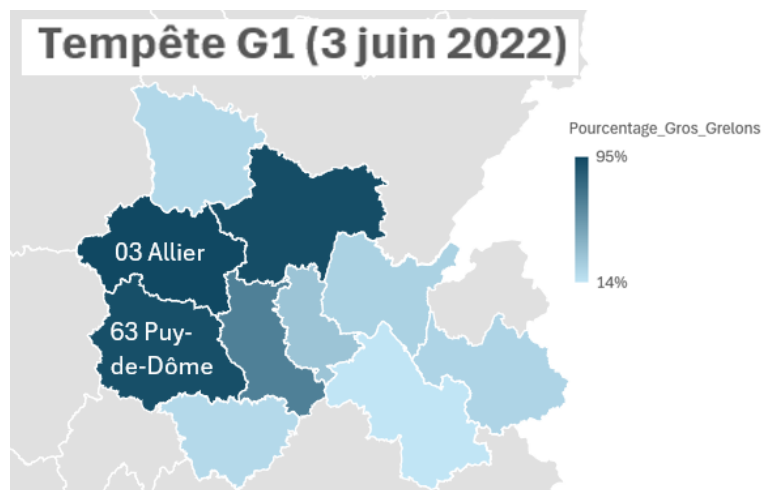
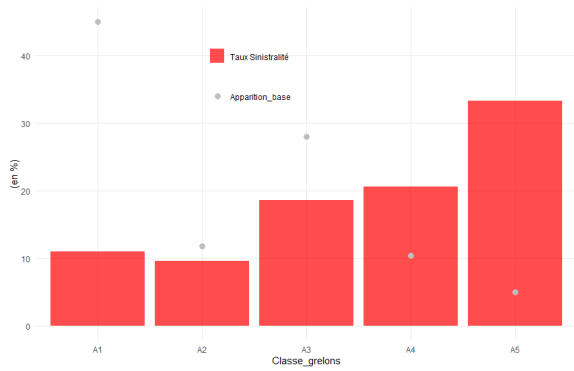


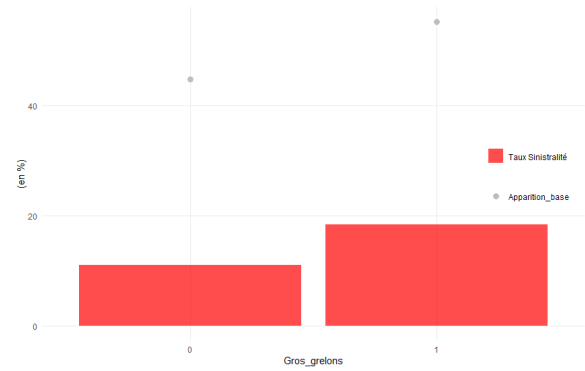
FIGURE 2.15 – Proportion de Gros grêlons par département pour la tempête G1

Pour être plus précis sur l'influence de la taille des grêlons sur la sinistralité, voici la répartition des proportions de sinistrés pour chaque classe de grêlons :

## CHAPITRE 2. MODÉLISATION DE LA CHARGE SINISTRE D'UNE TEMPÊTE MAJEURE SUR LE PORTEFEUILLE DE GRAA



(a) Proportion de nombre de sinistres pour la Tempête G1 en fonction de la variable classe\_grelons



(b) Proportion de nombre de sinistres pour la Tempête G1 en fonction de la variable Gros\_grelons

Il en ressort que la taille des grêlons impacte bien le nombre de sinistres sur la tempête G1 puisque le pourcentage de sinistrés croît au fur et à mesure de la classe des grêlons recensée est grande. Si l'on regarde le graphique de la variable *Gros\_grelons* on remarque également que plus de la moitié des assurés de notre base de données a été exposé à des gros grêlons. Nous détaillerons lors de l'élaboration de nos modèles la nécessité de sélectionner l'une ou l'autre de ces variables lors de l'entraînement de ceux-ci.

### Autres critères

Afin de définir précisément quels critères seront retenus comme variables explicatives qui serviront à entraîner nos modèles voici un récapitulatif des mêmes analyses que pour le vent et la taille des grêlons pour nos différentes variables de nos bases de données :



## CHAPITRE 2. MODÉLISATION DE LA CHARGE SINISTRE D'UNE TEMPÊTE MAJEURE SUR LE PORTEFEUILLE DE GRAA

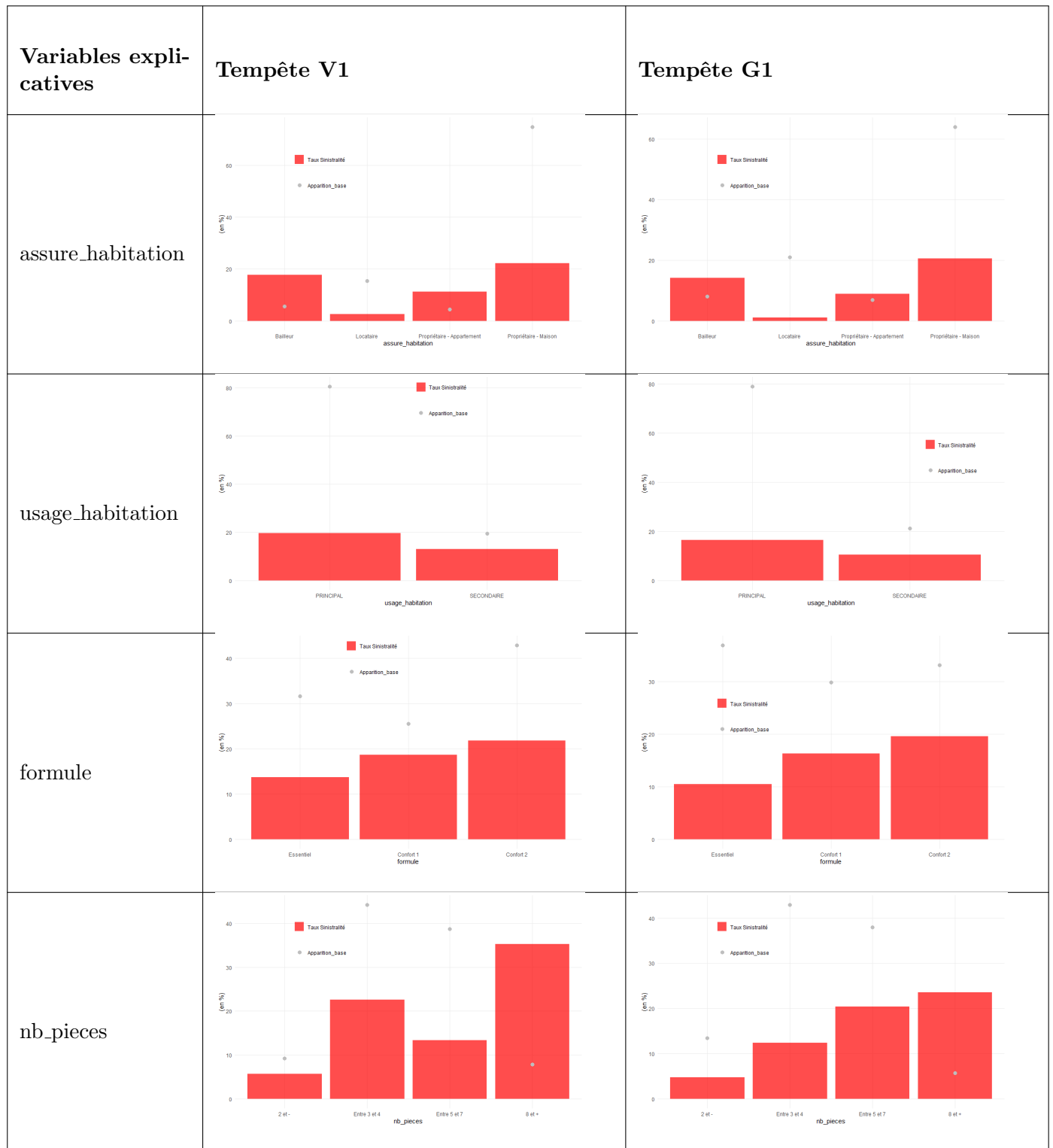


TABLE 2.2 – Tableau comparatif des variables explicatives pour les tempêtes de vent (V1) et de grêles (G1)

## CHAPITRE 2. MODÉLISATION DE LA CHARGE SINISTRE D'UNE TEMPÊTE MAJEURE SUR LE PORTEFEUILLE DE GRAA

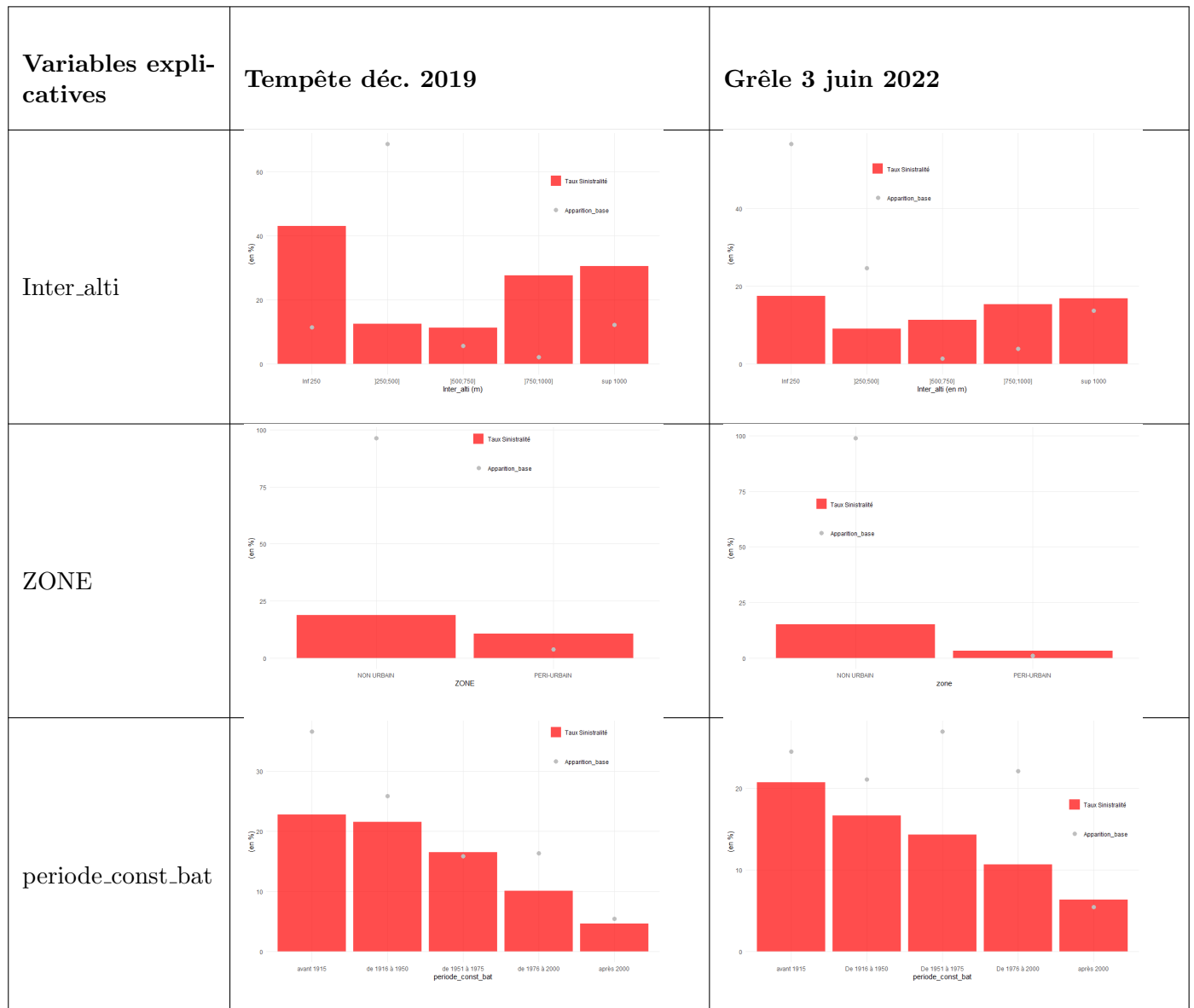


TABLE 2.3 – Tableau comparatif des variables explicatives pour les tempêtes V1 et G1

## CHAPITRE 2. MODÉLISATION DE LA CHARGE SINISTRE D'UNE TEMPÊTE MAJEURE SUR LE PORTEFEUILLE DE GRAA

L'analyse de ces graphiques nous permettent de choisir les variables qui seront utiles pour l'entraînement à la fois du modèle sur la tempête V1 et sur la tempête G1 :

- **assure\_Habitation : Nature de l'assure et de l'habitation**

On peut très clairement observer pour les deux tempêtes une sinistralité bien plus élevée chez les propriétaires de maison que pour les autres valeurs de la variable. Au delà du fait que cette catégorie représente plus de 60% du portefeuille, les maisons sont bien plus exposées au vent et à la grêle que les appartements.

- **usage\_habitation**

Les deux tempêtes ont montré que les habitations principales subissent plus de sinistres par rapport aux secondaires, probablement en raison de leur plus grande exposition aux aléas climatiques du quotidien.

- **formule**

La diversité des formules d'assurance montre que les formules intermédiaires (Confort 1) et premium (Confort 2) enregistrent une plus grande proportion de sinistres, dû à une couverture plus complète incitant les propriétaires à déclarer les dommages même mineurs.

- **nb\_pieces : Nombre de pièces composant l'habitation**

Les bâtiments avec un nombre plus élevé de pièces semblent avoir une proportion de sinistralité plus importante. Ceci peut s'expliquer par la taille et la complexité de la structure, qui augmentent les chances de dommages lors d'une tempête ou d'une grêle intense.

- **periode\_const\_bat : Date de construction de l'habitation**

Les bâtiments construits avant 1950 montrent une proportion plus élevée de sinistralité, particulièrement pour la tempête de Décembre 2019. Cela peut être attribué à des normes de construction moins robustes comparées aux constructions plus récentes.

En revanche certaines variables ne seront pas considérées comme explicatives de la fréquence de sinistre :

- **Altitude**

Le lien entre l'altitude où se situe les habitations et la probabilité de sinistralité n'est pas clairement définie sur ces graphiques, probablement dû au fait que même si des rafales de vent plus importantes peuvent être enregistrées en altitude, les constructions des bâtiments peuvent être plus robustes.

- **ZONE**

Cette variable n'est pas exploitable dans cette étude du fait de la proportion beaucoup trop élevée des habitations situées dans les zones définies comme "Non Urbaines" (96%). Ceci provient notamment du fait que Groupama est une société d'assurance très répandue dans les zones rurales et moins dans les zones péri-urbaines ou urbaines.

Maintenant qu'une base de données comportant une variable à expliquer (*Presence*) et plusieurs variables à priori explicatives (vitesse de vent, taille des grêlons, nature de l'assuré et de l'habitation, l'usage de l'habitation, la formule de contrat, le nombre de pièces, la période de construction de l'habitation et la quantité de précipitations relevée le jour des tempêtes) a été construite, plusieurs types de modèles vont pouvoir être entraînés et testés pour choisir celui qui donnera les meilleurs résultats, jugés à partir de métriques, par rapport aux probabilités qu'a chaque assuré d'être sinistré lors de ce genre de tempêtes.

## CHAPITRE 2. MODÉLISATION DE LA CHARGE SINISTRE D'UNE TEMPÊTE MAJEURE SUR LE PORTEFEUILLE DE GRAA

### 2.2.2 Métriques

Afin de comparer les performances des différents modèles étudiés, il est essentiel de définir des critères d'évaluation, appelés "métriques". Ces indicateurs permettent d'analyser la pertinence et la fiabilité des prédictions effectuées.

Avant d'examiner ces métriques en détail, il est nécessaire de structurer les résultats obtenus à l'aide d'un outil spécifique : la matrice de confusion.

#### Matrice de confusion

La matrice de confusion est un tableau qui permet d'évaluer les performances d'un modèle de classification en comparant ses prédictions avec les résultats réels. Selon une définition communément admise, cette matrice, également appelée tableau de contingence, mesure la fréquence des classifications correctes et erronées.

Un exemple de matrice de confusion est illustré dans le tableau ci-dessous :

	Négatif (Réal)	Positif (Réal)
Négatif (Prédiction)	Vrai Négatif (VN)	Faux Négatif (FN)
Positif (Prédiction)	Faux Positif (FP)	Vrai Positif (VP)

Interprétation de la matrice :

Vrais Négatifs (VN) : Le modèle ne prédit pas d'événement et aucun événement ne survient.

Faux Négatifs (FN) : Le modèle ne prévoit pas l'événement, mais celui-ci se produit.

Faux Positifs (FP) : Le modèle prévoit un événement qui ne se produit pas.

Vrais Positifs (VP) : Le modèle prévoit correctement l'événement lorsqu'il survient.

#### Accuracy

L'Accuracy est un indicateur simple qui mesure la proportion d'individus correctement classifiés par rapport à l'ensemble des observations. Elle se calcule selon la formule suivante :

$$\text{Accuracy} = \frac{VP + VN}{VP + FP + VN + FN}$$

Autrement dit, il s'agit du rapport entre le nombre de classifications correctes et le nombre total d'observations.

Cependant, cette métrique peut être trompeuse dans certains contextes, notamment lorsque les classes sont déséquilibrées. Dans cette étude, la classe majoritaire représente selon les tempêtes entre 80 et 85 % des observations. Ainsi, un modèle qui prédit systématiquement l'absence d'événement (classe 0) obtiendrait une Accuracy entre 80% et 85%, cette métrique peut donc être élevée sans pour autant garantir la pertinence des prédictions.

## CHAPITRE 2. MODÉLISATION DE LA CHARGE SINISTRE D'UNE TEMPÊTE MAJEURE SUR LE PORTEFEUILLE DE GRAA

L'objectif principal de cette étude étant la détection des événements rares (classe 1), l'Accuracy seule ne suffit pas à évaluer correctement la performance du modèle. D'autres métriques, mieux adaptées à ce contexte, ont donc été ajoutées.

### F1 score

Le calcul du F1-Score se fait à l'aide de autres métriques :

- La Précision : La précision est une métrique qui mesure la proportion d'individus correctement identifiés comme appartenant à la classe 1 par rapport au nombre total d'individus prédits dans cette classe. Elle correspond donc rapport entre le nombre de vrais positifs et le nombre total de prédictions positives ( $\frac{VP}{VP+FP}$ ).

La précision permet d'évaluer les individus de classe 0 mal classifiés (Faux Positifs). Une précision de 1 signifie que toutes les prédictions de classe 1 sont correctes, sans aucune fausse alerte.

