

**Mémoire présenté devant le Centre d'Etudes Actuarielles
pour la validation du
cursus à la Formation d'Actuaire du Centre d'Etudes Actuarielles
et l'admission à l'Institut des Actuaires
le 29 janvier 2015**

Par : LOPES David

Titre : Risque Tempête : Modélisation de la sinistralité et impact tarifaire.

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

*Membre présent du jury de
l'Institut des Actuaires :*

*Membres présents du jury du
Centre d'Études Actuarielles :*

Secrétariat :

Bibliothèque :

Entreprise :

Nom : MAAF Assurances

Signature : 

Directeur de mémoire en entreprise :

Nom : DE BARROS Fabien

Signature : 

Invité :

Nom : _____

Signature :

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

Signature du responsable entreprise



Signature(s) du candidat(s)



Résumé

Ce mémoire porte sur la modélisation du risque tempête affectant le produit Multirisque Habitation en France métropolitaine en dissociant l'aléa sur l'estimation de charge lors de la survenance d'une tempête et l'aléa sur la probabilité d'occurrence. Ces éléments contribuent à répondre aux besoins de COVEA pour la gestion des sinistres et l'estimation des coûts dans le cadre des arrêtés des comptes, ainsi qu'aux besoins de MAAF pour connaître le niveau de tarif nécessaire sur les garanties climatiques.

Après avoir exposé les connaissances météorologiques sur ce sujet, des modélisations sur la sinistralité lors de la survenance de tempête (en termes de fréquences et de coûts moyens) ont été proposées en tenant compte de l'intensité du vent, des normes de construction et des caractéristiques du portefeuille. En complément, une mesure de la volatilité de la moyenne a également été apportée. Enfin, pour juger du tarif actuel, des tempêtes sont simulées, en associant les périodes de retour et les coûts des tempêtes historiques. Cela permet d'estimer un coût annuel en intégrant des clauses spécifiques aux traités de cession climatiques, et ensuite de calculer les dotations des provisions pour égalisation.

Abstract

This thesis focuses on storm risk modeling affecting products housing in France by separating uncertainty on estimated cost when it occurs and hazard on the return period of a storm. These elements contribute to COVEA's needs for claims management and cost's estimation in relation to the accounts close, and MAAF's needs to know level required pricing on the climatic guarantee.

After identifying meteorological knowledge, modeling on loss upon the storm's occurrence (in terms of frequency and of average costs) have been proposed taking into account intensity of wind, building standards and portfolio characteristics. Then a measure of volatility around mean was also provided. Finally, to judge current pricing, storms are simulated, having previously associated return periods and costs historical storms. This allows to estimate an annual cost gross and net of reinsurance, and then calculate allocations of equalization provisions.

Remerciements

Je remercie tout d'abord Charles DUMARTINET, Willy LEPANTE, Dominique LEBEAU, Sandrine BOULET, Julien MEDARD et François LAIR pour leur expertise.

Merci également à mes collègues pour leur encouragement.

Et enfin, je tiens à remercier Fabien DE BARROS pour avoir tenu son rôle de directeur de mémoire ainsi que Aurélie LAGRE pour leurs conseils avisés.

NOTE DE SYNTHÈSE

INTRODUCTION

Le risque tempête est un enjeu majeur pour les assureurs. Ils se couvrent en transférant une partie de ce risque aux réassureurs pour protéger leurs fonds propres. Mais en cas de dépassement du plafond de réassurance, les fonds propres sont utilisés pour indemniser les sinistrés et la solvabilité de l'entreprise peut être ainsi remise en cause. Or, il est complexe pour un assureur de mettre en place un modèle propre. Ils peuvent alors avoir recours aux courtiers disposant de données de marché. Ce mémoire décrit ainsi la modélisation du risque tempête, tout d'abord, l'aléa sur l'estimation de charge au moment de la survenance de l'évènement. Il répond aux besoins de COVEA pour la gestion des sinistres et l'estimation des coûts dans le cadre des arrêtés des comptes. Ensuite, ce mémoire porte sur l'aléa de la probabilité que cet évènement survienne. L'objectif est de connaître le niveau du tarif nécessaire sur les garanties climatiques selon les conditions actuelles des traités de réassurance souscrits pour des besoins de MAAF.