- Le Rappel : Le rappel correspond au nombre d'individus correctement classés en classe 1 par rapport au nombre total d'individus appartenant à cette classe. Il correspond donc rapport entre le nombre de vrais positifs et le nombre total de réels positifs ( $\frac{VP}{VP+FN}$ ).

Le rappel accorde une importance particulière aux individus de classe 1. Une valeur de 1 signifie que toutes les tempêtes ont été correctement détectées, sans omission. Étant donné le déséquilibre de la variable cible dans l'étude, le modèle peut générer un nombre important de Faux Négatifs. L'objectif est donc de maximiser le rappel pour minimiser ces erreurs. Cependant, un rappel élevé avec une faible précision entraîne un nombre excessif de Faux Positifs, ce qui peut fausser l'interprétation des résultats.

Le F1 Score est une mesure qui combine la précision et le rappel en une moyenne harmonique, selon la formule suivante :

$$F1 = \frac{2 * Rappel * Précision}{Rappel + Précision}$$

L'intérêt principal du F1 Score est qu'il équilibre ces deux métriques, offrant ainsi une évaluation plus fiable des performances du modèle.

Dans cette étude, le F1 Score est pertinent car il prend en compte à la fois la capacité du modèle à prédire correctement les tempêtes et sa robustesse en minimisant le nombre de sinistres non détectées.

### AUC

L'AUC (AUC, Area Under the Curve) désigne l'aire sous la courbe ROC, c'est un indicateur essentiel pour évaluer la performance d'un modèle de classification, notamment dans des domaines comme la sélection des risques en assurance ou la détection de fraudes. Cet indicateur mesure la capacité du modèle à distinguer entre deux classes, par exemple les sinistrés et les non-sinistrés, en traçant la courbe ROC qui représente le taux de vrais positifs (Sensibilité) en fonction du taux de faux positifs (1 - Spécificité).

## CHAPITRE 2. MODÉLISATION DE LA CHARGE SINISTRE D'UNE TEMPÊTE MAJEURE SUR LE PORTEFEUILLE DE GRAA

Voici un graphique illustrant les formes des courbes ROC de modèles performants et non performants :

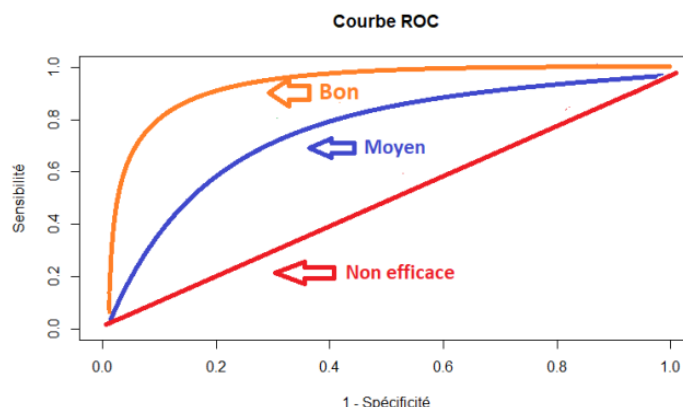


FIGURE 2.17 – Illustration de différents types de courbes ROC

Un AUC proche de 1 indique un modèle performant, tandis qu'une valeur proche de 0,5 suggère une classification aléatoire.

En résumé, la qualité de nos 3 différents modèles seront jugés à partir de trois indicateurs clés : l'Accuracy, le F1-Score et l'AUC.

### 2.2.3 Arbre de décision

#### Principe du modèle

Les arbres de décision sont des modèles de classification et de régression largement utilisés en apprentissage automatique pour leur interprétabilité et leur efficacité. Ils fonctionnent en divisant l'ensemble des données en sous-groupes homogènes en fonction des différentes variables explicatives.

Un arbre de décision repose sur une structure hiérarchique constituée de nœuds de décision, où chaque division se fait selon une règle basée sur une variable explicative. L'objectif est de partitionner les données de manière à minimiser l'incertitude et maximiser la pureté des sous-groupes obtenus. Dans le cadre de la prévision des sinistres d'un assuré, l'arbre de décision utilise plusieurs caractéristiques de l'assuré (âge, historique de sinistres, type de véhicule, localisation, etc.) pour estimer la probabilité d'occurrence d'un sinistre.

#### — Construction de l'Arbre

L'apprentissage d'un arbre de décision repose sur un algorithme récursif qui sélectionne, à chaque niveau, la variable explicative qui permet la meilleure séparation des données. Cette séparation est généralement basée sur des critères tels que :

L'indice de Gini, qui mesure l'impureté des groupes formés.

L'entropie, qui quantifie l'incertitude avant et après une division.

Le processus continue jusqu'à atteindre une condition d'arrêt, comme une profondeur maxi-

## CHAPITRE 2. MODÉLISATION DE LA CHARGE SINISTRE D'UNE TEMPÊTE MAJEURE SUR LE PORTEFEUILLE DE GRAA

male de l'arbre ou un nombre minimal d'observations par feuille.

### — Paramètres de l'arbre

Dans un modèle d'arbre de décision deux paramètres permettent de réguler la complexité et la performance du modèle :

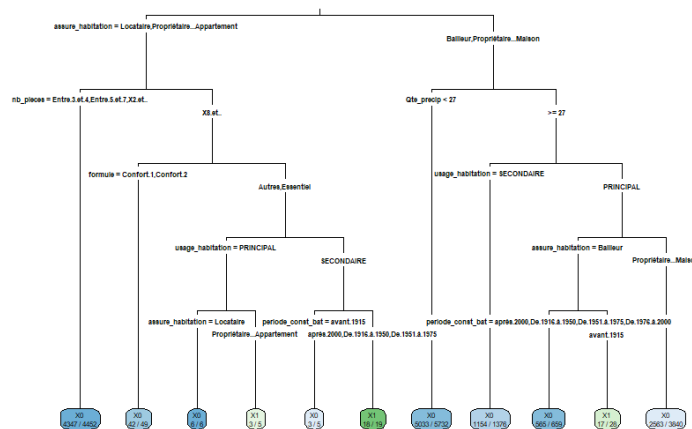
- Complexity Parameter (cp) : Le paramètre cp contrôle la taille de l'arbre en imposant une contrainte sur la réduction minimale de l'erreur nécessaire pour effectuer une division. Une valeur plus élevée de cp empêche la création de divisions mineures, limitant ainsi le surajustement (overfitting), tandis qu'une valeur trop faible peut conduire à un arbre excessivement complexe et spécifique aux données d'entraînement.
- Taille maximale de l'arbre (maxdepth) : Ce paramètre fixe la profondeur maximale de l'arbre, influençant directement la granularité de la segmentation. Une profondeur trop élevée risque de surajuster les données, alors qu'une profondeur trop faible peut mener à un modèle sous-ajusté (underfitting).

L'optimisation conjointe de ces deux paramètres, souvent via validation croisée, est essentielle pour obtenir un bon compromis entre biais et variance dans un modèle d'arbre de décision.

### — Interprétation des résultats

Une fois l'arbre construit, il est possible de visualiser les décisions sous forme d'un diagramme hiérarchique. Chaque branche correspond à une règle conditionnelle et chaque feuille représente une classification finale. Par exemple, un assuré ayant un historique de sinistres élevé et conduisant un véhicule puissant peut être classé dans un groupe à forte probabilité de sinistre.

Voici à quoi ressemble ce diagramme hiérarchique pour une profondeur d'arbre égale à 5 récolté à l'aide de la base de donnée de la tempête G1 :



## CHAPITRE 2. MODÉLISATION DE LA CHARGE SINISTRE D'UNE TEMPÊTE MAJEURE SUR LE PORTEFEUILLE DE GRAA

bailleur et qui sont dans des communes qui ont été exposées à des quantités de précipitations supérieures à 27mm, 5033 n'ont pas déclarés de sinistre.

### — Avantages et limites

L'arbre de décision présente plusieurs avantages :

- Facilité d'interprétation : les décisions sont exprimées sous forme de règles explicites.
- Adaptabilité aux données catégoriques et numériques : il fonctionne avec une grande variété de variables.
- Faible prétraitement des données : peu de transformation des variables est requise.

Cependant, il présente aussi des limites :

- Sensibilité au sur-apprentissage : un arbre trop profond risque de mémoriser les bruits des données d'apprentissage.
- Rigidité des règles : une légère modification des données peut entraîner des changements significatifs dans la structure de l'arbre.

Pour pallier ces limitations, des approches comme les forêts aléatoires ou le boosting peuvent être utilisées afin d'améliorer la robustesse et la généralisation du modèle.

L'arbre de décision constitue donc un outil performant pour estimer la probabilité qu'un assuré ait un sinistre ou non, en exploitant les caractéristiques individuelles et historiques du profil de l'assuré.

### Application aux bases de données

Dans un premier temps, une séparation des bases correspondant aux tempêtes V1 et G1 est réalisée. Ainsi, chacune des deux bases est coupée aléatoirement pour former :

- Une base d'entraînement : correspondant à 80% des bases d'origines et qui constituera la base sur laquelle s'appuyer pour construire le modèle.
- Une base de test : correspondant à 20% des bases d'origines et qui sera utilisée pour calculer les métriques associées aux différents modèles testés.

### Arbres maximaux sans optimisation des paramètres

Voyons tout d'abord comment se comportent nos modèles en prenant les arbres maximaux, c'est-à-dire en choisissant les paramètres qui maximisent le nombre de branches de l'arbre (un cp très bas et un maxdepth très haut) :



## CHAPITRE 2. MODÉLISATION DE LA CHARGE SINISTRE D'UNE TEMPÊTE MAJEURE SUR LE PORTEFEUILLE DE GRAA

Tempête	cp	maxdepth	Mat confusion
V1	0,05%	30	<pre> Reference Prediction x0 x1 x0 688 81 x1 31 81 </pre>
G1	0,05%	30	<pre> Reference Prediction 0 1 0 3198 370 1 233 241 </pre>

TABLE 2.4 – Matrices de confusion des arbres maximaux

Tempête	cp	maxdepth	Accuracy	F1-Score	AUC
V1	0,05%	30	0,87	0,59	0,73
G1	0,05%	30	0,85	0,44	0,70

TABLE 2.5 – Métriques des arbres maximaux

A noter que toutes matrices de confusion de cette étude s'obtiennent en attribuant aux assurés dont le modèle leur associe une probabilité supérieure ou égale à 0,5 le critère positif (Presence\_prédite = 1) et le critère négatif (Presence\_prédite = 0) sinon.

Les résultats donnés par les arbres ne sont sans surprise pas concluant. En effet, pour la tempête G1 un F1-Score de 0,44 signifie que le modèle réalise des prédictions très aléatoires. Ces résultats sont principalement dûs aux choix des hyperparamètres. Un cp de 0,05% signifie que l'arbre continue à se développer tant qu'une nouvelle division réduit l'erreur relative d'au moins 0,05%, Cette valeur est très basse, ce qui signifie que l'arbre peut devenir très détaillé, augmentant ainsi le risque de surajustement (overfitting). Quant au maxdepth, 30 est une profondeur élevée, ce qui permet à l'arbre de capturer des relations complexes, mais cela augmente aussi le risque de surajustement.

### Arbres optimaux

Pour améliorer les performances des modèles et régler ces problèmes de surajustement, une optimisation des hyperparamètres est réalisée par validation croisée. En voici le processus :

1. Division du jeu de données : Les données sont séparées en k sous-ensembles (ou "folds"), généralement k prend les valeurs 5 ou 10 (dans notre cas le choix de k = 5 a été fait). Chaque fold servira successivement d'ensemble de validation, tandis que les k-1 folds restants seront utilisés pour entraîner le modèle.
2. Définition d'une grille de valeurs pour cp et maxdepth : Un ensemble de valeurs candidates pour les hyperparamètres est choisi. Ici par exemple :  
cp = seq(0.0005, 0.1, by = 0.0005)  
maxdepth = seq(1, 30, by = 1) Entraînement et évaluation du modèle
3. Aggrégation des performances : Pour chaque combinaison de cp et maxdepth, un modèle d'arbre de décision est entraîné sur les k-1 folds et testé sur le fold de validation. Les métriques de performance sont alors calculés pour chacune de ces combinaisons.

## CHAPITRE 2. MODÉLISATION DE LA CHARGE SINISTRE D'UNE TEMPÊTE MAJEURE SUR LE PORTEFEUILLE DE GRAA

4. Sélection des meilleurs hyperparamètres : La combinaison (cp, maxdepth) qui maximise la performance du modèle sur la validation croisée est alors identifiée. En cas de performances similaires, le modèle le plus simple est privilégié pour éviter le surajustement.

Après avoir réalisé ces étapes sur chacune de nos bases, voici les hyperparamètres qui minimisent l'erreur et optimisent les performances des modèles :

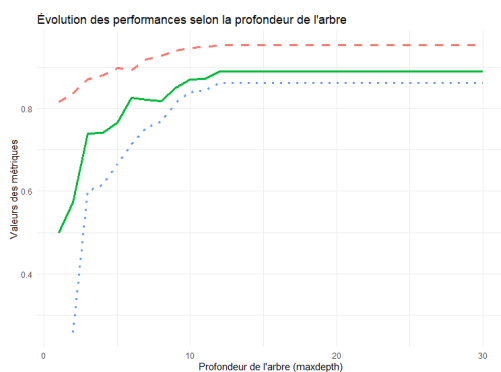
Tempête	cp	maxdepth	Mat confusion
V1	0,1%	12	<pre> Reference Prediction x0 x1 x0 712 34 x1 7 128 </pre>
G1	0,1%	18	<pre> Reference Prediction 0 1 0 3176 295 1 255 316 </pre>

TABLE 2.6 – Matrices de confusion des arbres maximaux

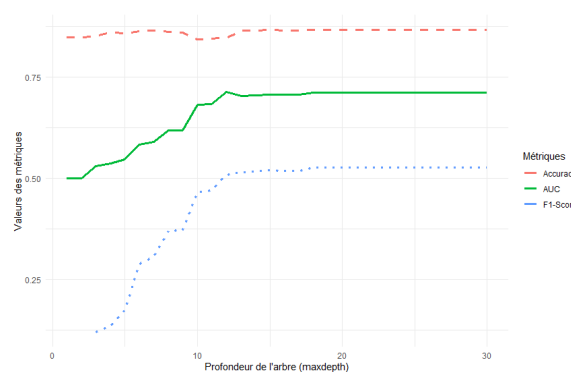
Tempête	cp	maxdepth	Accuracy	F1-Score	AUC
V1	0,1%	12	0,95	0,86	0,89
G1	0,1%	18	0,86	0,54	0,83

TABLE 2.7 – Métriques des arbres optimaux

En effet si l'on observe les courbes des métriques de nos modèles (en fixant le paramètre cp avec celui retenu) en fonction de la profondeur maximale de l'arbre :



(a) Métriques selon la variable maxdepth de la tempête V1



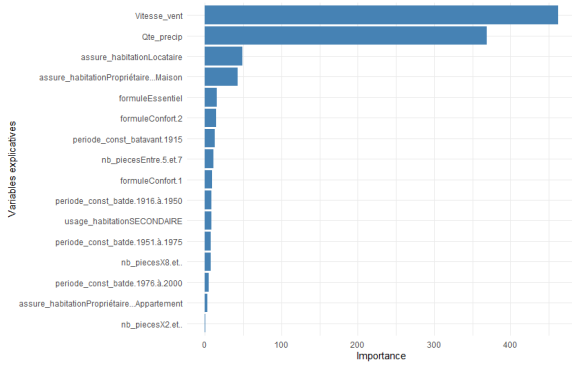
(b) Métriques selon la variable maxdepth de la tempête G1

On observe bien qu'à partir d'un certain nombre de branches, les performances des modèles n'évoluent plus, le fait de prendre le maxdepth le plus faible évite le surajustement et ne dégrade donc pas la qualité du modèle.

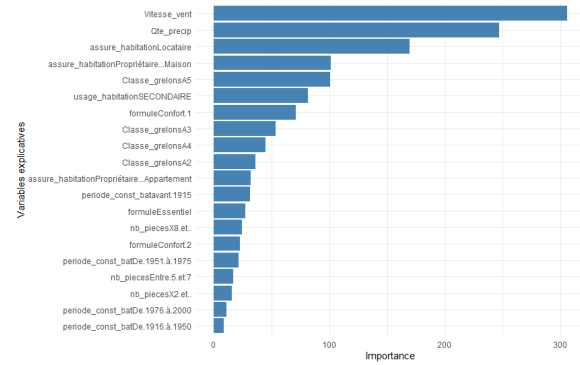
Pour finir si l'on se penche sur l'importance qu'accorde les modèles aux différentes variables explicatives :

On observe que la variable explicative qui influe le plus les deux modèles d'arbre de décision est

## CHAPITRE 2. MODÉLISATION DE LA CHARGE SINISTRE D'UNE TEMPÊTE MAJEURE SUR LE PORTEFEUILLE DE GRAA



(a) Importance variables tempête V1



(b) Importance variables tempête G1

la variable *Vitesse\_vent*. Or il a été remarqué lors des analyses sur les bases de données que cette variable n'était pas la plus significative par rapport au nombre de sinistres pour la tempête G1. Alors qu'au contraire la variable *Classe\_grelons*, qui semblaient beaucoup influencer sur la probabilité de sinistralité d'un assuré, influe moins sur les résultats de l'arbre. Ce point est sans-doute un élément qui expliquerait le faible F1-Score donné par le modèle optimal définit pour cette tempête comparé à celui sur la tempête V1.

Maintenant que les modèles d'arbres de décision optimaux ont été construits, passons désormais à un autre type de modèle de classification mais qui est néanmoins en lien avec les arbres de décision : les forêts aléatoires.

### 2.2.4 Forêts aléatoires

#### Principe du modèle

La méthode Random Forest (Forêts aléatoires) est une extension des arbres de décision visant à améliorer leur performance et leur robustesse en combinant plusieurs arbres de décision dans un modèle d'ensemble. Cette approche permet de limiter les risques de sur-apprentissage et d'augmenter la précision des prédictions.

Random Forest repose sur le concept de bagging (Bootstrap Aggregating), qui consiste à entraîner plusieurs arbres de décision indépendamment à partir d'échantillons aléatoires des données d'apprentissage. Chaque arbre prend des décisions de manière autonome, et la prédiction finale est obtenue par un vote majoritaire (pour la classification) ou une moyenne (pour la régression).

Dans le cadre de la prédiction des sinistres d'un assuré, Random Forest permet également d'estimer la probabilité de survenue d'un sinistre en exploitant les différentes caractéristiques des assurés.

#### — Différences avec un Arbre de Décision Simple

Bien qu'un arbre de décision unique soit facile à interpréter et rapide à entraîner, il présente certaines limites, notamment une forte sensibilité aux variations des données et un risque de sur-ajustement. Random Forest pallie ces défauts en combinant plusieurs arbres de décision et

## CHAPITRE 2. MODÉLISATION DE LA CHARGE SINISTRE D'UNE TEMPÊTE MAJEURE SUR LE PORTEFEUILLE DE GRAA

en introduisant de l'aléatoire dans deux aspects clés :

- La sélection des données : Chaque arbre est entraîné sur un sous-échantillon différent des données d'origine.
- La sélection des variables : à chaque nœud de décision, un sous-ensemble aléatoire de variables explicatives est considéré pour réaliser la séparation.

Ces mécanismes permettent d'obtenir un modèle plus robuste et moins sujet aux fluctuations des données.

### — Paramètres de l'arbre

Dans un modèle de forêt aléatoire, il existe deux paramètres déterminants pour la qualité des prédictions.

- Nombre de variables sélectionnées (mtry) : Le paramètre mtry fixe le nombre de variables sélectionnées de manière aléatoire à chaque division d'un arbre. Une valeur trop faible risque de limiter la capacité du modèle à capturer des relations pertinentes, tandis qu'une valeur trop élevée réduit la diversité entre les arbres, compromettant ainsi la puissance de la méthode.
- Nombre d'arbres (ntree) : ce paramètre correspond au nombre d'arbres générés dans la forêt. Un nombre insuffisant peut entraîner une forte variabilité des prédictions, tandis qu'un nombre trop élevé allonge inutilement le temps de calcul sans réel gain en précision.

L'ajustement optimal de ces deux paramètres permet d'obtenir un modèle robuste, performant et généralisable.

### — Interprétation des Résultats

Bien que la structure de Random Forest soit plus complexe qu'un arbre de décision unique, il est possible d'évaluer son importance en analysant :

- L'importance des variables : tout comme dans les modèles d'arbre de décisions simples, certaines caractéristiques des assurés ont plus d'influence que d'autres sur la prédiction des sinistres.
- Les votes des arbres : l'analyse des décisions prises par chaque arbre permet de mieux comprendre les résultats du modèle.

### — Avantages et Limites

Ce type de modèle présente plusieurs avantages :

- Amélioration de la précision : en combinant plusieurs modèles, Random Forest réduit les erreurs de prédiction.
- Réduction du sur-apprentissage : les arbres générés sont moins sensibles aux variations des données.
- Capacité à gérer les données manquantes et les variables bruitées : le modèle est plus résilient aux incohérences des données.

Mais également certaines limites :

- Moins interprétable qu'un arbre unique : bien que performant, Random Forest est une boîte noire plus difficile à expliquer.
- Temps de calcul plus élevé : en raison du nombre d'arbres générés, l'entraînement et

## CHAPITRE 2. MODÉLISATION DE LA CHARGE SINISTRE D'UNE TEMPÊTE MAJEURE SUR LE PORTEFEUILLE DE GRAA

l'inférence peuvent être plus coûteux en ressources.

En conclusion, Random Forest est une alternative puissante aux arbres de décision pour modéliser la probabilité d'un sinistre. Il combine les avantages des arbres de décision tout en corrigeant leurs faiblesses, ce qui en fait un modèle adapté aux problématiques de classification dans le domaine de l'assurance.

### Application aux bases de donnée

Concernant la séparation des données, l'utilisation de la fonction *set.seed* permet d'obtenir le même partitionnement que pour le modèle d'arbre de décision et ainsi pouvoir avoir une comparaison plus significative des résultats. Ainsi les mêmes bases d'entraînement et de test sont formées

### Forêts maximales sans optimisation des hyperparamètres

Voyons tout d'abord comment se comportent nos modèles en prenant les forêts maximales, c'est-à-dire en choisissant les paramètres qui maximisent le nombre d'arbres (un *ntree* très haut) et le nombre de variables sélectionnées à chaque division (*mtry* maximale) :

Tempête	mtry	ntree	Mat confusion												
V1	7	500	<table><tr><td></td><td colspan="2">Reference</td></tr><tr><td>Prediction</td><td>0</td><td>1</td></tr><tr><td>0</td><td>709</td><td>25</td></tr><tr><td>1</td><td>10</td><td>137</td></tr></table>		Reference		Prediction	0	1	0	709	25	1	10	137
	Reference														
Prediction	0	1													
0	709	25													
1	10	137													
G1	8	500	<table><tr><td></td><td colspan="2">Reference</td></tr><tr><td>Prediction</td><td>0</td><td>1</td></tr><tr><td>0</td><td>3226</td><td>340</td></tr><tr><td>1</td><td>205</td><td>271</td></tr></table>		Reference		Prediction	0	1	0	3226	340	1	205	271
	Reference														
Prediction	0	1													
0	3226	340													
1	205	271													

TABLE 2.8 – Matrices de confusion des forêts maximales

Tempête	mtry	ntree	Accuracy	F1-Score	AUC
V1	7	500	0,96	0,88	0,91
G1	8	500	0,86	0,50	0,69

TABLE 2.9 – Métriques des forêts maximales

Ces premiers modèles donnent dès lors de bons résultats pour la tempête V1 puisque sur les 162 sinistrés de la bases, 137 reçoivent une probabilité supérieure ou égale à 0,5, ce qui donne un rappel de 85%. Et sur les 147 sinistres prédits par le modèle, les mêmes 137 ont effectivement été sinistrés, ce qui donne une précision de 93%.

Les résultats pour la tempête sont encore une fois inférieurs à ceux de la tempête V1.

Voyons désormais si un changement des hyperparamètres permettrait de d'améliorer les performances de ces modèles.

## CHAPITRE 2. MODÉLISATION DE LA CHARGE SINISTRE D'UNE TEMPÊTE MAJEURE SUR LE PORTEFEUILLE DE GRAA

### Forêts optimales

Pour améliorer les performances des modèles une optimisation des hyperparamètres est une nouvelle fois réalisée par validation croisée. Cette fois-ci les différentes valeurs des `mtry` et `ntree` testées sont :

- `mtry = seq(1, 7, by = 1)` pour la tempête V1 et `mtry = seq(1, 8, by = 1)` pour la tempête G1.
- `ntree = seq(10, 500, by = 1)`

Après avoir réalisé les étapes du processus sur chacune de nos bases, voici les hyperparamètres qui optimisent les performances des modèles :

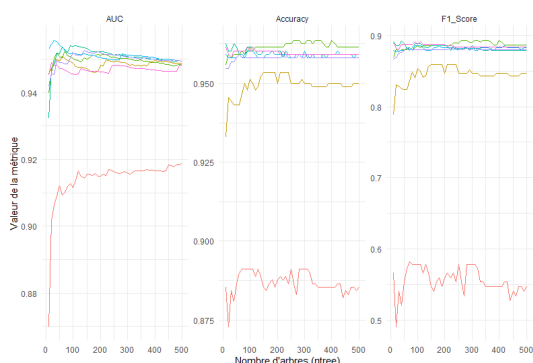
Tempête	mtry	ntree	Mat confusion
V1	6	90	<pre> Reference Prediction  0  1            0 715 28            1   4 134 </pre>
G1	5	220	<pre> Reference Prediction  0  1            0 3228 331            1  203 280 </pre>

TABLE 2.10 – Matrices de confusion des forêts maximales

Tempête	mtry	ntree	Accuracy	F1-Score	AUC
V1	6	90	0,96	0,89	0,95
G1	5	220	0,87	0,51	0,70

TABLE 2.11 – Métriques des forêts optimales

Le réduction du nombre d'arbres dans les forêts aléatoires ainsi que du nombre de variables utilisées à chaque noeud a donc en effet permis de limiter les erreurs de prédiction et d'améliorer légèrement les performances des modèles, voici pour chaque modèle l'évolution des métriques en fonction des valeurs de `mtry` et `ntree` :



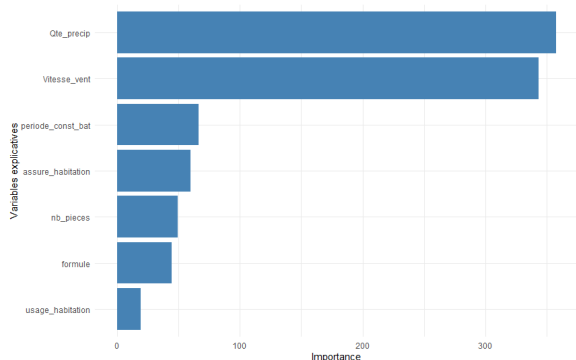
(a) Métriques selon les valeurs de `mtry` et `ntree` de la tempête V1



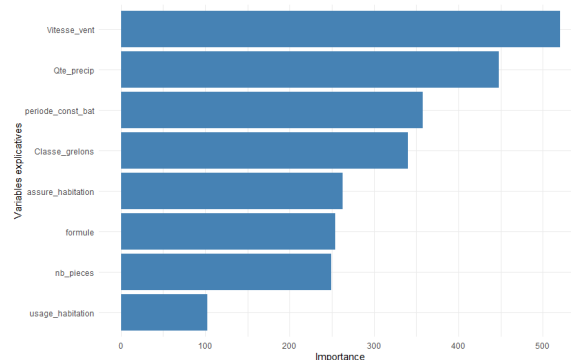
(b) Métriques selon les valeurs de `mtry` et `ntree` de la tempête G1

Pour finir si l'on se penche sur l'importance qu'accorde les modèles aux différentes variables explicatives :

## CHAPITRE 2. MODÉLISATION DE LA CHARGE SINISTRE D'UNE TEMPÊTE MAJEURE SUR LE PORTEFEUILLE DE GRAA



(a) Importance variables tempête V1



(b) Importance variables tempête G1

On observe cette fois ci que le modèle pour la tempête V1 accorde légèrement plus d'importance à la quantité de précipitation qu'à la vitesse de vent.

Quant au modèle associé à la tempête G1, on remarque que l'importance des variables est plus équilibrée que lors du modèle d'arbre de décision simple, cependant cela n'a pas permis d'améliorer les performances de prédictions.

Passons désormais au dernier type de modèle de classification qui sera testé dans cette étude, un modèle linéaire généralisé.

### 2.2.5 Modèles linéaires généralisés

#### Principe du modèle

Les modèles linéaires généralisés (GLM) constituent une extension des modèles linéaires classiques. Contrairement aux modèles linéaires qui supposent que la variable réponse suit une distribution normale, les GLM permettent d'établir un lien entre la variable réponse et les prédicteurs à l'aide d'une fonction de lien.

Dans un modèle linéaire généralisé, la distribution de la variable réponse  $Y$  appartient à la famille des distributions exponentielles. Ils sont particulièrement adaptés aux problèmes de classification et de prévision de probabilités, comme la modélisation du risque de sinistre en assurance.

Un modèle linéaire généralisé repose sur trois éléments fondamentaux :

- La variable réponse  $Y$  : également appelée *composante aléatoire*, elle suit une distribution appartenant à la famille exponentielle.
- Une fonction de lien : elle relie la moyenne de la variable réponse à une combinaison linéaire des variables explicatives.
- Une distribution de probabilité : contrairement à la régression linéaire classique, les MLG peuvent s'adapter à différentes distributions (binomiale, poisson, gamma, etc.).

Dans le cadre de l'assurance, le GLM permet de modéliser la probabilité qu'un assuré ait un sinistre en fonction de ses caractéristiques personnelles (âge, historique de sinistres, type de véhicule, etc.).

#### — Modèle de régression logistique et Modèle de Poisson

Deux types de MLG sont couramment utilisés pour la modélisation des sinistres :

- La régression logistique : adaptée aux problèmes de classification binaire (0 ou 1), elle permet d'estimer la probabilité qu'un assuré ait un sinistre ou non. La fonction de lien utilisée est

## CHAPITRE 2. MODÉLISATION DE LA CHARGE SINISTRE D'UNE TEMPÊTE MAJEURE SUR LE PORTEFEUILLE DE GRAA

la fonction logistique :

$$p \mapsto \ln \left( \frac{p}{1-p} \right)$$

associée à une distribution de Bernoulli donnée par :

$$P(Y = y) = p^y(1-p)^{1-y}, \quad \text{pour } y \in \{0, 1\}.$$

Où  $p$  est la probabilité d'occurrence d'un des 2 critères.

- Le modèle de Poisson : utilisé pour prédire le nombre de sinistres attendus sur une période donnée, il repose sur une fonction de lien logarithmique :

$$g(\mu) = \ln(\mu), \quad \text{pour } \mu \in R^*.$$

et une distribution de Poisson :

$$P(Y = k) = \frac{\lambda^k e^{-\lambda}}{k!}, \quad \text{pour } k \in \mathbb{N}.$$

### — Interprétation des Résultats

Une fois le modèle ajusté, l'interprétation repose sur l'analyse des coefficients, qui indiquent l'impact de chaque variable explicative sur la probabilité de sinistre. Par exemple, un coefficient positif pour l'âge pourrait signifier qu'un assuré plus âgé a un risque accru de sinistre.

La significativité des coefficients se mesure à l'aide de la p-value, elle mesure la probabilité d'observer une statistique de test aussi extrême que celle obtenue, sous l'hypothèse nulle où le coefficient du prédicteur est nul. Une p-value faible (généralement inférieure à un seuil de 0.05) suggère que le prédicteur a un effet significatif sur la variable réponse, tandis qu'une p-value élevée indique que l'effet observé pourrait être dû au hasard.

Les p-values sont calculées à partir du test de Wald ou du test du rapport de vraisemblance (Likelihood Ratio Test), en fonction de la distribution choisie pour la variable réponse. Contrairement aux modèles linéaires classiques où l'inférence repose sur la normalité des erreurs, les MLG utilisent la vraisemblance pour estimer les paramètres et tester leur significativité. Il est donc essentiel de bien interpréter ces p-values en tenant compte du choix de la fonction de lien et de la distribution de la variable réponse, afin d'éviter des conclusions erronées sur l'importance des prédicteurs.

### — Avantages et Limites

Les modèles linéaires généralisés possèdent de nombreux avantages :

- Modèle simple et interprétable : les coefficients permettent une analyse directe de l'impact des variables.
- Flexibilité : adaptation à différentes distributions de la variable réponse.
- Efficacité computationnelle : rapide à entraîner et bien adapté aux grands ensembles de données.

Mais également quelques limites :

- Hypothèses strictes : nécessite une bonne spécification de la fonction de lien et de la distribution des résidus.



## CHAPITRE 2. MODÉLISATION DE LA CHARGE SINISTRE D'UNE TEMPÊTE MAJEURE SUR LE PORTEFEUILLE DE GRAA

- Moins performant pour des relations non linéaires complexes : contrairement aux méthodes comme Random Forest ou les réseaux de neurones, les GLM ne capturent pas facilement des interactions complexes.

En conclusion, les modèles linéaires généralisés sont des outils puissants et facilement interprétables pour la modélisation du risque de sinistre en assurance. Ils permettent d'estimer des probabilités de sinistre de manière robuste tout en conservant une facilité d'interprétation.

### Application à la base de donnée

Les mêmes bases d'entraînement et de test que pour les modèles précédents sont une nouvelle fois récupérées.

Comme expliqué dans la présentation des modèles linéaires généralisés, la première étape consiste en la sélection d'une fonction de lien et d'une loi de distribution qui serviront à prédire efficacement les probabilités associées à chaque type d'assuré, en sachant que dans le cas de cette étude la variable réponse est la variable *Presence*, qui est une variable binaire.

Au vue de la présentation des divers modèles qu'il existait, le modèle de régression logistique est dans notre cas le plus adapté, ce qui signifie que la fonction de lien des modèles sera la fonction logistique et la loi de distribution sera la loi binomiale.

### Régressions complètes

En laissant toutes nos variables explicatives telles qu'elles sont dans les bases de données, le modèle de régression se présente comme suit :

$$Presence_i \sim \text{Bernoulli} \left( \frac{1}{1 + \exp(-(\beta_1 + \text{Vitesse\_vent} \cdot \beta_2 + \text{Qte\_Precip} \cdot \beta_3 + \text{assure\_habitation} \cdot \beta_{4,5,6,7} + \dots))} \right).$$

Avec  $\beta_i$  le coefficient associé au  $i$ -ème critère dans le modèle.

Ces modèles présentent les résultats suivant :

Tempête	Mat confusion		
V1	Reference		
	Prediction	0	1
	0	623	70
	1	96	92
G1	Reference		
	Prediction	0	1
	0	3106	382
	1	325	229

TABLE 2.12 – Matrices de confusion des régressions globales

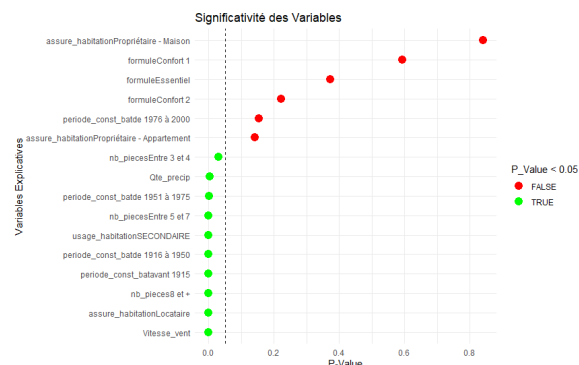
Tempête	Accuracy	F1-Score	AUC
V1	0,81	0,53	0,72
G1	0,86	0,39	0,65

TABLE 2.13 – Métriques de la régression globale

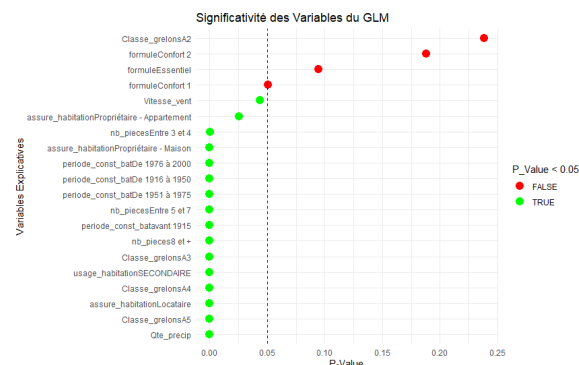
## CHAPITRE 2. MODÉLISATION DE LA CHARGE SINISTRE D'UNE TEMPÊTE MAJEURE SUR LE PORTEFEUILLE DE GRAA

Les résultats obtenus nous indiquent que les modèles prédits sont très aléatoires, des changements au niveau des variables est donc nécessaire pour améliorer leurs performances.

Voyons tout d'abord les p-values associées à chacune des caractéristiques de nos bases :



(a) Significativité des critères pour la régression complète de la tempête V1



(b) Significativité des critères pour la régression complète de la tempête G1

On remarque dans un premier temps que l'ensemble des critères de la variable *formule* possède une p-value supérieure au seuil de significativité de 0,05. Cela signifie que l'effet observé de cette variable sur celle à expliquer (*Presence*) est probablement dû au hasard.