Chapitre 1. Les tempêtes : un risque à maîtriser

Le premier chapitre montre l'importance de maîtriser le risque tempête en abordant les connaissances assurantielles et météorologiques. L'exclusion de certains évènements climatiques (tempêtes, grêle, neige) du régime des catastrophes naturelles depuis 1990 nous oblige à bien appréhender les aléas de la nature. Les traités de réassurance ne proposent pas de garantie illimitée pour ces périls contrairement aux catastrophes naturelles. L'enjeu majeur de la solvabilité des assureurs porte alors sur ces risques et plus particulièrement la tempête. Ce risque tempête se caractérise par l'aléa sur l'estimation de la charge lors de sa survenance et par l'aléa sur la période de retour des tempêtes. Les logiciels de marché sont des outils qui agissent comme des « boîtes noires » et ne sont pas adaptés aux risques français. Néanmoins, le principe de leurs modules revient à évaluer le risque tempête sous plusieurs angles :

- le module « aléa » sur la détermination de la vitesse de vent par commune et sur la détermination de la période de retour d'une tempête ;
- le module « enjeu » sur la récolte des caractéristiques du portefeuille d'assurés et des communes ;

- le module « vulnérabilité » combinant les informations de deux modules précédents en déterminant des courbes d'endommagement en fonction de la vitesse de vent ou de la période de retour ;
- le module « résultats » sur les prévisions de nombre et de charge lors de la survenance d'une tempête et sur l'estimation de charges probables annuelles par simulation de tempête.

Ce type d'outil permet de répondre à deux besoins :

- sachant la configuration de la tempête et notre portefeuille, quelles seront le nombre de sinistre à gérer et le coût engendré ;
- sachant notre portefeuille, quelle est notre exposition annuelle au risque.

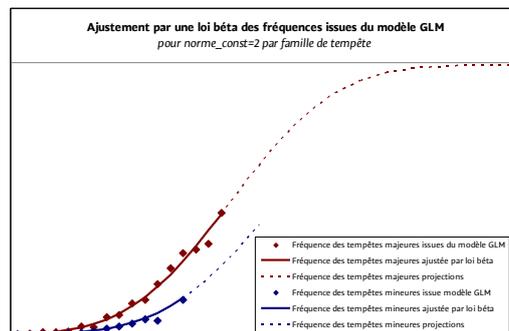
Chapitre 2. Application d'un modèle lors de la survenance d'une tempête

L'objectif de ce chapitre est de répondre au premier besoin à savoir la prévision des nombres de sinistres et des coûts engendrés lors de l'arrivée d'une tempête. Pour ce faire, une modélisation de la sinistralité est réalisée à l'aide des modèles linéaires généralisés.

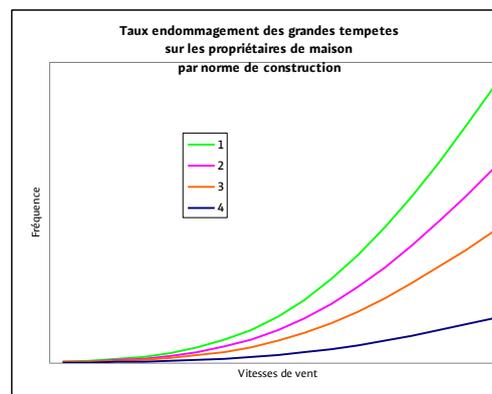
La sinistralité de 6 tempêtes historiques a été analysée par commune à travers les fréquences et les coûts moyens. Les variables analysées sont l'intensité du vent, la structure du portefeuille (qualité juridique, type d'habitat, démographie), la zone de construction ... Ces variables ont été introduites dans un modèle linéaire pour juger de leur pertinence et ainsi exclure certaines variables ou regrouper certaines modalités dans l'optique d'une application pour une prochaine tempête.

Les nombres de sinistres pour une commune dépendent de la qualité juridique (propriétaires, locataires), du type d'habitat (maison, appartement), des normes de constructions de l'habitat assuré, de la densité de la commune, de la vitesse de vent et de la qualification de la tempête. La loi retenue sur le nombre de sinistres est poissonnienne avec pour contrainte le nombre de contrats en portefeuille et pour fonction de lien, le logarithme.

Les paramètres sur les vitesses de vent correspondent à une moyenne des observations et présentent des irrégularités de fréquence (rapport entre le nombre de sinistres et le nombre de contrats) en fonction de l'intensité du vent. C'est pourquoi, ces paramètres ont été ajustés, lissés, selon une loi de la famille exponentielle, permettant de projeter des vitesses de vent non encore enregistrées ou insuffisamment robustes. Pour ce faire, les coefficients des vitesses de vent du modèle GLM sont convertis en fréquences. Ces fréquences sont ensuite ajustées par une loi bêta à l'aide de la méthode des moindres carrés. Cet ajustement a pour contrainte d'atteindre une fréquence de 100% à 350 km/h. Dès lors, les fréquences ajustées sont converties en coefficients. Ces derniers sont ensuite introduits dans le GLM comme contrainte.