Ensuite, pour la tempête V1, deux des critères sur les quatre qui composent la variable *assure\_habitation* ont également une p-value trop élevée. Cette variable peut donc aussi être non-significative dans la prédiction des modèles.

Pour le reste des variables, les p-values sont assez basses pour ne pas avoir à les retirer des modèles.

### Régressions corrigées

Pour améliorer les performances des modèles plusieurs variables ont été modifiées :

- La variable *formule* a été retirée des 2 modèles.
- La variable *assure\_habitation* a été retirée uniquement du modèle prédisant la tempête V1.
- Les variables *Vitesse\_vent* et *Qte\_precip* ont été remplacées par les variables *Inter\_vent* et *Inter\_precip* pour pallier les erreurs d'extraction liées à la distance qui sépare le lieu du sinistre et la station météo la plus proche.

Après avoir réalisé ces modifications, voici les métriques associées aux modèles des deux tempêtes :

Tempête	Mat confusion
V1	Reference
	Prediction 0 1
	0 636 73
	1 83 89
G1	Reference
	Prediction 0 1
	0 3056 259
	1 375 352

TABLE 2.14 – Matrices de confusion des régressions optimales

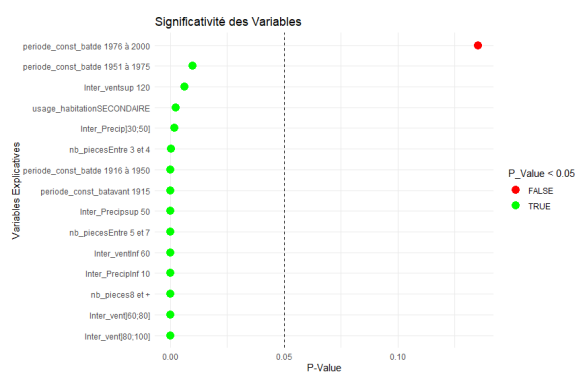
## CHAPITRE 2. MODÉLISATION DE LA CHARGE SINISTRE D'UNE TEMPÊTE MAJEURE SUR LE PORTEFEUILLE DE GRAA

Tempête	Accuracy	F1-Score	AUC
V1	0,82	0,53	0,72
G1	0,84	0,53	0,70

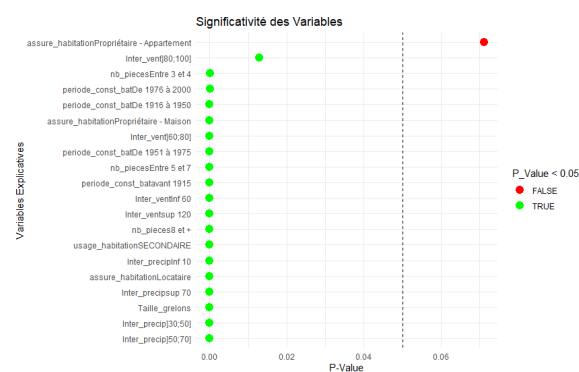
TABLE 2.15 – Métriques des régressions optimales

On observe une nette amélioration des performances par rapport aux régressions globales.

Et si l'on extrait les p-value pour chacun des critères :



(a) Significativité des régressions optimales tempête V1



(b) Significativité des régressions optimales tempête G1

On s'aperçoit que seulement un critère par modèle a une p-value supérieure au seuil, ce qui est suffisant pour considérer que chacune des variables est significative dans la prédiction des probabilités de survenance des sinistres dans ces modèles

Des trois modèles de classification qui ont été traitées, le plus performant sera garder pour prédire les nombres de sinistres provoqués par une tempête de vent ou de grêle.

### 2.2.6 Analyse des résultats

#### Choix du meilleur modèle

Voici un résumé des différents indicateurs qui ont été obtenus pour les trois modèles testés.

**Pour la tempête V1 :**

Métriques	Arbre de décision	Forêts aléatoires	MLG
Accuracy	0,95	0,96	0,82
F1-Score	0,86	0,89	0,53
AUC	0,89	0,95	0,72

TABLE 2.16 – Métriques des modèles pour la tempête V1

On constate très distinctement que le modèle le plus efficace pour prédire le nombre de sinistres

## CHAPITRE 2. MODÉLISATION DE LA CHARGE SINISTRE D'UNE TEMPÊTE MAJEURE SUR LE PORTEFEUILLE DE GRAA

de la tempête V1 est le modèle de Forêts aléatoires.

**Pour la tempête G1 :**

Métriques	Arbre de décision	Forêts aléatoires	MLG
Accuracy	0,86	0,87	0,84
F1-Score	0,54	0,51	0,53
AUC	0,83	0,70	0,70

TABLE 2.17 – Métriques des modèles pour la tempête G1

En revanche celui qui correspond le mieux aux données de la tempête G1 et le modèle d'arbre de décision. Les modèles entraînés par cette tempête sont à priori beaucoup moins performants que ceux pour la tempête V1, probablement lié au fait que cette tempête a provoqué un nombre plus important de sinistres (3055 contre 802) et comporte plus d'assurés exposés (20000 contre 4000), ce qui rend les probabilité de sinistres pour les assurés plus aléatoires.

### Cohérence des modèles

Pour vérifier la cohérence de nos modèles, un nouveau test est effectué, c'est-à-dire que l'on va appliquer nos modèles sur l'entièreté des bases correspondantes aux tempêtes V1 et G1 et comparer le nombre de sinistres par département prédit par les modèles (somme des probabilités associées aux assurés) au nombre de sinistres réel (somme de la variable *Presence*).

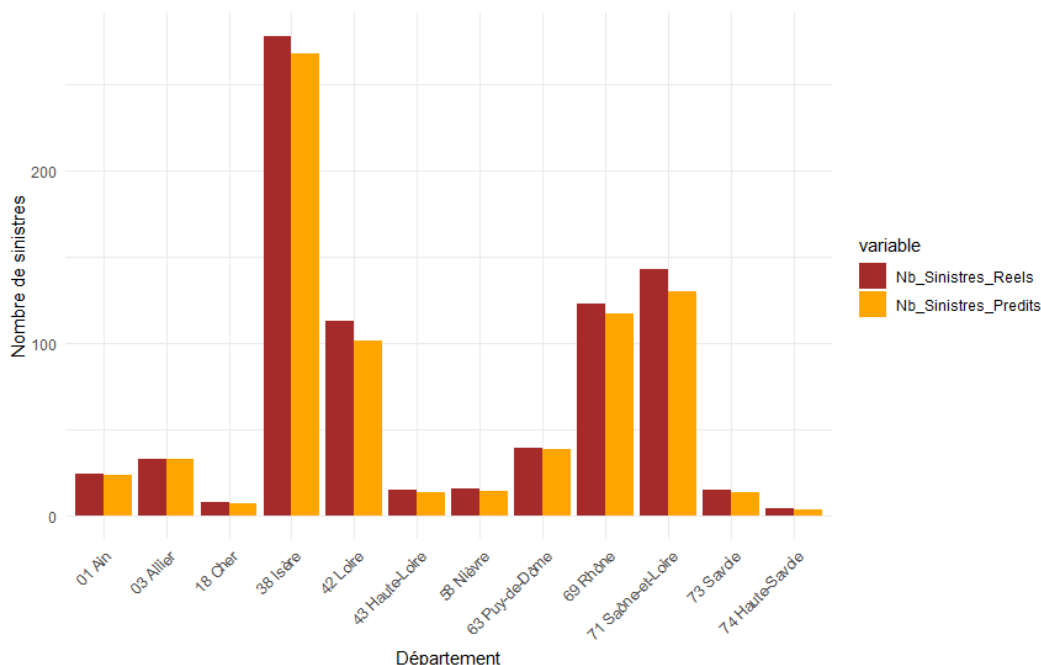


FIGURE 2.25 – Comparaison des nombres de sinistres réels et prédits par département pour la tempête V1

## CHAPITRE 2. MODÉLISATION DE LA CHARGE SINISTRE D'UNE TEMPÊTE MAJEURE SUR LE PORTEFEUILLE DE GRAA

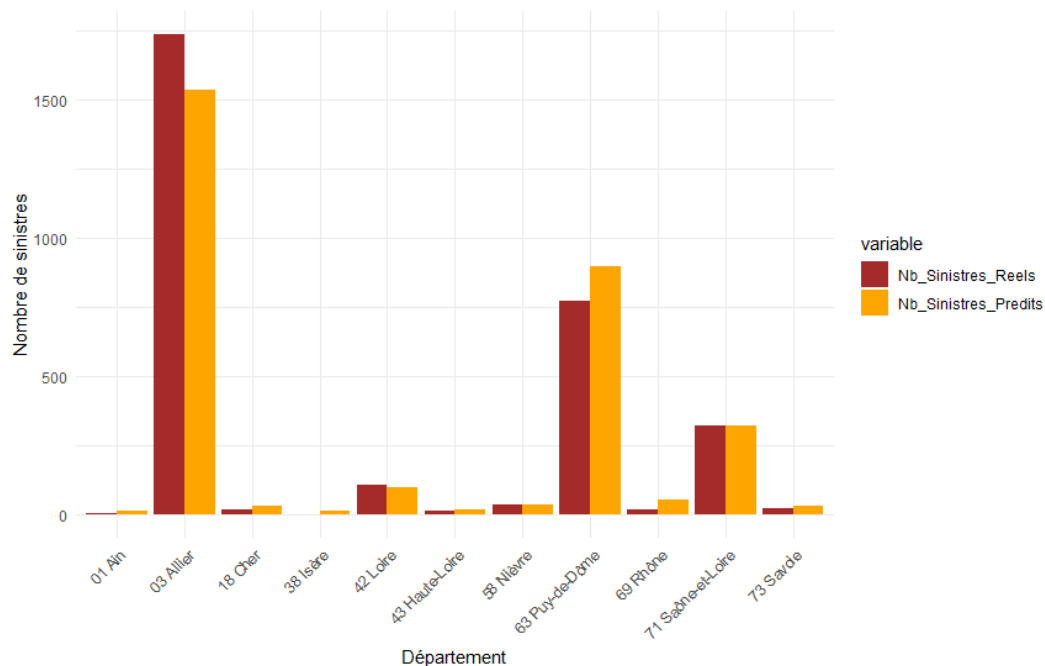


FIGURE 2.26 – Comparaison des nombres de sinistres réels et prédits par département pour la tempête G1

Que ce soit au niveau de la tempête V1 ou G1, les modèles sous-estiment très légèrement le nombre de sinistres. Cependant on peut noter que les valeurs suivent de manière très correcte celles pour chaque département. Les résultats sont donc satisfaisant surtout en notant que la variable liée aux départements n'a pas été intégrée aux modèles en temps que variable explicative puisque le but de cette étude est de pouvoir appliquer nos modèles à d'autres tempêtes que celles servant à entraîner les modèles. C'est pourquoi seules les caractéristiques liées aux profils d'assurés et des habitations ainsi que les conditions météo auxquelles ils ont été exposés ont été retenues.

### Résultats du modèle de fréquence sur d'autres tempêtes

Voyons désormais si nos modèles sont performants pour prédire le nombre de sinistres de tempêtes différentes de celles avec lesquelles ils ont été entraînés.

Pour cela, deux autres bases de données d'assurés exposés à des tempêtes ont été extraites. Une correspondant à la tempête de grêle survenue le 21 juin 2022 provoquant 16,6M€ de sinistres en gardant un seuil de 5% de sinistrés. Et une tempête liée au vent datant du 22 au 24 février 2024 dont la charge sinistres est de 1,4M€. Les proportions de sinistrés pour ces tempêtes qui seront notées G2 et V2 sont respectivement de 15,7% et 7,7%.

Phénomène	Date	Notation	Nombre d'assurés exposés	Nombre de sinistres	Montant de sinistres
Grêle	21 juin 2022	G2	8011	1259	16,6M€
Vent	22-24 février 2024	V2	9775	757	1,4M€

TABLE 2.18 – Présentation des tempêtes V2 et G2

Testons donc les modèles formés sur ces tempêtes V2 et G2 :

## CHAPITRE 2. MODÉLISATION DE LA CHARGE SINISTRE D'UNE TEMPÊTE MAJEURE SUR LE PORTEFEUILLE DE GRAA

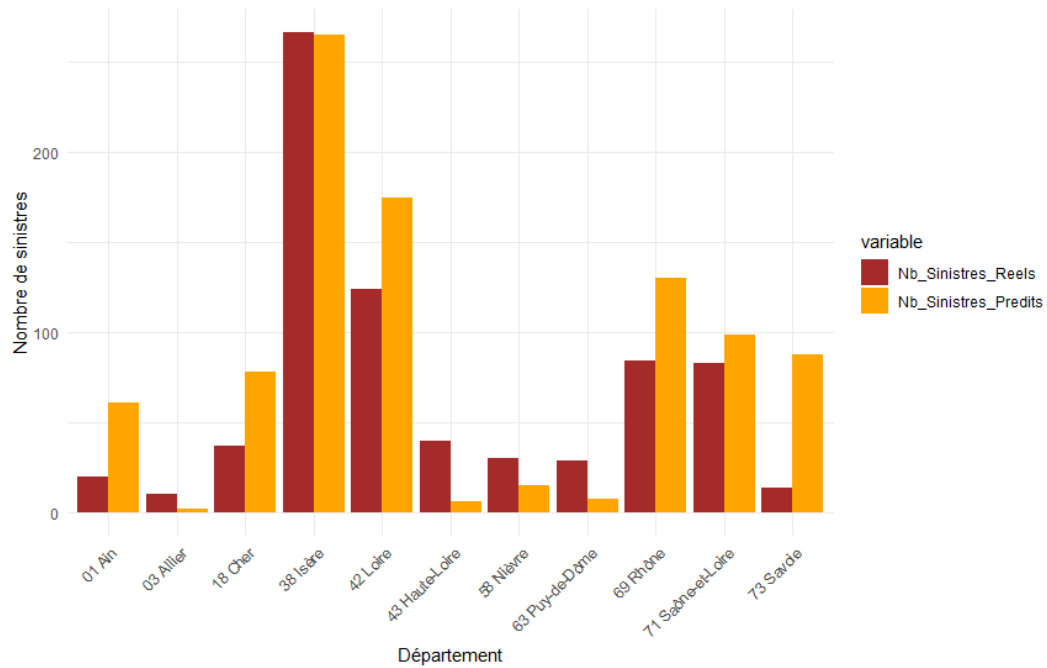


FIGURE 2.27 – Nombre de sinistres par département réels et prédits pour la tempête V2

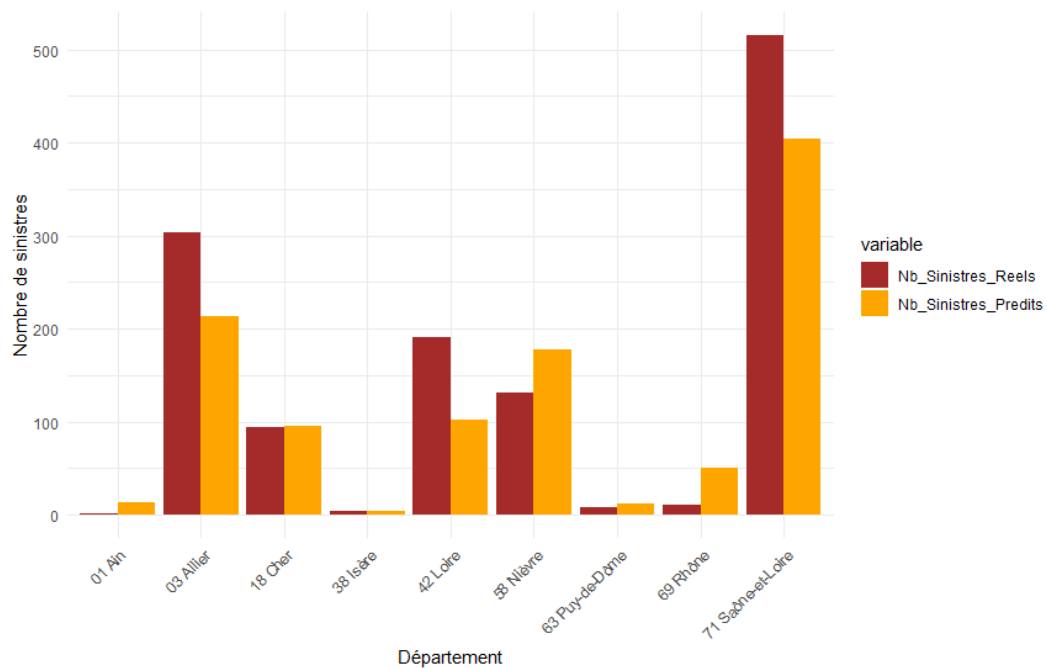


FIGURE 2.28 – Nombre de sinistres par département réels et prédits pour la tempête G2

Les résultats globaux sont rassemblés dans ce tableau :

## CHAPITRE 2. MODÉLISATION DE LA CHARGE SINISTRE D'UNE TEMPÊTE MAJEURE SUR LE PORTEFEUILLE DE GRAA

Tempête	Nombre Sinistres Réel	Nombre Sinistres Prédit
G2	1259	1072
V2	757	914

TABLE 2.19 – Présentation des résultats des modèles sur les tempêtes V2 et G2

Les résultats des prédictions sur ces deux tempêtes sont satisfaisant puisqu'ils suivent globalement bien les valeurs réelles, avec malgré tout, à l'image des tests sur les tempêtes V1 et G1, une sous-estimation du nombre de sinistres pour la modélisation sur la tempête G2. Cependant on observe également une sur-estimation au niveau de la tempête V2, ce qui écarte l'hypothèse que le modèle sous-estime systématiquement les valeurs.

### 2.3 Estimation de la charge sinistre totale des tempêtes

#### 2.3.1 Estimation du cout moyen des sinistres

Les charges prédites pour chaque tempête seront calculées avec la formule suivante.

$$\text{Charge totale prédite} = \sum_{i=1}^n P(\text{Presence}_i = 1) \times \mathbb{E}[\text{Coût}_i | X_i]$$

- $n$  représente le nombre total d'assurés dans la base de données.
- $P(\text{Presence}_i = 1)$  est la probabilité que l'assuré  $i$  soit sinistré lors de la tempête (ces probabilités ont été calculées dans la partie précédente)
- $\mathbb{E}[\text{Coût}_i | X_i]$  désigne le coût moyen des sinistres déclarés par les assurés possédant les mêmes caractéristiques  $X$  que l'assuré  $i$

La prochaine étape sera donc de déterminer les variables de nos bases de données qui ont une influence significative sur les coûts des sinistres. Pour cela, analysons les coûts moyens associés aux critères de chaque variable :

## CHAPITRE 2. MODÉLISATION DE LA CHARGE SINISTRE D'UNE TEMPÊTE MAJEURE SUR LE PORTEFEUILLE DE GRAA

Variables explicatives	Tempête V1	Tempête G1
Inter_vent		
Inter_precip		
Classe_grelons		
nb_pieces		

TABLE 2.20 – Tableau comparatif des effets des variables sur les coûts moyens des tempêtes



## CHAPITRE 2. MODÉLISATION DE LA CHARGE SINISTRE D'UNE TEMPÊTE MAJEURE SUR LE PORTEFEUILLE DE GRAA

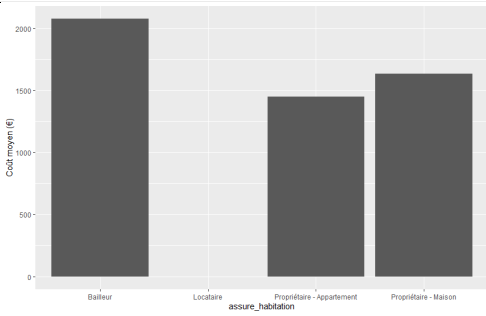
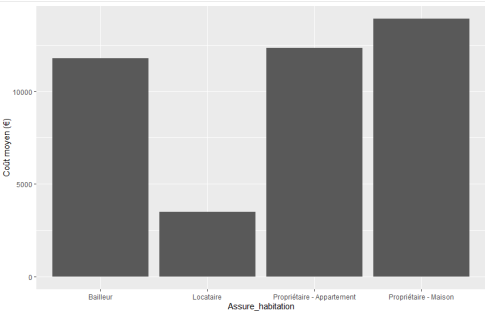
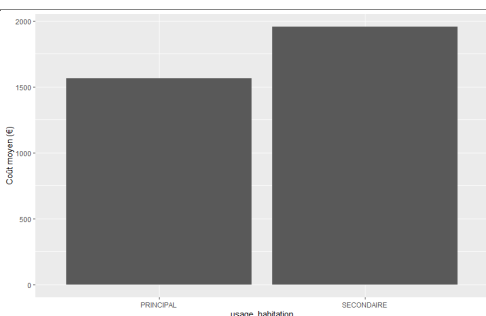

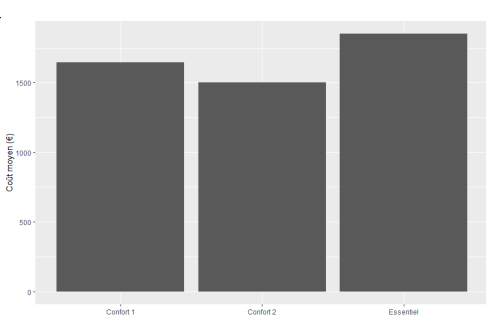
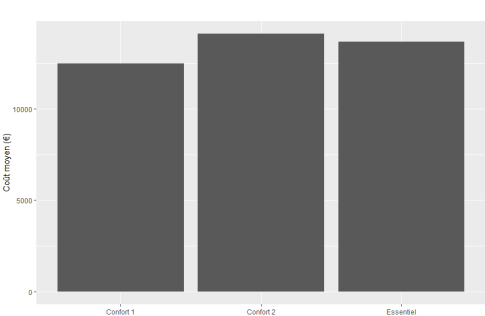
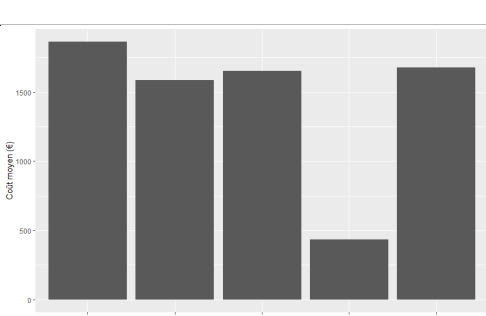
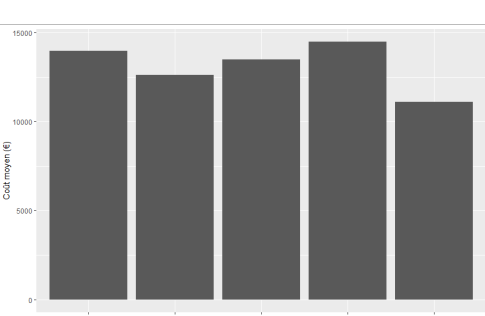
Variables explicatives	Tempête V1	Tempête G1
assure_habitation		
usage_habitation		
formule		
periode_const_bat		

TABLE 2.21 – Tableau comparatif des effets des variables sur les coûts moyens des tempêtes

Ces tableaux permettent de visualiser les variables qui ont une influence sur les coût des sinistres pour les deux tempêtes :

— **Inter\_precip** : les montants augmentent bien avec les quantités de précipitations recensées,

## CHAPITRE 2. MODÉLISATION DE LA CHARGE SINISTRE D'UNE TEMPÊTE MAJEURE SUR LE PORTEFEUILLE DE GRAA

avec une nette augmentation à partir de 30mm.

- **nb\_pieces** : on observe assez logiquement que plus le nombre de pièces (qui peut être associé à la superficie de l'habitation) est élevé, plus les sinistres déclarés sont importants. Cette variable donc également jugée explicative.

Celle qui n'influe que sur la tempête lié au vent :

- **Inter\_vent** : En effet la vitesse des rafales de vent ont une incidence sur les coûts de sinistres de la tempête V1. Ce qui n'est en revanche pas le cas pour la tempête G1

Et celle qui n'influe que sur la tempête lié à la grêle :

- **Classe\_grêlons** : La taille des grêlons est également sans surprise un élément qui explique les montants élevés ou non des sinistres de la tempête G2. Notons particulièrement une différence nette entre les coûts moyens associés aux classes A1 et A2 et ceux des classes A3, A4 et A5.

Les analyses graphiques associées aux autres variables de l'étude ne montrent pas une significativité assez prononcée pour que celles-ci soient jugées influentes dans les valeurs des coûts des sinistres. Les coûts moyens des sinistres associés à chaque seront donc calculés à partir de trois variables pour chaque tempête.

Pour chaque assuré  $i$  exposé à la tempête V1 :

$$\mathbb{E}[\text{Coût}_i | X_i] = \text{Mean}(\text{CTP\_ajust} | \text{Inter\_vent} = \text{Inter\_vent}_i \cap \text{Inter\_precip} = \text{Inter\_precip}_i \cap \text{nb\_pieces} = \text{nb\_pieces}_i)$$

Pour chaque assuré  $j$  exposé à la tempête G1 :

$$\mathbb{E}[\text{Coût}_j | X_j] = \text{Mean} \left( \text{CTP\_ajust} \left| \begin{array}{l} \text{Classe\_grêlons} = \text{Classe\_grêlons}_j \\ \cap \text{Inter\_precip} = \text{Inter\_precip}_j \\ \cap \text{nb\_pieces} = \text{nb\_pieces}_j \end{array} \right. \right)$$

Maintenant que les coûts moyens de sinistres ont été associés à chaque profil d'assuré et que les probabilités de survenance des sinistres sont connues. La charge totale estimée pour chaque tempête peut être calculée.

### 2.3.2 Obtention des estimations de charges sinistres

#### Estimation de la charge sinistre des tempêtes d'entraînement

Comme expliqué précédemment, en multipliant pour chaque assuré sa probabilité d'être sinistré et son coût moyen de sinistre, on obtient sa charge de sinistre estimée.

En sommant les charges prédites pour les assurés de chaque département exposé aux tempêtes V1 et G1, les résultats suivants sont obtenus :

## CHAPITRE 2. MODÉLISATION DE LA CHARGE SINISTRE D'UNE TEMPÊTE MAJEURE SUR LE PORTEFEUILLE DE GRAA

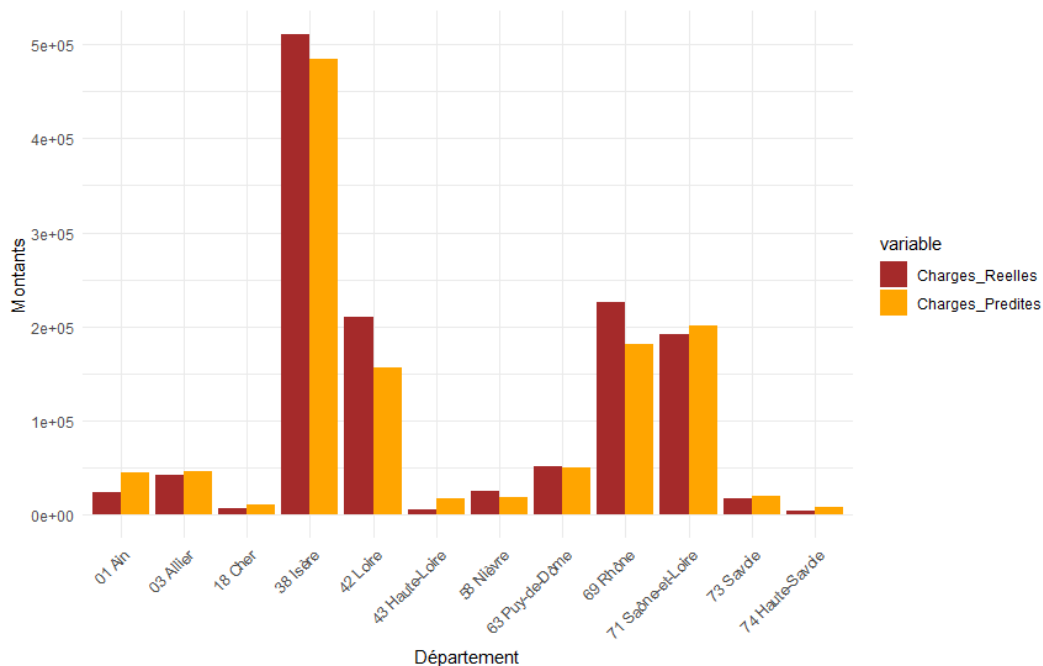


FIGURE 2.29 – Charge de sinistres par département réels et prédits pour la tempête V1

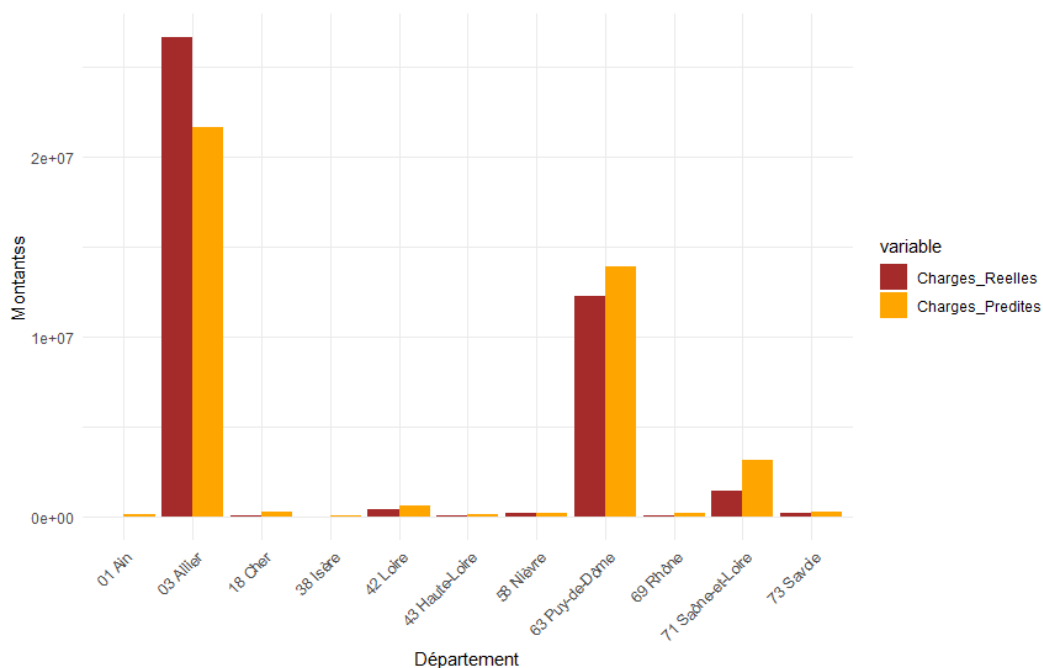


FIGURE 2.30 – Charge de sinistres par département réels et prédits pour la tempête G1

On observe une cohérence assez forte des résultats prédits avec les valeurs réelles. Les écarts observés sont similaires à ceux observés pour les nombres de sinistres, ce qui est logique puisque les coûts moyens ont été calculés à partir des sinistres provoqués par ces deux tempêtes. L'application de ces modèles à des bases d'assurés exposés à d'autres tempêtes permettra de vérifier la pertinence ou non des valeurs recensées.

## CHAPITRE 2. MODÉLISATION DE LA CHARGE SINISTRE D'UNE TEMPÊTE MAJEURE SUR LE PORTEFEUILLE DE GRAA

### Résultats du modèle de charge sur d'autres tempêtes

Les modèles de fréquence de sinistres construits dans la partie précédente associés aux coûts moyens calculés à partir des sinistres des tempêtes V1 et G1 ont été appliqués aux bases d'assurés associées aux tempêtes V2 et G2.

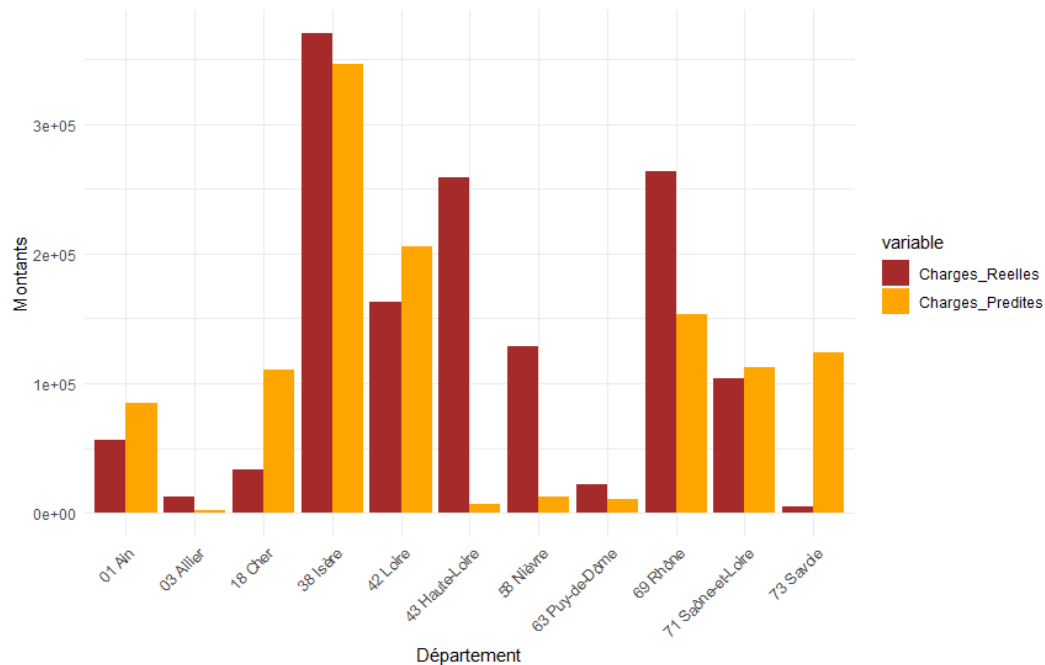


FIGURE 2.31 – Charges de sinistres par département réelles et prédites pour la tempête V2

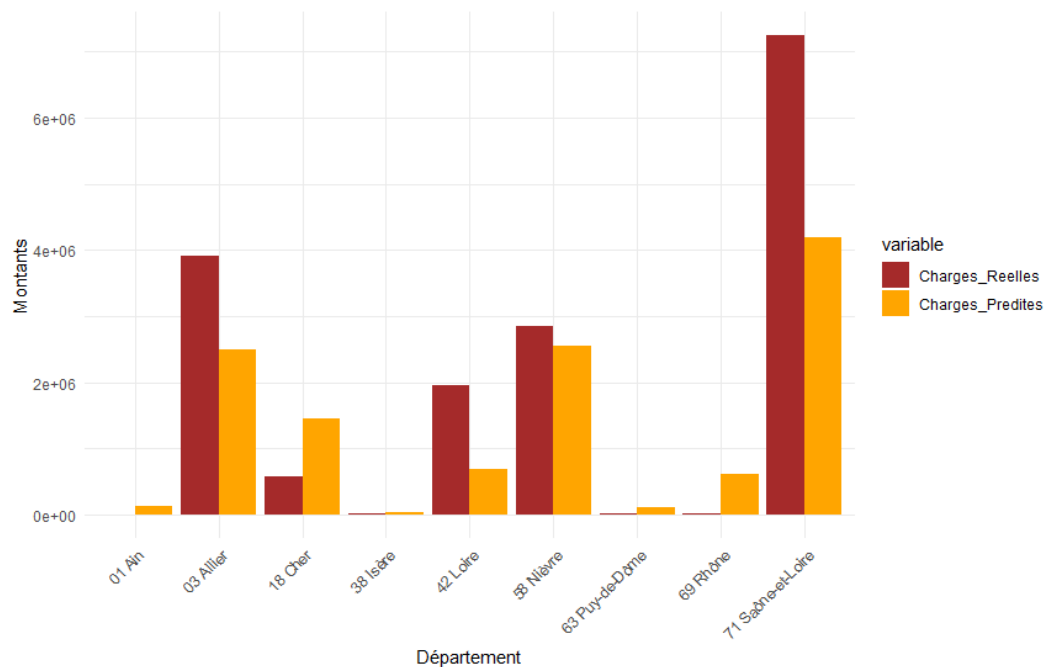


FIGURE 2.32 – Charges de sinistres par département réelles et prédites pour la tempête G2

## CHAPITRE 2. MODÉLISATION DE LA CHARGE SINISTRE D'UNE TEMPÊTE MAJEURE SUR LE PORTEFEUILLE DE GRAA

Les montants prédits ont été rassemblés dans le tableau ci-dessous :

Tempête	Charge Sinistres Réelle	Charge Sinistres Prédite
G2	16,6M€	12,1M€
V2	1,4M€	1,2M€

TABLE 2.22 – Présentation des résultats des modèles de charge sur les tempêtes V2 et G2

Les résultats sont satisfaisants puisque même si les deux modèles sous-estiment les charges totales de sinistres, les valeurs obtenues suivent globalement les ordres de grandeur des valeurs réelles. On observe toutefois des écarts importants au niveau de certains départements (Haute-Loire et Nièvre pour la tempête V2 et Saône-et-Loire pour la tempête G2).