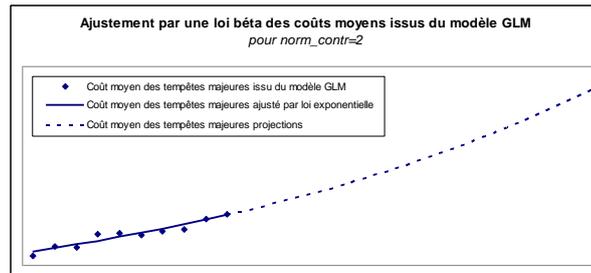


Le modèle des nombres de sinistres peut s'illustrer de la manière suivante :

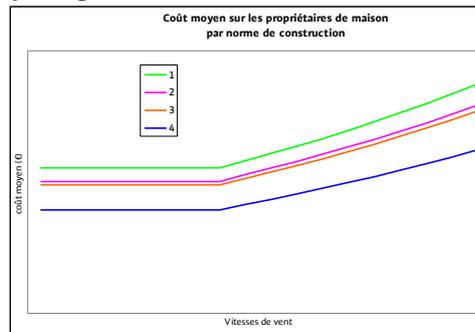


Quant aux coûts moyens, les critères explicatifs sont identiques à ceux des nombres de sinistres à l'exception de l'importance de la tempête et en ajoutant l'enseigne (MAAF, GMF et MMA). On retient dans ce cas la loi gamma avec une fonction de lien logarithmique.

Le modèle GLM retranscrit un coût moyen relativement stable jusqu'à une certaine vitesse de vent et au-delà, le coût moyen progresse de façon exponentielle. Un coût moyen constant est ainsi proposé jusqu'à la vitesse de vent seuil. Au-delà, un ajustement linéaire est réalisé sur les logarithmes des coûts moyens proposés par le modèle GLM. Ces coûts moyens ajustés sont alors convertis en coefficients et introduits comme contrainte dans le modèle GLM.



Le modèle sur les coûts moyens peut s'illustrer de la manière suivante :



Les courbes sur les fréquences de sinistres et sur les coûts moyens en fonction des vitesses de vent permettent ainsi d'estimer les nombres de sinistres et les charges engendrés lors d'une tempête, dès lors que les vitesses de vent par commune sont communiquées. Les faibles écarts d'estimations sur les 6 tempêtes historiques ont permis de valider les courbes.

Chapitre 3. Incertitude autour de la moyenne

Afin de mesurer l'incertitude des estimations autour de la moyenne, deux méthodes sont utilisées : celle sur les écarts à la prédiction (MSEP) et celle sur les bootstrap. Ce sont des méthodes souvent employées pour calculer les incertitudes dans le cadre du provisionnement.

L'erreur quadratique moyenne de prédiction, le MSEP, est le carré de la différence entre la variable aléatoire et son prédicteur. La valeur du MSEP se déduit des valeurs de la volatilité de la variable aléatoire et des valeurs de volatilité de la prédiction. Ces volatilités sont déterminées à partir des hypothèses et des données des modèles linéaires généralisés.

La méthode du bootstrap permet de connaître les inférences. La méthode du bootstrap non paramétrique conduit à bootstraper les résidus de Pearson standardisés afin d'établir un nouveau jeu de données, sur lequel le modèle linéaire généralisé est ensuite appliqué.

Ces 2 méthodes permettent de sécuriser les prédictions réalisées précédemment. Après mis en œuvre, pour une sécurisation à 95% il faut appliquer aux prédictions un coefficient de 1.1% pour les nombres de sinistres, de 2.9% pour les coûts moyens et de 3.5% pour les charges.

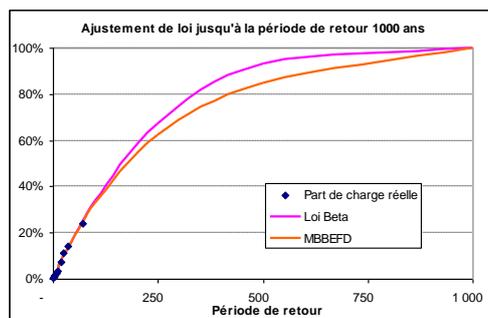
Chapitre 4. Définition de la distribution de la sinistralité et impact tarifaire

Ce chapitre vise à estimer une distribution de la charge annuelle des garanties climatiques puis la rentabilité de ces garanties. Etant donnée la complexité de l'exercice, une approche pragmatique a été retenue sans avoir connaissance à priori des événements. Pour cela, on distingue la charge des tempêtes exceptionnelles et des tempêtes récurrentes, ainsi que les autres événements climatiques (gel, neige ...) et les lois de chacune d'elle.

La loi des tempêtes exceptionnelles (période de retour de plus de 5 ans) est donnée par une courbe de charge de sinistres en fonction de la probabilité de survenance (la période de retour).