Dans l'annexe de ce mémoire se trouvent des graphiques des charges sinistres réelles et prédites pour ces tempêtes en fonction de variables explicatives dans les modèles (la vitesse de vent pour la tempête V2 et la taille de grêlons pour la tempête G2). Ceux-ci montrent que les écarts obtenus sur la tempête V2 proviennent principalement des assurés exposés à de faibles vitesses de vent dont le modèle sous-estime le nombre et la charge des sinistres. Quant à la sous-estimation du modèle prédisant la tempête G2, elle provient surtout des assurés exposés à de grosses tailles de grêlons.

### 2.4 Incertitudes liées aux modèles : Méthode Bootstrap

Les modèles développés dans cette partie sont destinés à évaluer l'ampleur de la charge sinistre de tempêtes, qui sont des événements par nature rares et fortement variables, ce qui rend la prévision incertaine.

Il est donc essentiel de quantifier rigoureusement l'incertitude entourant les prévisions produites par un modèle. La méthode bootstrap s'impose comme un outil particulièrement adapté dans ce contexte, notamment en raison de sa souplesse et de son faible niveau d'exigence en hypothèses distributionnelles.

#### 2.4.1 Principe de la méthode

##### Processus

La méthode bootstrap repose sur une stratégie de rééchantillonnage avec remise à partir d'un échantillon de données observées, dans le but de reproduire la variabilité inhérente aux données. Contrairement aux approches paramétriques, qui reposent sur une distribution théorique des erreurs (souvent supposée normale), le bootstrap permet d'approcher empiriquement la distribution de l'estimateur ou de la variable d'intérêt.

Le processus générique est le suivant :

1. On ajuste un modèle prédictif (régression linéaire, arbre de décision, forêts aléatoires, etc.) sur l'échantillon de données historiques.
2. On en extrait les résidus ou les bruits d'erreur entre les observations réelles et les valeurs prédites.
3. On génère un grand nombre d'échantillons bootstrap en rééchantillonnant ces résidus (ou éventuellement les données entières selon le type de modèle).

## CHAPITRE 2. MODÉLISATION DE LA CHARGE SINISTRE D'UNE TEMPÊTE MAJEURE SUR LE PORTEFEUILLE DE GRAA

4. Pour chaque échantillon bootstrap, on recalibre le modèle, puis on prédit la charge sinistre de la tempête simulée.
5. On construit une distribution empirique des prévisions, à partir de laquelle on peut dériver des intervalles de confiance (à 95% par exemple)

### Intérêts

L'application de la méthode Bootstrap présente de nombreux avantages :

- La flexibilité : la méthode s'adapte à tout modèle prédictif, y compris non linéaire ou non paramétrique.
- Le réalisme empirique : elle repose sur les données réellement observées, sans hypothèses fortes sur la distribution des erreurs.
- Une évaluation complète : elle permet de quantifier l'incertitude totale de la prévision (paramétrique + aléatoire).

Mais en revanche ne permet pas d'effacer certaines limites :

- La dépendance aux données historiques : si les données passées ne sont pas représentatives de la diversité des tempêtes futures, le bootstrap peut sous-estimer l'incertitude.
- La sous-estimation de l'aléa extrême : le rééchantillonnage ne permet pas de générer des sinistres plus extrêmes que ceux observés dans l'échantillon.
- Le coût computationnel : l'approche nécessite de recalculer le modèle un grand nombre de fois, ce qui peut être coûteux pour des modèles complexes.

### 2.4.2 Application aux modèles obtenus

Dans le cadre de cette étude, la méthode bootstrap est appliquée pour quantifier l'incertitude autour de l'estimation de la charge sinistrée d'un portefeuille d'assurés exposés à des tempêtes majeures, en s'appuyant sur des données issues d'un événement précédent similaire (tempêtes V1 et G1).

### Modèles de fréquence

Les modèles de fréquence sont d'abord ajustés sur les bases d'apprentissage (tempête V1 et G1) pour prédire, pour chaque assuré, la probabilité  $P(\text{Presence}_i = 1)$  d'être sinistré en fonction de ses caractéristiques (type d'habitation, usage, date de construction, intensité du vent, quantité de précipitations, etc.).

Ce modèle est ré-estimé à chaque itération bootstrap afin de capturer la variabilité liée à l'apprentissage du modèle. Les modèles utilisés sont l'arbre de décision (tempête G1) et la forêt aléatoire (tempête V1), qui permet d'estimer des probabilités individuelles sans hypothèse paramétrique forte.

### Estimations des coûts moyens

Le coût moyen en cas de sinistre n'est pas fixé a priori. Il est estimé empiriquement, à partir des sinistres survenus dans les échantillons bootstrapés de V1 et G1, en calculant un coût moyen

## CHAPITRE 2. MODÉLISATION DE LA CHARGE SINISTRE D'UNE TEMPÊTE MAJEURE SUR LE PORTEFEUILLE DE GRAA

par groupe homogène d'assurés (défini par exemple selon le nombre de pièces, la vitesse du vent, les précipitations...).

En cas d'absence de sinistre dans un groupe donné pour un échantillon bootstrap (ce qui rend le coût inestimable), un coût de référence préalablement calculé sur l'ensemble de V1 est utilisé comme substitution.

Ainsi pour chaque assuré du portefeuille associé à la tempête dont on souhaite estimer le coût :

- On simule la survenue d'un sinistre selon une loi de Bernoulli de paramètre  $p = P(\text{Presence}_i = 1)$ , estimé à une itération bootstrap  $b$
- Si un sinistre est simulé, un coût moyen lui est affecté selon son groupe.
- La charge totale simulée est alors obtenue par agrégation des coûts individuels simulés

### 2.4.3 Obtention d'intervalles de confiance

Comme expliqué précédemment, le Bootstrap permet d'obtenir une moyenne de charge sinistres estimée ainsi qu'un intervalle de confiance à 95% en prenant comme borne les quantiles à 2,5 et 97,5% de toutes les charges obtenues lors du Bootstrap.

#### Modèle associé aux tempêtes de vent

En appliquant cette méthode au portefeuille d'assurés exposés à la tempête V2, on obtient une charge moyenne estimée à 1,38M€ ainsi qu'un intervalle de confiance à 95% de [1,21 ; 1,57].

Pour rappel, la charge obtenue sans l'utilisation du Bootstrap était de 1,24M€, qui est situé donc dans la partie basse de l'intervalle de confiance.

La charge de sinistres réelle recensée pour la tempête V2 (1,42M€) est elle aussi dans l'intervalle de confiance.

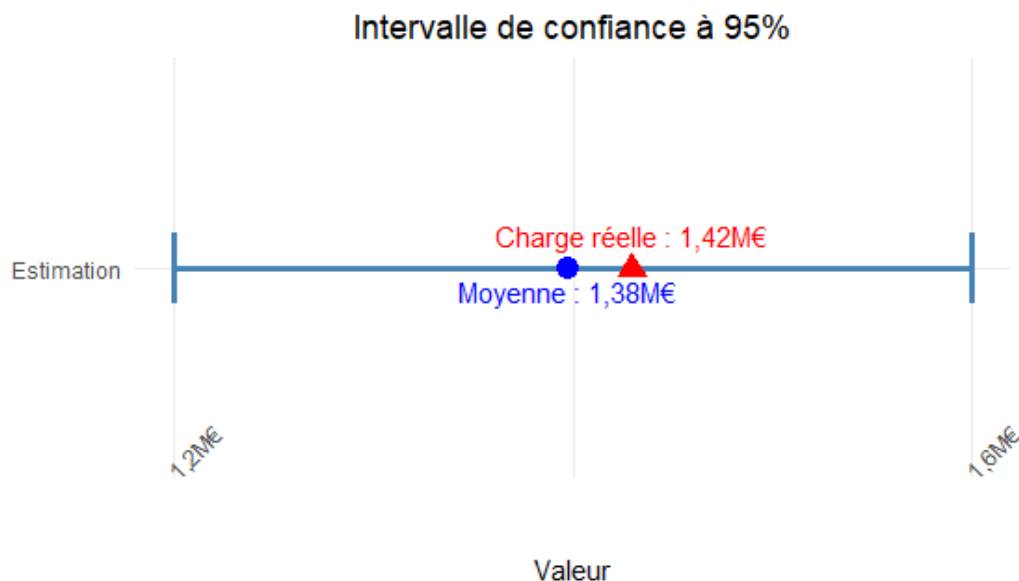


FIGURE 2.33 – Intervalle de confiance associé à la tempête V2

## CHAPITRE 2. MODÉLISATION DE LA CHARGE SINISTRE D'UNE TEMPÊTE MAJEURE SUR LE PORTEFEUILLE DE GRAA

Il en ressort que la méthode bootstrap donne une valeur estimée plus proche de la réalité qu'en appliquant simplement les modèles de fréquence et de coût moyen sur le portefeuille d'assurés.

### Modèle associé aux tempêtes de grêles

Voyons si l'on observe la même remarque avec le portefeuille d'assurés exposés à la tempête G2. La charge moyenne estimée atteint 16,8M€ et l'intervalle de confiance à 95% donne [11,6M€ ; 22,5M€]. Pour rappel, la charge obtenue sans l'utilisation du Bootstrap était de 12,1M€, qui est donc également situé dans la partie basse de l'intervalle de confiance.

La charge de sinistres réelle recensée pour la tempête G2 (16,6M€) est elle aussi dans l'intervalle de confiance.

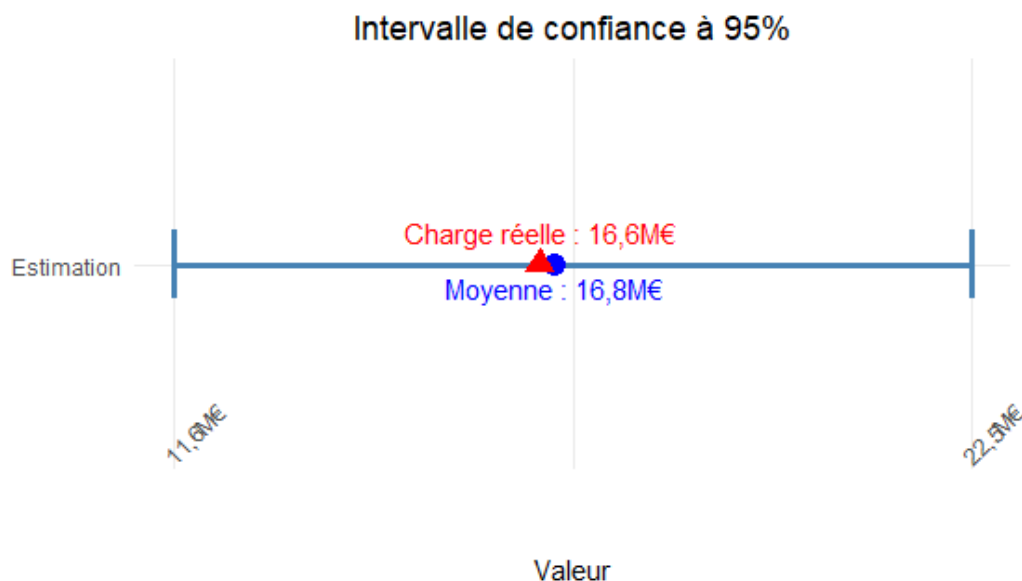


FIGURE 2.34 – Intervalle de confiance associé à la tempête G2

Ce graphique donne lieu à 2 remarques.

Dans un premier temps, nous pouvons voir que la charge de sinistres obtenue est encore une fois bien plus proche de la réalité que la charge obtenue sans la méthode bootstrap.

Dans un second temps, on observe un intervalle de confiance à 95% plutôt large puisque la borne supérieure représente près du double de la borne inférieure, ce qui signifie que les incertitudes concernant les modèles de fréquence et de coût moyen sont élevées. Ceci met en évidence les limites de ces modèles présentées précédemment que la méthode bootstrap ne permet pas d'effacer.

### 2.4.4 Limites de la méthode

Au cours de cette étude, plusieurs axes et choix qui ont été faits peuvent expliquer les écarts observés ainsi que les incertitudes élevées lors de la prédiction de charge sinistre :

- **Extraction des données météorologiques**

Un relevé plus précis des vitesses de vent, des quantités de précipitations et des classes est un



## CHAPITRE 2. MODÉLISATION DE LA CHARGE SINISTRE D'UNE TEMPÊTE MAJEURE SUR LE PORTEFEUILLE DE GRAA

premier point qui permettrait d'améliorer la qualité des modèles. En effet, malgré l'ajustement des données en fonction de la distance des sites météo avec les lieux de survenance des sinistres, une incertitude quant à la qualité de ces données persiste et cette incertitude se répercute sur les modèles.

- **Historique de tempêtes faible :**

En effet les tempêtes de vent et de grêle jugées comme majeures représentent des aléas climatiques récents, la base de données de GRAA contient donc peu d'évènements, ce qui constitue un facteur d'augmentation de l'incertitude au niveau des résultats. Les tempêtes à venir constitueront donc des possibilités d'enrichir l'entraînement des modèles et donc de les rendre plus efficace.

- **Sélection des meilleurs modèles :**

Lors de la construction des modèles de fréquence des sinistres, le choix qui a été fait est de retenir pour chaque tempête le modèle qui présentait les meilleurs métriques lorsque l'on testait les modèles sur les 20% de nos bases mis de côté. Ce choix a permis d'obtenir de très bon résultats en retestant les modèles sur les bases des assurés exposés aux tempêtes sur lesquelles ils avaient été entraînés et testés mais a renforcé le potentiel sur-ajustement des modèles et ainsi diminuer le caractère adaptatif de ces modèles à d'autres tempêtes. Par exemple, la méthode nous a amené à sélectionner le modèle d'arbre de décision pour la tempête G1 alors que le risque de sur-ajustement sur ce type de modèle est bien plus important que sur la forêt aléatoire ou la régression logistique.

- **Estimation des coûts moyens à partir des tempêtes historiques :**

Une dernière limite de cette méthode réside dans l'attribution du coût moyen d'un sinistre pour chaque catégorie d'assuré. En effet seuls 3 critères pour chacun des modèles ont été jugés significatifs par rapport au coût d'un sinistre, la recherche d'autres critères pourrait être une solution pour affiner les groupes d'assurés et ainsi réduire l'incertitude liée à ces estimations.

## Conclusion

Pour conclure sur ce chapitre proposant une méthode d'estimation de la charge sinistre d'une tempête dès lors que les données relevant de celle-ci sont connues, cette estimation a été réalisée en combinant des modèles de prédiction de la fréquence des sinistres et des coûts moyens des sinistres, basés sur les caractéristiques des assurés et des conditions météorologiques. Les résultats ont montré une bonne cohérence entre les valeurs réelles et les valeurs estimées, malgré une légère sous-estimation globale. Les modèles appliqués aux tempêtes différentes de celles utilisées pour l'entraînement ont également donné des résultats satisfaisants, confirmant la robustesse et la pertinence des méthodes utilisées. La modélisation de la fréquence des sinistres a été particulièrement efficace avec les forêts aléatoires pour les tempêtes de vent, et les arbres de décision pour les tempêtes de grêle. Quant à l'estimation des coûts moyens des sinistres, les analyses ont permis de déterminer les variables influentes, telles que la vitesse du vent, la quantité de précipitations, et la taille des grêlons. Ces variables ont été intégrées dans les modèles pour affiner les prédictions et obtenir des résultats plus précis. L'application des modèles à d'autres tempêtes a validé leur efficacité, bien que l'intervalle de confiance à 95% élevé montre que les incertitudes liées à ces modèles doivent encore être réduites. En approfondissant cette étude au niveau de certains points comme une extraction plus précise des

## CHAPITRE 2. MODÉLISATION DE LA CHARGE SINISTRE D'UNE TEMPÊTE MAJEURE SUR LE PORTEFEUILLE DE GRAA

données météorologiques, elle peut servir de base de travail dans des perspectives d'amélioration des processus de gestion du risque tempête et grêle chez Groupama Rhône-Alpes Auvergne.

## Chapitre 3

# Outils de protection contre les tempêtes majeures

### 3.1 Outils de Réassurance

#### 3.1.1 Différents traités de réassurance

La réassurance est un mécanisme juridique qui repose sur un contrat conclu entre une compagnie d'assurance, appelée cédante, et une entreprise de réassurance, appelée cessionnaire. Ce contrat instaure un échange entre les parties impliquant :

- Le versement d'une prime par l'assureur cédant, appelée "prime cédée".
- Une indemnisation, par le réassureur, d'une partie des sinistres subis par l'assureur.

L'objectif principal de la réassurance est de réduire la volatilité des résultats financiers et de protéger les fonds propres des compagnies d'assurance en transférant une partie du risque à des réassureurs.

Les différents types de traités de réassurance se classent en deux grandes catégories : la réassurance proportionnelle et la réassurance non proportionnelle.

#### Réassurance Proportionnelle

Dans le cadre de la réassurance proportionnelle, l'assureur cède au réassureur une partie prédéfinie de ses primes et, en contrepartie, le réassureur prend en charge la même proportion des indemnisations. Cette catégorie comprend principalement deux formes :

Le Traité en Quote-Part (QP) : L'assureur transfère un pourcentage fixe des primes d'un portefeuille donné au réassureur, qui s'engage à rembourser le même pourcentage des sinistres. Ce type de réassurance est fréquemment utilisé pour couvrir les risques de catastrophes naturelles.

Le Traité en Excédent de Capitaux ou de Pleins (XP) : Un taux de cession est déterminé pour chaque police d'assurance lorsque le capital assuré dépasse un seuil défini (le "plein de conservation"). Ce mécanisme est particulièrement adapté aux contrats d'assurance décès et aux assurances indivi-

## CHAPITRE 3. OUTILS DE PROTECTION CONTRE LES TEMPÊTES MAJEURES

duelles contre les accidents.

### Réassurance Non Proportionnelle

Dans un traité de réassurance non proportionnelle, l'intervention du réassureur ne dépend pas d'un partage fixe des primes et sinistres, mais plutôt d'un seuil de déclenchement. Trois principaux types de réassurance non proportionnelle existent :

**L'Excédent de Sinistres (XS) :** Le réassureur intervient uniquement pour les sinistres dont le coût excède un seuil défini (priorité) et jusqu'à une limite supérieure (plafond). Par exemple, un traité "30 XS 50" signifie que l'assureur prend en charge les 50 premiers millions d'euros de sinistres et que le réassureur couvre les 30 millions suivants, jusqu'à un plafond de 80 millions d'euros. Ce type de traité est couramment utilisé pour la responsabilité civile et certains événements climatiques. Dans cet exemple le "30" est appelé la capacité de tranche, c'est-à-dire que le réassureur prend en charge la tranche entre 50 et 80 millions (soit un maximum de 30 millions).