Cette courbe est calibrée selon 3 étapes :

- une estimation des charges de 25 tempêtes historiques en fonction de la structure actuelle du portefeuille et à l'aide des courbes d'endommagement sur les fréquences de sinistres et sur les coûts moyens en fonction des vitesses de vent,
- des hypothèses de période de retour pour chacune de ces tempêtes
- et enfin, un ajustement de courbe selon une loi bêta en ajoutant comme contrainte d'atteindre une charge définie pour une période de retour de 1000 années.



Quant aux charges des tempêtes récurrentes (période de retour de moins de 5 ans), elles ont été approximées selon une loi normale, calibrée sur l'historique des primes pures historiques de notre portefeuille. Il n'a pas été noté de tendance sur les différentes années d'observation.

Afin de mesurer les impacts de différentes protections (réassurance et provision pour égalisation), le périmètre doit être élargi aux autres périls climatiques (gel, neige ...) et à l'ensemble des produits (Automobile, Professionnel ...). La charge des autres périls a également été approximée selon une loi normale.

Pour estimer la charge annuelle des tempêtes exceptionnelles, un nombre moyen de tempêtes annuelles a été modélisé selon une loi binomiale négative avec un tirage aléatoire d'une distribution uniforme sur la courbe calibrée. Pour les autres charges, un tirage uniforme est réalisé sur les lois normales. Ainsi plusieurs évènements sont possibles dans l'année.

Pour une charge annuelle nette de cession, les conditions des traités climatiques ont été appliquées à chacune des simulations. Cette charge est ensuite ajoutée aux cotisations cédées et aux charges d'exploitation puis rapportée aux cotisations brutes de cession pour établir un ratio combiné. Avec les simulations, l'analyse des résultats des provisions pour égalisation a montré le ratio combiné nécessaire pour être en capacité de doter et a permis de juger du niveau de tarif.

CONCLUSION

L'étude sur la prédiction des nombres de sinistres lors de la survenance d'une tempête a permis de valider globalement les courbes d'endommagement actuellement appliquées. Quant à la prévision sur les coûts moyens, elle permettra à COVEA d'affiner ces estimations de charges.

L'analyse sur l'impact tarifaire a amené MAAF à réviser le tarif des garanties climatiques sur les produits MRH. Néanmoins, l'étude de sensibilité sur les périodes de retour estimées par les courtiers a montré un impact important sur la prime pure et donc sur le tarif. Les pistes d'approfondissement pour estimer les périodes de retour peuvent être d'analyser les données sur la distribution des vitesses de vent par station météorologique et les corrélations entre ces stations ou d'influer sur les paramètres météorologiques des tempêtes historiques (trajectoires, déplacement et vitesse du courant-jet ...). Pour cela, il est nécessaire alors de se rapprocher des experts météorologiques.

Les réflexions sur ce mémoire peuvent contribuer à des études complémentaires : sur les programmes de réassurance en ajustant des priorités et des plafonds différents afin de connaître les incidences sur les résultats des garanties climatiques ; sur le tarif de réassurance économique ; sur le capital requis du risque catastrophes naturelles dans le cadre de solvabilité 2.

EXECUTIVE SUMMARY

INTRODUCTION

Storm risk is a major issue for insurers. They are covered by transferring some of the risk to reinsurers to protect their own proper. In case of exceed limit's reinsurance, the capital is used to compensate victims and therefore the solvency of the company can be challenged. However, it is complex for an insurer to be set up a proper model. They can then use brokers with market data. This thesis describes risk storm modeling, first of all, hazard on the estimated cost at the time of occurrence of an event. It meets COVEA's needs for claims management and cost's estimation in relation to the accounts close. Secondly, this study focuses on hazard of the probability that this event occurs. The aim is to determine level of required pricing on the climatic guarantee under the current conditions of reinsurance treaties purchased for purposes of MAAF.

Chapter 1 : A storm Risk to manage

The first chapter shows the importance of controlling storm risk by addressing insurance-storm and meteorological knowledge. Exclusion of some weather events (storms, hail, snow) from the system of natural disasters since 1990 requires to understand well the natural hazard. Reinsurance treaties do not offer unlimited guarantee for these perils unlike natural disasters. Major issue about solvency of insurers is based on risks and particularly the storm. This storm risk is characterized by the uncertainty on estimated cost when it occurs and by the hazard on the return period of a storm. Software market act as "black boxes" remains unclear and seems not to be adapted to French risks.

Nevertheless, their modules should estimate storm risk from several angles :

- the "random" module on the determination of the wind speed by municipality and determination of return period of a storm ;
- the "issue" module harvest characteristics of portfolio insured and municipalities ;
- the "vulnerability" module combining information from two previous modules determining damage curves on the wind speed or the return period;

- the "results" module projects number and cost upon the occurrence and estimates an annual probable cost by simulated event.