**L'Excédent de Perte ou Stop-Loss (SL) :** Fonctionnant sur le même principe que l'excédent de sinistres, ce mécanisme repose sur un ratio technique (S/C) au lieu de montants fixes. Il est principalement utilisé pour couvrir les risques cycliques tels que les catastrophes naturelles.

Pour minimiser le risque de défaillance des réassureurs, les traités sont souvent diversifiés et répartis entre plusieurs sociétés de réassurance.

### Clauses Spécifiques des Traités de Réassurance

Les contrats de réassurance incluent plusieurs clauses permettant d'ajuster les modalités d'intervention :

- **Clause Obligatoire :** Définit la priorité et la portée du traité (sur quels risques porte le traité), précisant si l'intervention se fait par risque ou par événement.
- **Clause Aggregate :** Comprend deux sous-types :
  - Clause de franchise annuelle : L'assureur conserve une partie des sinistres au-delà de la priorité pour limiter le coût de la réassurance.
  - Clause de plafond annuel : Fixe le montant maximum d'engagement du réassureur.
- **Clause de Reconstitution :** : Prévoit la reconstitution de la couverture après un sinistre important, moyennant une prime complémentaire.
- **Clause de Participation Bénéficiaire :** : L'assureur peut recevoir une part des bénéfices si le solde du traité est excédentaire.
- **Clause de "No Claim Bonus" :** : Une réduction de prime est appliquée si aucune indemnisation n'a été versée pendant l'année.

#### 3.1.2 Formules de réassurance à GRAA

La réassurance est un domaine crucial dans la gestion des risques pour les caisses régionales comme Groupama Rhône-Alpes Auvergne (GRAA). En particulier, la réassurance des risques liés aux tempêtes et à la grêle est essentielle pour garantir la stabilité financière de la caisse et assurer une couverture adéquate face à des phénomènes naturels imprévisibles.

## CHAPITRE 3. OUTILS DE PROTECTION CONTRE LES TEMPÊTES MAJEURES

### Pool de rétrocession

Le pool de rétrocession chez Groupama Rhône-Alpes Auvergne est conçu pour répartir de manière optimale la part des risques nécessitant une réassurance externe importante entre les caisses régionales et la caisse nationale. Voici les points clés :

- **Objectif** : Augmenter la capacité de conservation globale des caisses régionales et de la caisse nationale en répartissant les risques de manière optimale.
- **Composition** : Il existe deux types de pools : le pool permanent de rétrocession et le pool des risques souscrits par la caisse nationale.
  - Pool permanent de rétrocession : Concerne les risques nécessitant une réassurance externe significative, comme les risques climatiques. Les cotisations et sinistres rétrocédés sont nets de la part des protections externes. Cette réassurance est proportionnelle, les caisses régionales transfèrent un pourcentage de leur prime reçu qui sera noté  $R\%$  contre une indemnisation à hauteur de ce même pourcentage sur les montants de sinistres nets de la réassurance externe.
  - Pool des risques souscrits par la caisse nationale : Permet la participation des caisses régionales aux résultats des risques souscrits ou acceptés au plan national.
- **Conditions de rétrocession** : Elles varient selon les branches d'assurance (par exemple, auto, RC des professionnels, tempêtes, etc.) et sont spécifiquement détaillées en fonction des risques et des protections externes en vigueur.
- **Rétrocession spécifique** : La caisse nationale peut être amenée à souscrire des risques spécifiques et les rétrocède ensuite aux caisses régionales selon des modalités précises.

En résumé, le pool de rétrocession chez GRAA permet une gestion communautaire et solidaire des risques, en optimisant les capacités de réassurance interne et externe des caisses régionales et de la caisse nationale.

### Risques tempêtes-grêle-neige

La réassurance des risques liés aux tempêtes et à la grêle est structurée principalement autour de protections en excédent de sinistre par événement et par risque. Cette approche permet de limiter l'exposition financière de GRAA en cas de survenance de ces phénomènes naturels.

Les chiffres réels présents sur les contrats de réassurance signés par GRAA étant confidentiels, des notations seront utilisées dans la suite de cette partie.

Le schéma général de réassurance pour les risques de tempête et de grêle est le suivant :

- **Excédent de sinistre par événement** : Ce type de protection intervient à partir de seuils d'intervention définis en amont (que l'on notera  $S_{e1}$ ,  $S_{e2}$  et  $S_{e3}$ ) et couvre le cumul des sinistres occasionnés par un même événement. Le premier de ces seuils d'intervention est un pourcentage des cotisations acquises brutes des branches Incendie et Tempête.

Les tranches de cession seront notées de la manière suivante :

- $T_{e1}\%$  pour les coûts entre le seuils  $S_{e1}$  et  $S_{e2}$ .

### CHAPITRE 3. OUTILS DE PROTECTION CONTRE LES TEMPÊTES MAJEURES

- $T_{e_2}\%$  pour les coûts entre les seuils  $S_{e_2}$  et  $S_{e_3}$ .
- $T_{e_3}\%$  pour les coûts au delà du seuil  $S_{e_3}$ .

• **Excédent de sinistre par risque** : Cette protection fonctionne uniquement par sinistre et par risque, couvrant les dommages occasionnés à un seul emplacement assuré par une seule police. Les seuils d'intervention pour les protections en excédent de sinistre par risque seront notés  $S_{r_1}, S_{r_2}$  et  $S_{r_3}$ .

Les tranches de cession seront notées de la manière suivante :

- $T_{r_1}\%$  pour les sinistres entre les seuils  $S_{r_1}$  et  $S_{r_2}$ .
- $T_{r_2}\%$  pour les sinistres entre les seuils  $S_{r_2}$  et  $S_{r_3}$ .
- $T_{r_3}\%$  pour les sinistres au-delà du seuil  $S_{r_3}$ .

Voici un schéma extrait de la *Convention de Réassurance 2024*, qui synthétise les conditions de réassurance des risques tempêtes-grêles chez GRAA :

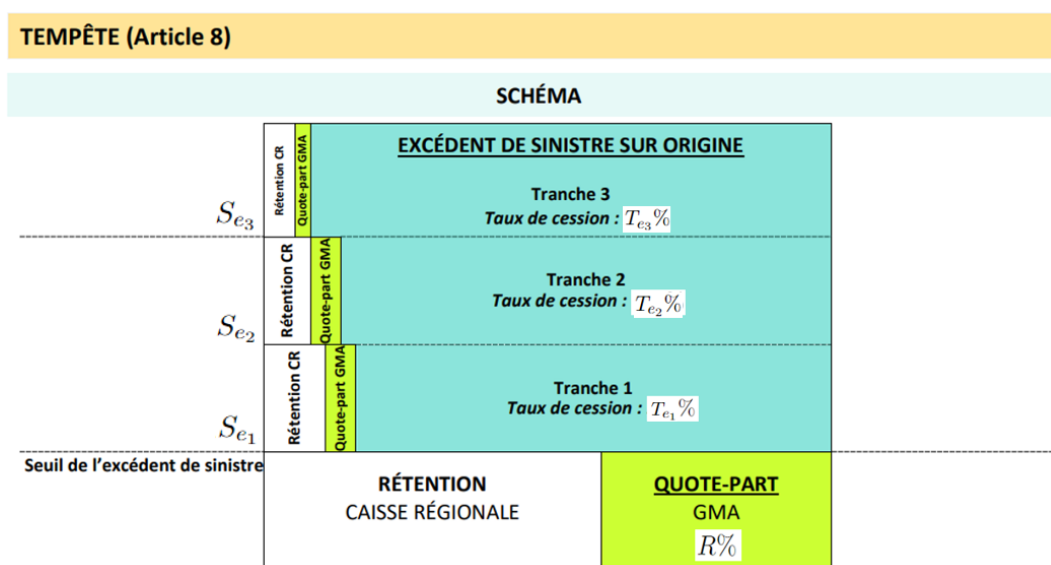


FIGURE 3.1 – Schéma présentant la réassurance du risque tempête à GRAA

On observe bien dans ce graphique la combinaison de la réassurance interne, une réassurance proportionnelle avec la Quote-Part cédée à la caisse nationale qui intervient sur toutes les tranches de l'excédent de sinistre par évènement (réassurance externe), qui constitue une réassurance non-proportionnelle puisque les taux de cessions différent en fonction des tranches de coûts de sinistres.

Ces différents types de réassurance sont détaillés ci-dessous :

## CHAPITRE 3. OUTILS DE PROTECTION CONTRE LES TEMPÊTES MAJEURES

PARAMÈTRES D'APPLICATION 2024 - Caisses métropolitaines						
1/ EXCÉDENT DE SINISTRE PAR RISQUE SUR ORIGINE						
TRANCHE 1	Seuil (€) :	$S_{r_1}$	Plafond (€) :	$S_{r_2}$	Part cédée :	$T_{r_1}\%$
TRANCHE 2	Seuil (€) :	$S_{r_2}$	Plafond (€) :	$S_{r_3}$	Part cédée :	$T_{r_2}\%$
TRANCHE 3	Seuil (€) :	$S_{r_3}$	Plafond (€) :	(*)	Part cédée :	$T_{r_3}\%$
(*) Dans la limite des plafonds de garantie de chaque contrat						
2/ EXCÉDENT DE SINISTRE PAR ÉVÈNEMENT SUR ORIGINE						
SEUIL $S_{e_1}$	des cotisations acquises brutes année N-1 des branches Incendie + Attentats + Tempête (voir page suivante : seuils monétaires 2024)					
TRANCHE 1	Seuil :	$S_{e_1}$	Plafond :	$S_{e_2}$	Part cédée :	$T_{e_1}\%$
TRANCHE 2	Seuil :	$S_{e_2}$	Plafond :	$S_{e_3}$	Part cédée :	$T_{e_2}\%$
TRANCHE 3	Seuil :	$S_{e_3}$	Plafond :	-	Part cédée :	$T_{e_3}\%$
TAUX DE COTISATION	Au total dont protections externes dont rémunération des fonds propres					
ASSIETTE DES COTISATIONS	Cotisations acquises brutes des branches Incendie (yc Attentats) et Tempête multipliées par le Coefficient de vulnérabilité (voir page suivante)					
3/ QUOTE-PART SUR RÉTENTIONS						
TAUX DE CESSION QUOTE-PART			$R\%$			
(à appliquer aux cotisations acquises origine de la branche Tempête nettes d'excédent de sinistre Tempête)						
Taux de commission brute						
Taux de frais GMA (base 100%)						
TAUX DE COMMISSION NET						
(à appliquer à la cotisation de quote-part brute)						

FIGURE 3.2 – Tableau présentant la réassurance du risque tempête à GRAA

### 3.1.3 Application aux charges de sinistres des tempêtes modélisées

#### Application du système de réassurance de GRAA

Utilisons les quatre tempêtes pour lesquelles ont été estimés les charges de sinistres totales pour voir le passage entre le coût brut (avant application de la réassurance) et le coût net (après application de la réassurance) de ces événements.

Pour obtenir les graphiques ci-dessous, les vrais taux de cession et les vrais seuils d'excédent de sinistre ont été utilisés, c'est pourquoi les montants nets ne sont pas montrés mais cela donne néanmoins un aperçu de l'effet de la réassurance sur les coûts réels des événements pour GRAA.

### CHAPITRE 3. OUTILS DE PROTECTION CONTRE LES TEMPÊTES MAJEURES

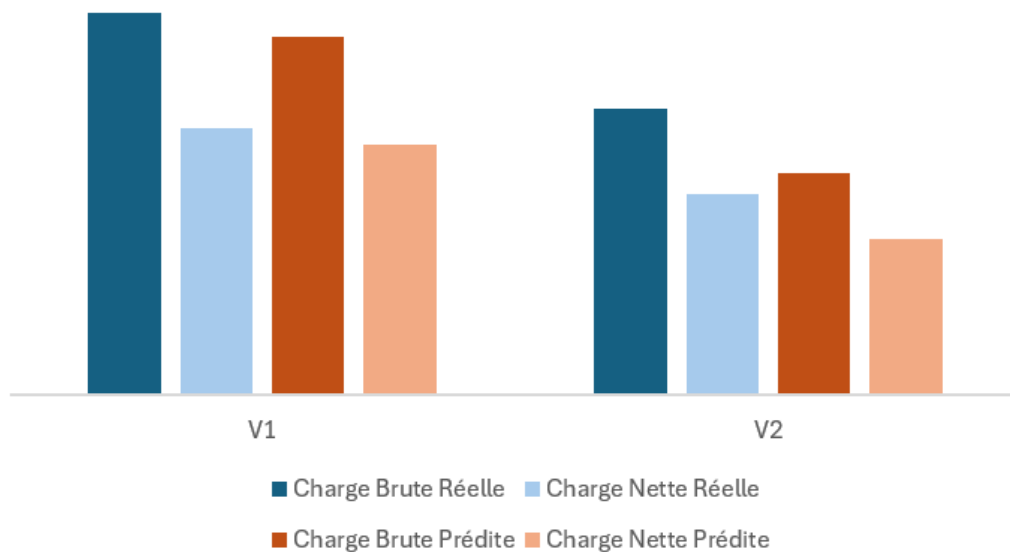


FIGURE 3.3 – Charges brutes et nettes des tempêtes V1 et V2

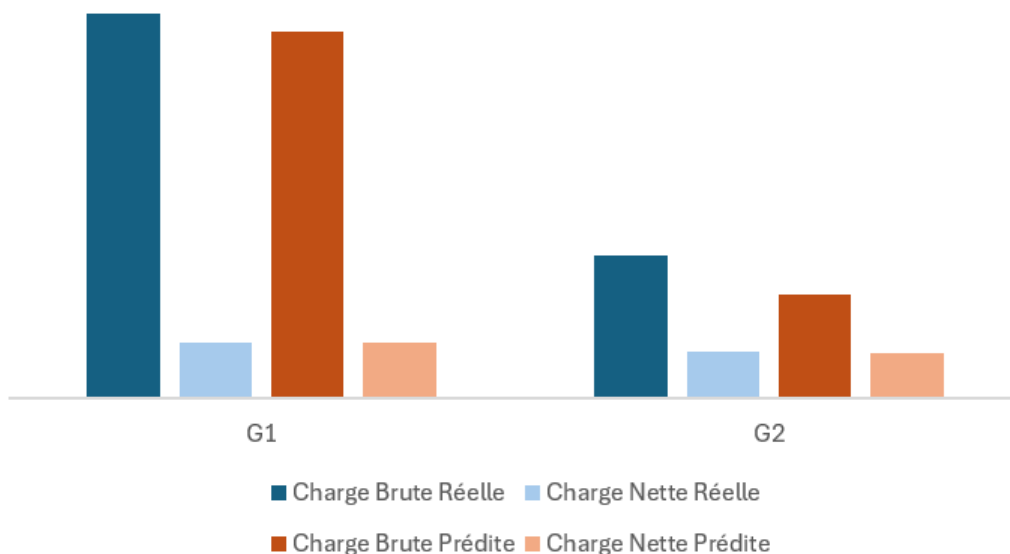


FIGURE 3.4 – Charges brutes et nettes des tempêtes G1 et G2

L'analyse de ces graphiques montrent la forte influence de l'application du système de réassurance sur les coûts des sinistres pour un événement tempête, en particulier pour les événements avec des charges sinistres très importantes comme les tempêtes G1 et G2 (les deux images n'étant pas à la même échelle). En effet, les taux de réassurance sont plus élevés pour les tranches hautes de réassurance. Le deuxième qui peut être remarqué est la réduction des écarts lorsque l'on passe de la charge brute à la nette entre les valeurs réelles et prédites (encore une fois en particulier pour les grosses tempêtes). A noter que ces montants nets ne sont pas totalement les charges versées par GRAA puisqu'une partie des sinistres du pool de rétrocession (à l'intérieur duquel  $R\%$  des coûts des sinistres liés aux tempêtes



## CHAPITRE 3. OUTILS DE PROTECTION CONTRE LES TEMPÊTES MAJEURES

ont été versés) reviendra à la charge de la caisse régionale.

### 3.2 Outils de prévention

Les tempêtes et la grêle sont des phénomènes météorologiques pouvant entraîner des dommages matériels importants. Pour aider ses assurés à mieux anticiper ces aléas climatiques et ainsi limiter ces risques, Groupama Rhône-Alpes Auvergne met en place plusieurs outils de prévention. Ces dispositifs s'articulent autour de trois axes principaux : un système d'alertes météorologiques, des programmes d'incitations et de réduction de primes et des actions d'éducation et de sensibilisation.

#### 3.2.1 Les alertes météorologiques

##### Qu'est-ce qu'une alerte météo ?

Il est important de différencier l'alerte météo de la vigilance météorologique. La vigilance météo, mise en place en 2001, est une information publique diffusée par Météo-France afin d'anticiper l'arrivée de phénomènes météorologiques dangereux. Elle permet aux autorités et aux habitants d'évaluer les risques potentiels et de prendre les précautions nécessaires.

En revanche, une alerte météo est déclenchée par les pouvoirs publics lorsqu'un phénomène météorologique représente un danger immédiat. Cette alerte vise à informer la population des mesures à adopter pour assurer sa sécurité et limiter les dégâts matériels.

##### Le service Alerte Météo de Groupama

Pour aider ses assurés à mieux anticiper les aléas climatiques, GRAA propose un service d'alerte météo en partenariat avec Predict Services. Ce dispositif permet aux adhérents de recevoir des notifications par SMS ou e-mail en cas d'intempéries imminentes (quelques heures avant en général).

Grâce à ce service, les assurés bénéficient d'une information rapide et personnalisée, ce qui leur permet d'adopter les bons réflexes face aux risques météorologiques. Ce dispositif profite donc à la fois aux assurés et à l'entreprise, qui enregistre un nombre et des coûts de sinistres moins importants. En complément, un lien vers un espace dédié est fourni pour obtenir des conseils de prévention adaptés à la situation .

##### Les situations nécessitant une alerte météo

Les alertes météo concernent différents types de phénomènes climatiques pouvant engendrer des dommages :

- Vents violents : risque de chutes d'arbres et d'objets, coupures d'électricité.
- Risques de grêle : danger accru selon les saisons.
- Grand froid : gel intense pouvant impacter la circulation et les infrastructures.
- Inondations : risque de débordement de cours d'eau et dégâts sur les habitations.