This type of tool is used to address two needs:

- knowing storm's configuration and our portfolio, what will be number of claims manage and cost generated;
- knowing our portfolio, what is our risk exposure.

Chapter 2 Modelling upon the occurrence of a storm

The purpose of this chapter is to answer for the first hazard to predict numbers of claims and claims costs incurred during a storm. Therefore, a model of loss is carried out using generalized linear models.

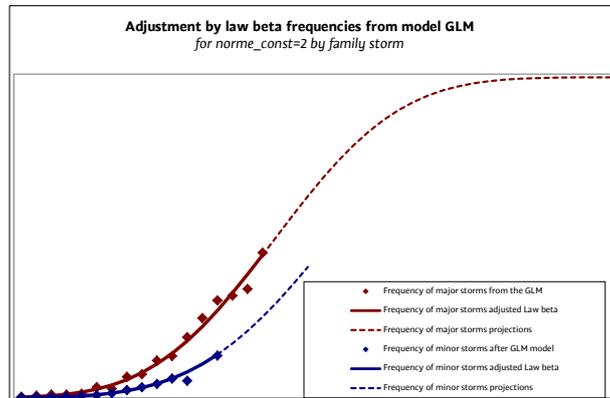
Six historical storms were analyzed by municipality through frequencies and average costs. Variables analyzed are the intensity of the wind, the portfolio structure (legal quality, habitat type, demographics), the construction area ... These variables were introduced in a linear model to judge their relevance and thus exclude them or group of them in the object of an application for an next storm.

Number of claims for municipality depend on legal quality (owners, tenant), on habitat type (house, apartment), on construction standards of housing provided, on density of municipality, on importance of the storm and on wind speed. Selected law representative number of claims is Poisson law with number of contract as constraint and link function, logarithm function.

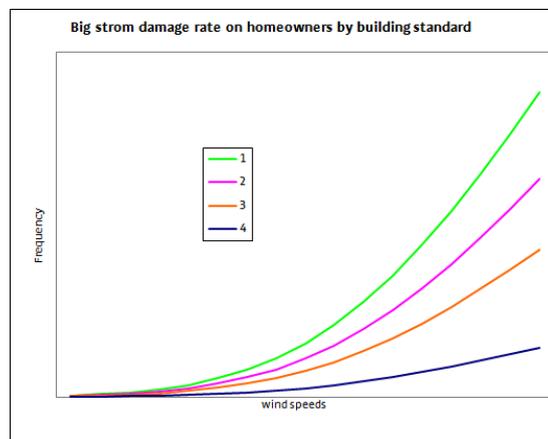
Factors on wind speed correspond to an average of observations and have irregularities frequency (ratio between the number of claims and the number of contract) depending on wind intensity. That's why, these parameters were adjusted according to a law of the exponential family, to project wind speeds not yet registered or not robust enough.

To do that, coefficients wind speeds of GLM are converted to frequencies. Frequencies are then adjusted by a beta law using least squares method. This adjustment has constraint to

achieve a frequency of 100% to 350 km/h. Therefore, adjusted frequencies are converted into coefficients. They are then introduced into the GLM as a constraint.

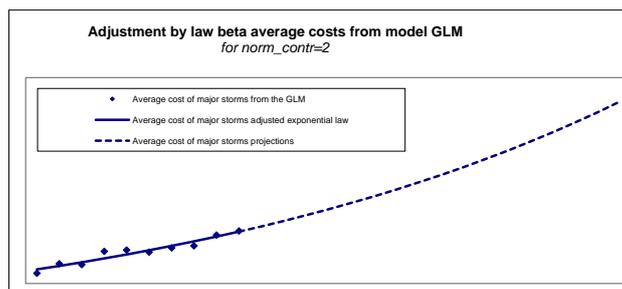


The model on numbers of claims can be illustrated as follows:

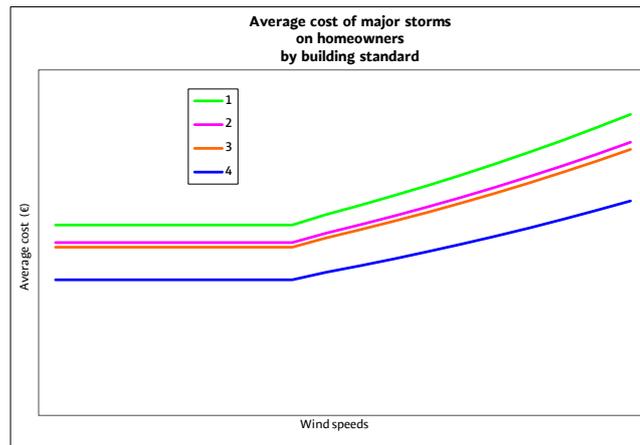


Regarding average costs, explanatory criteria are identical those of the number of claims, except the importance of the storm and by adding the trading name (MAAF, GMF and MMA). It holds in this case law gamma with logarithmic link function.

The GLM transcribes a relatively stable average cost up to a certain wind speed and beyond average cost progressing exponential. A flat average cost is proposed to the wind speed threshold. Beyond, a linear fit is made on logarithms of average costs proposed by the GLM. These adjusted average costs are then converted into coefficients and introduced as constraint in the GLM.



The model on average costs can be illustrated as follows:



Curves on frequencies and on average costs according to wind speeds by municipality allow us to estimate numbers and costs of claims caused during a storm, when wind speeds are provided by municipality. Small differences of estimates on the 6 historic storms validated the curves.

Chapter 3: Uncertainty around the mean

In order to measure the uncertainty of estimates around the mean, two methods are used : one on prediction disparities (MSEP) and one on bootstrap. These are methods widely used to calculate uncertainty in provisioning.

The Mean Square Error of Prediction (MSEP) is the square of difference between the random variable and its predictor. The value of MSEP is derived from volatility values of the random variable and volatility values of the prediction. Volatilities are determined from hypothesis and data from generalized linear models.

The bootstrap method is used to find inferences. Nonparametric bootstrap method leads to bootstrap standardized Pearson residuals with the aim of establishing a new data set which the generalized linear model is then applied.

These two methods allow secure precedent predictions. After implemented, to secure at 95%, it should be applied to a prediction coefficient of 1.1% for numbers claims, 2.9% for average costs and 3.5% for costs of claims.

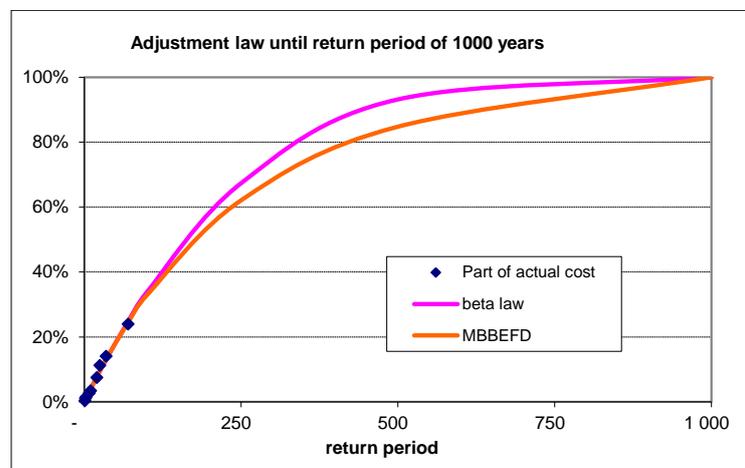
Chapter 4 Defining distribution of claims and impact on pricing

This chapter aims to estimate an annual charge's distribution of climatic guarantees and profitability of these guarantees. Given complexity of exercise, a pragmatic approach was taken without a priori knowledge of events. For this, there are charge of exceptional and recurrent storms and other climate (frost, snow ...) events and laws of each.

The law of exceptional storms (return period of more than 5 years) is given by a curve of costs claims based on the probability of occurrence (return period).

This curve is calibrated in three stages :

- estimated costs by 25 historical storms based on the current structure of the portfolio and with the damage curves on frequency and on average costs of claims according to wind speeds,
- assumptions return period for each of these storms
- and finally, an adjustment of curve as a beta law by adding a constraint to achieve a cost defined a return period of 1000 years.



As for recurring expenses storms (return period of less than 5 years), they were approximated by a normal distribution, calibrated on historical pure premiums in our portfolio. It was not noted trend on the different years of observation.

To measure impact of various protection (reinsurance and equalization provisions), scope should be expanded to other climatic hazards (frost, snow ...) and all products (Automotive, Professional ...). Costs of other perils were also approximated by a normal distribution.

To estimate annual cost of exceptional storms, average annual number of storms was modeled using a negative binomial distribution with a random draw from a uniform distribution on the calibrated curve. For other loads, uniform draw is made on a normal distribution. So many events are possible in the year.

To have an annual net cost of reinsurance, conditions of climate treaties were applied to each simulation. This cost is then added to ceded premium and operating expenses to gross premiums reported to establish a combined ratio. With simulations, analyzing results of equalization provisions showed combined ratio needed to be capacity to staff and helped judge the level of price.

CONCLUSION

Work on the prediction number of claims during an occurrence allows to valid the damage curves currently applied. As projection on average costs, it will allow COVEA refine estimates of costs.