## CHAPITRE 3. OUTILS DE PROTECTION CONTRE LES TEMPÊTES MAJEURES

Les niveaux de vigilance météorologique Météo-France utilise un système de vigilance à quatre niveaux pour informer la population de la gravité des phénomènes à venir :

- Vert : pas de vigilance particulière.
- Jaune : il faut être attentif, notamment en cas de pratique d'activités extérieures.
- Orange : vigilance importance, des phénomènes dangereux sont prévus.
- Rouge : vigilance absolue, des intempéries exceptionnelles sont attendues.

Les alertes sont déclenchées à partir du niveau orange, un pictogramme spécifique indique la nature du danger (vent violent, pluie, grêles, canicule, etc.).

### 3.2.2 Programme d'incitation et de réduction de primes

En plus d'alerter les assurés lorsqu'un aléa climatique se prépare, Groupama a mis en place un système qui permet aux assurés d'obtenir des avantages ou des réductions de primes s'ils mettent en place des dispositifs réduisant le risque de sinistres en cas de tempête ou de grêles.

#### Exemple de dispositif : La bâche anti-grêle

Groupama collabore avec HB Collection, une entreprise spécialisée dans les housses de protection, pour offrir des solutions de protection contre les aléas climatiques, y compris les orages de grêle. Les housses anti-grêle sont faciles à installer et se posent en quelques minutes. Elles sont conçues pour protéger les véhicules contre les impacts de grêle et les petits débris.

Dans le cadre de ce partenariat, les assurés peuvent ainsi prétendre à deux types de réductions :

- Une réduction sur le prix des bâches : les sociétaires de GRAA bénéficient de 15% de réduction sur les housses auto anti-grêle.
- La franchise offerte : l'assuré possédant la bâche subit un sinistre.



FIGURE 3.5 – Bâche anti-grêle auto HB Collection

Ce type de partenariat profite donc aux trois parties : l'assuré qui protège ses biens tout en obtenant

## CHAPITRE 3. OUTILS DE PROTECTION CONTRE LES TEMPÊTES MAJEURES

des avantages monétaires, la compagnie d'assurance qui réduit sa charge sinistre en cas de tempête et l'entreprise qui offre ce service profitant de l'exposition de GRAA pour se développer.

### 3.2.3 Education et sensibilisation

La prévention passe également par une meilleure information des assurés sur les risques climatiques et les gestes à adopter en cas de tempête ou de grêle. Groupama Rhône-Alpes Auvergne met en place plusieurs actions de sensibilisation, destinées aussi bien aux particuliers qu'aux professionnels :

- Programmes de prévention agricole : GRAA met en œuvre des programmes de sensibilisation aux risques climatiques, y compris les tempêtes et la grêle, pour les assurés agricoles. Ces programmes incluent des conseils pratiques et des mesures de prévention pour minimiser les dommages.
- Utilisation de la technologie : GRAA utilise des outils de visualisation des risques pour aider les assurés à comprendre le niveau de risque auquel leurs biens sont exposés. Par exemple, des cartes et des scores de risque sont calculés pour orienter les efforts de prévention et de tarification.
- Communication proactive : GRAA envoie régulièrement des communications par email pour rappeler aux assurés l'importance de la prévention et les informer des bonnes pratiques à adopter. Par exemple, des emails sur la cybersécurité et d'autres risques sont envoyés pour sensibiliser les collaborateurs et les assurés.

## Conclusion

Pour conclure sur cette partie concernant les dispositifs mis en place par GRAA pour se protéger contre le risque tempête-grêle, deux catégories de dispositifs ont été détaillés. La première concernant la partie purement financière avec le recours à un système de réassurance pour limiter l'impact des événements climatiques exceptionnels. La deuxième s'appuyant sur différents dispositifs d'informations transmis aux assurés, qui permettent de réduire à la fois la probabilité de sinistralité et les coûts des sinistres lorsque qu'une tempête survient. Ces dispositifs ont également une utilité pour les assurés puisqu'ils se protéger et protéger leurs biens plus efficacement et ainsi profiter de réductions sur les cotisations ou les franchises.

# Conclusion Générale

En conclusion, ce mémoire sur la maîtrise du risque Tempête-Grêle au sein de Groupama Rhône-Alpes Auvergne met en lumière l'importance croissante des risques naturels dans le contexte actuel de changements climatiques. Les événements climatiques extrêmes, tels que les tempêtes et la grêle, représentent des défis majeurs pour les compagnies d'assurance, nécessitant des stratégies de gestion et des outils de modélisation robustes pour anticiper et atténuer leurs impacts. Tout au long de cette étude, plusieurs aspects cruciaux ont été abordés.

En premier lieu, une analyse détaillée des pratiques de gestion des risques chez Groupama Rhône-Alpes Auvergne a été réalisée. Cette analyse a permis notamment de comprendre les limites que posent les processus d'estimation actuels des charges sinistres des tempêtes et grêles qui sont jugées comme des événements à sinistralités élevées. C'est pourquoi la construction d'une autre méthode a été entreprise.

La deuxième partie de l'étude s'est donc concentrée sur cette construction d'une méthode alternative de modélisation des coûts des sinistres pour les tempêtes de grande envergure, en tenant compte :

- Des spécificités du portefeuille d'assurance de Groupama Rhône-Alpes Auvergne.
- De l'historique de sinistralité au niveau du risque tempête-grêle.
- Des données climatiques propres à cette région.

L'utilisation de modèles de classification tels que les forêts aléatoires et les arbres de décision s'est avérée efficace pour prédire la fréquence des sinistres, tandis que l'analyse de l'influence de chaque variable sur les coûts moyens des sinistres a permis de prédire un montant total de sinistres pour une tempête donnée. La méthode bootstrap a quant à elle permis de recalibrer les modèles et d'obtenir des intervalles de confiance. Les résultats obtenus ont montré une bonne cohérence entre les valeurs réelles et les valeurs estimées, bien que des ajustements puissent encore être effectués pour affiner les prédictions puisque la méthode "fréquence fois coûts moyens" utilisée ainsi que les différents choix faits pour mettre en place ces modèles présentent des limites.

Enfin, il a été constaté que la gestion de ces risques pouvait également se faire en amont de l'arrivée des phénomènes climatiques, avec en particulier l'impact fort du système de réassurance sur les événements de grandes envergures. Cet outil permettant de limiter leur coût net pour GRAA et assurer ainsi la solvabilité de la caisse régionale. La compagnie joue le rôle d'assureur mais également de préventeur face aux risques climatiques en interagissant directement avec les assurés, à l'aide d'alertes SMS sur les événements à venir, de sensibilisation et de programmes d'incitation à se protéger au mieux face à ces risques.

En somme, ce mémoire constitue une contribution significative à la compréhension et à la gestion des risques climatiques dans le secteur de l'assurance. En abordant les défis posés par les tempêtes et

### CHAPITRE 3. OUTILS DE PROTECTION CONTRE LES TEMPÊTES MAJEURES

la grêle, et en proposant des solutions pratiques, il ouvre la voie à une gestion plus résiliente et efficace des risques naturels, assurant la pérennité et la stabilité financière des compagnies d'assurance face aux aléas climatiques de plus en plus fréquents et sévères.

Au-delà des points d'améliorations mentionnés dans ce rapport, les modèles d'estimation de la charge sinistre des tempêtes sur les habitations peuvent servir de base pour l'estimation des montants pour d'autres types de contrats concernant des bâtiments comme les collectivités ou les entreprises.

# Bibliographie

- [1] France Assureurs. *Assurance des catastrophes naturelles*. Disponible à : <https://www.franceassureurs.fr/assurance-protège-finance-et-emploi/assurance-protège/assurance-en-pratique-pour-les-particuliers/assurance-des-catastrophes-naturelles/>.
- [2] Ministère de la Transition écologique et de la Cohésion des territoires. *Site des statistiques du développement durable*. Disponible à : <https://www.statistiques.developpement-durable.gouv.fr/>.
- [3] Météo France. *Vents violents et tempêtes : quelles sont les différences ?*. Disponible à : <https://meteofrance.com/actualites-et-dossiers/comprendre-la-meteo/le-vent/vents-violents-et-tempetes-queelles-sont-les-differences#:~:text=La%20formation%20des%20temp%C3%Aates%20Une%20temp%C3%Aate%20est%20toujours,en%20altitude%2C%20en%20particulier%20de%20l%27%C3%A9tat%20du%20courant-jet..>
- [4] Légifrance. *Exposé des motifs du projet de loi n° 400*. Disponible à : [https://www.legifrance.gouv.fr/dossierlegislatif/JORFDOLE000025625784/?detailType=EXPOSE\\_MOTIFS&detailId=](https://www.legifrance.gouv.fr/dossierlegislatif/JORFDOLE000025625784/?detailType=EXPOSE_MOTIFS&detailId=).
- [5] ORGANISATION MÉTÉOROLOGIQUE MONDIALE. *Atlas de la mortalité et des pertes économiques dues à des phénomènes météorologiques, climatiques et hydrologiques extrêmes (1970-2019)*. 2021.
- [6] France ASSUREUR. *Impact du changement climatique 2050*. Octobre 2021.
- [7] Covéa. *Livre blanc : Risques Climatiques*. Février 2022. Disponible à : [https://www.covea.com/sites/default/files/2023-11/202202\\_Livre\\_Blanc\\_Cov%C3%A9a-Risques\\_Climatiques.pdf](https://www.covea.com/sites/default/files/2023-11/202202_Livre_Blanc_Cov%C3%A9a-Risques_Climatiques.pdf).
- [8] rroupama Rhône-Alpes Auvergne. *Convention de Réassurance 2024*. Avril 2024
- [9] LOPES, David. *Risque Tempête : Modélisation de la sinistralité et impact tarifaire*. Mémoire d'actuariat, 2015.
- [10] Météo France. *Plateforme de données météorologiques*. Disponible à : <https://meteo.data.gouv.fr/>.
- [11] Gouvernement Français. *Plateforme des données publiques françaises*. Disponible à : <https://www.data.gouv.fr/>.
- [12] INSEE. *Statistiques locales*. Disponible à : <https://statistiques-locales.insee.fr/>.

# Annexe

Nombres de sinistres réels et prédits tempêtes V2 et G2

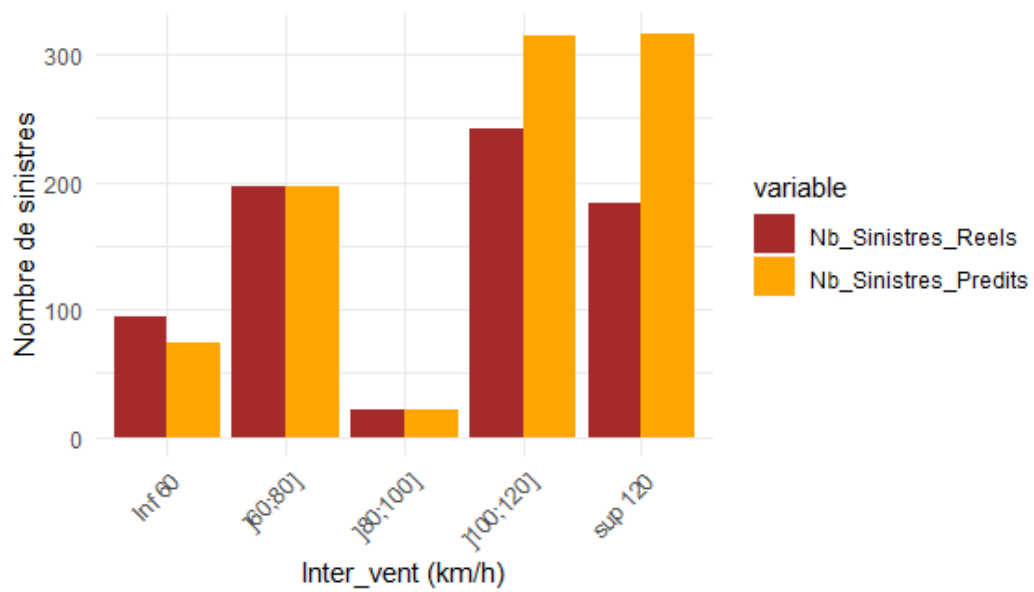


FIGURE 3.6 – Nombres de sinistres par vitesses de vent réels et prédits pour la tempête V2

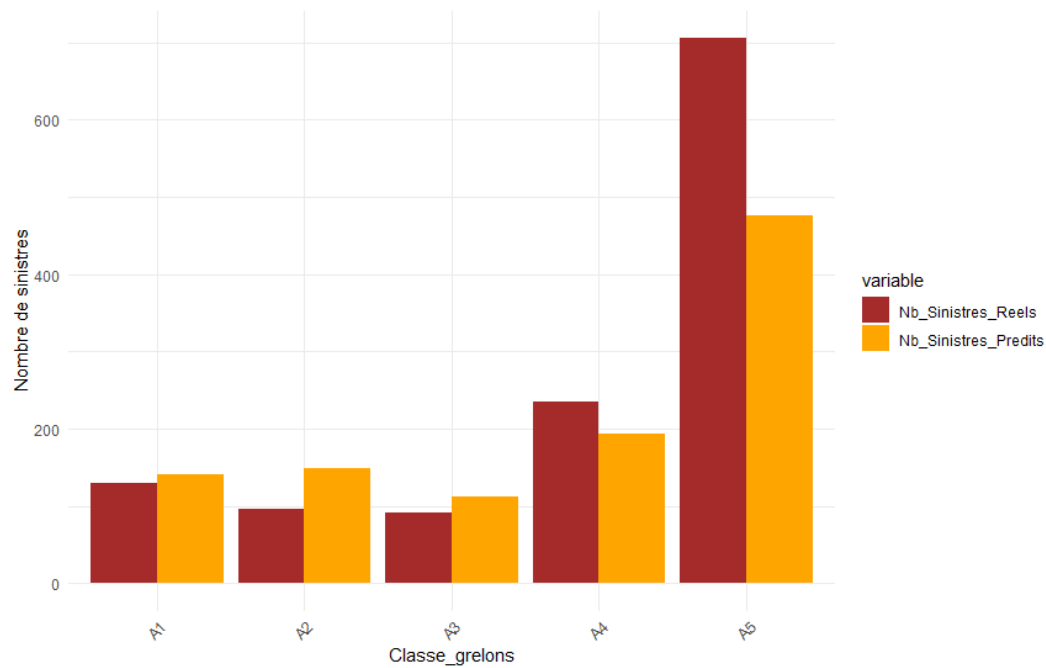


FIGURE 3.7 – Nombres de sinistres par taille des grêlons réels et prédits pour la tempête G2

Charges de sinistres réelles et prédites tempêtes V2 et G2

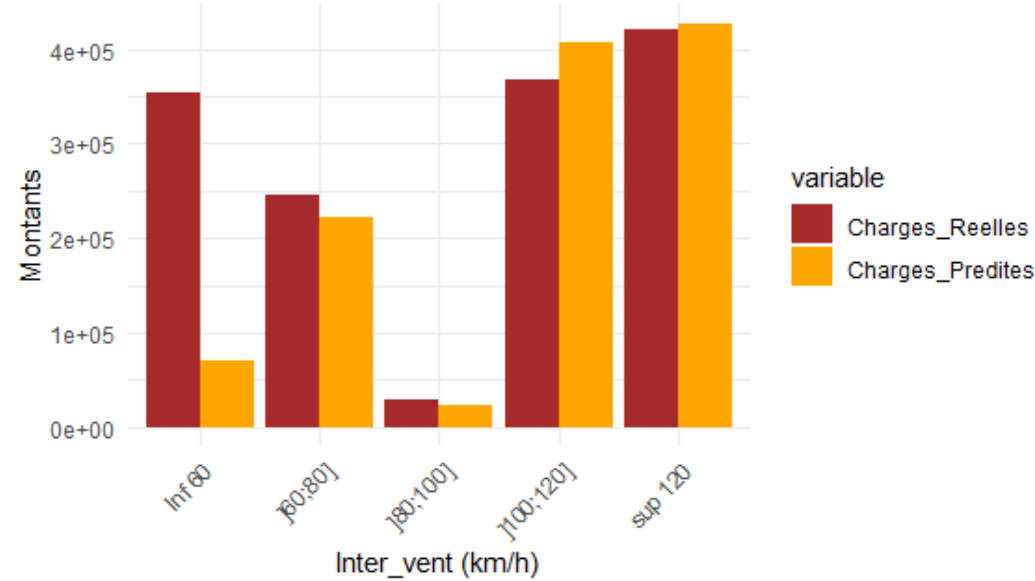


FIGURE 3.8 – Charges de sinistres par vitesses de vent réelles et prédites pour la tempête V2



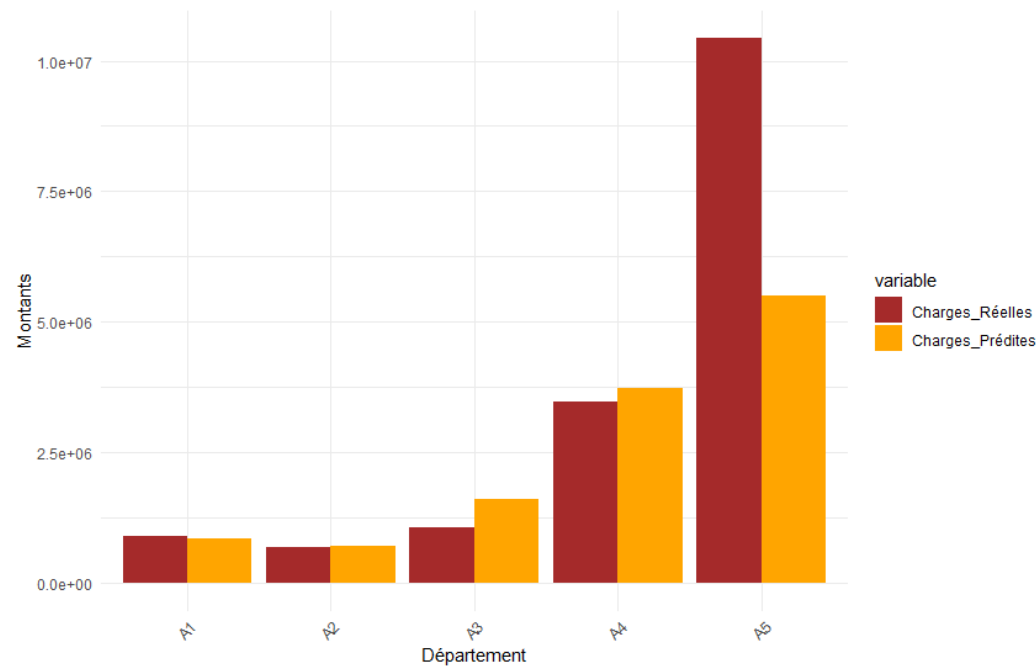


FIGURE 3.9 – Charges de sinistres par taille des grêlons réelles et prédites pour la tempête G2