Analysis of the impact on pricing brought MAAF to revise price of climate guarantees on products housing. However, sensitivity analysis on return periods estimated by brokers showed significant impact on pure premium and hence the price. Tracks to estimate the return periods may be to analyze the data on distribution of wind speeds by weather station and correlations between these stations or influence on weather parameters of historic storms (trajectories, displacement and jet stream ...). For this, it is necessary to approach weather experts.

Reflections on the memory may contribute to further studies : for example, programs reinsurance adjusting priorities and different ceilings in order to know impact on the results of climate guarantees, reinsurance costs and economic capital requirements (SCR) of natural disasters risk in the context of Solvency 2.

Sommaire

Résumé	2
Abstract	2
Remerciements	3
Note de synthèse	4
Executive summary	10
Sommaire	16
Introduction	19
Chapitre 1. Les tempêtes : un risque à maîtriser.....	20
I. Cadre	20
I.1. COVEA	20
I.2. Frontière entre catastrophes naturelles et Tempête, Grêle, Neige	20
I.2.1 Catastrophes Naturelles	21
I.2.2 Tempête, grêle et neige (TGN)	21
II. Tempête : Risque à enjeu.....	22
II.1. Typologie des risques d'assurance	22
II.2. Pourquoi maîtriser le risque tempête ?	22
II.3. Evaluation du risque tempête	23
II.3.1 Besoins lors de la survenance d'une tempête.....	23
II.3.2 Utilités de la prime pure annuelle moyenne	23
II.4. Objet du mémoire	24
III. Outils de marché : une solution ?	25
III.1. Courbes de charge et période de retour	25
III.2. Présentation des différents modules	26
III.2.1 Le module aléa : vitesse de vents et période de retour	27
III.2.2 Le module enjeu : caractéristiques du risque assuré et du code postal.....	33
III.2.3 Le module vulnérabilité : modélisations des aléas	35
III.2.4 Le module résultat : mise en application.....	37
IV. Synthèse.....	38
Chapitre 2. Application d'un modèle lors de la survenance d'une tempête	39
I. Objectif.....	39
II. Recherche des critères discriminants	40
II.1. Présentation des données.....	40
II.1.1 Tempêtes historiques	40
II.1.2 Extraction et présentation des critères	40
II.2. Profil des courbes d'endommagement.....	43
II.2.1 Courbe de prime pure.....	43

II.2.2	<i>Courbe de fréquence (N)</i>	45
II.2.3	<i>Courbe de coût moyen (Y)</i>	48
II.3.	Corrélation entre les variables.....	51
III.	Modélisation des tempêtes historiques	52
III.1.	Méthode d'estimation de la sinistralité	52
III.2.	Rappel sur la théorie des modèles linéaires généralisés	52
III.2.1	<i>Régressions Classiques</i>	52
III.2.2	<i>Modèle Linéaire Généralisé</i>	55
III.3.	Modélisation des nombres et des coûts de sinistres	61
III.3.1	<i>Modélisations des Fréquences</i>	61
III.3.2	<i>Modélisations des Coûts Moyens (Y)</i>	73
III.3.3	<i>Résultats des modélisations en terme de charge</i>	82
IV.	Synthèse	83
Chapitre 3.	Incertitude autour de la moyenne	84
I.	Objectif	84
II.	Erreur quadratique moyenne de prédiction	85
II.1.	Définition	85
II.2.	Volatilité des nombres et des coûts moyens de sinistres	86
II.2.1	<i>Volatilité des variables aléatoires</i>	86
II.2.2	<i>Volatilité des prédictions</i>	88
II.3.	Volatilité des charges de sinistres.....	89
II.3.1	<i>Volatilité de la variable aléatoire</i>	90
II.3.2	<i>Volatilité des prédictions</i>	91
II.4.	MSEP des indicateurs de sinistralité et sécurisations	91
II.5.	Limite de la méthode MSEP.....	92
III.	Bootstrap	93
III.1.	Méthode	93
III.2.	Bootstrap des résidus de Pearson standardisés	93
III.2.1	<i>Principe général</i>	93
III.2.2	<i>Etapas de construction</i>	94
III.3.	Mise en application et résultats.....	96
III.4.	Limite de la méthode du bootstrap	97
IV.	Synthèse	98
Chapitre 4.	Définition de la distribution de la sinistralité et impact tarifaire	99
I.	Objectif	99
II.	Prime pure annuelle moyenne du risque tempête	100
II.1.	Méthode de la constitution de la courbe de sinistralité en fonction des périodes de retour.....	100

II.1.1	Coût as if sur le portefeuille MRH.....	100
II.1.2	Ajustement des coûts as if en fonction des périodes de retour.....	102
II.2.	Distinction des tempêtes récurrentes et exceptionnelles.	104
II.2.1	Analyse des primes pures des tempêtes récurrentes MRH.....	104
II.2.2	Simulation des tempêtes exceptionnelles MRH.....	104
III.	Impact tarifaire	109
III.1.	Mécanismes de protection.....	109
III.1.1	les traités de réassurance.....	109
III.1.2	les provisions pour égalisation	114
III.1.3	Illustration globale des mécanismes de protection	115
III.2.	Détermination des ratios combinés des évènements climatiques	116
III.2.1	Hypothèses sur les autres périmètres	116
III.2.2	Analyse des résultats des traités et des provisions pour égalisation	117
III.2.3	Compte de résultat par branche.....	119
IV.	Synthèse.....	120
	Conclusion Générale.....	121
	Bibliographie	122

Introduction

Les tempêtes Lothar et Martin de 1999 ont révélé que le risque tempête était sous-estimé par les compagnies d'assurance. Elles ont vu leur résultat fortement se réduire, voire même devenir déficitaire. Pour y faire face, certains assureurs ont augmenté leur tarif de manière significative.

Aujourd'hui, ce risque est mieux appréhendé mais les assureurs ne disposent pas de modèle propre et leur vision du risque tempête reste à approfondir. Un capital économique très important est nécessaire aux assureurs non vie pour faire face à ce risque. Il leur faut se couvrir en transférant une partie de ce risque aux réassureurs et ainsi protéger leurs fonds propres. Les capacités de réassurance depuis les tempêtes Lothar et Martin ont fortement progressé de part l'augmentation des plafonds qui restent toutefois limités.

Cette étude a pour objectif de proposer un modèle sur le risque tempête pour répondre aux besoins de COVEA en estimant le nombre et le coût de la sinistralité de la Multirisque Habitation lors de la survenance d'une tempête en France métropolitaine, et aux besoins de MAAF Assurances en jugeant de la pertinence du tarif climatique.

Le premier chapitre montre l'importance de maîtriser le risque tempête en abordant les connaissances assurantielles et météorologiques. Nous verrons que le risque tempête comporte deux aléas : l'un sur l'estimation de la charge lors de la survenance et l'autre sur la probabilité de survenance d'une tempête.

L'objectif du second chapitre est de répondre au premier aléa pour prévoir les nombres de sinistres et les coûts engendrés lors de l'arrivée d'une tempête. Pour ce faire, une modélisation de la sinistralité est réalisée à l'aide de modèles linéaires généralisés.

Le but du troisième chapitre est de connaître pour le premier aléa la volatilité des estimations moyennes en se basant sur les méthodes MSEP et bootstrap.

Enfin, le dernier chapitre vise à répondre au deuxième aléa en définissant une fonction entre les charges et les périodes de retour des tempêtes pour simuler des événements. Après une présentation des traités de réassurance et des provisions pour égalisation, une prime pure est estimée pour l'ensemble des périls climatiques hors catastrophes naturelles de tous les produits IARD afin de constituer un compte de résultat et ainsi connaître la pertinence du tarif actuel à travers le ratio combiné net de réassurance.

Chapitre 1. LES TEMPETES : UN RISQUE A MAITRISER

I. Cadre

I.1. COVEA

La MAAF a été créée en 1950 afin d'offrir aux artisans la possibilité de s'assurer mutuellement, par souci d'économie et esprit de solidarité pour l'obtention d'une assurance automobile. Elle se développe rapidement et propose d'autres produits d'assurance : habitation, professionnel puis épargne, santé et prévoyance. Après des difficultés, la MAAF renoue avec la croissance depuis les années 1990.

En 2003, la Société de Groupe d'Assurance Mutuelle (SGAM) nommée COVEA est créée. Elle regroupe les mutuelles MAAF et MMA puis GMF en 2005 afin de réaliser des économies d'échelle et de mutualiser les investissements.

Un français sur 4 a rejoint le groupe COVEA, 1er en assurance de biens et de responsabilités, 3ème en entreprises et 4ème en santé individuelle.

I.2. Frontière entre catastrophes naturelles et Tempête, Grêle, Neige

Les activités de l'assurance se distinguent entre les assurances de personnes et les assurances de biens et de responsabilités. Ce dernier couvre notamment l'assurance automobile, l'assurance habitation, et l'assurance des entreprises (ou professionnelle).

Ces contrats d'assurance de biens doivent contenir entre autres les garanties aléas naturels (tempête, grêle, neige) et autres périls (inondation, sécheresse, tremblement de terre...).

1.2.1 Catastrophes Naturelles

Concernant les autres périls, les sinistres peuvent relever du régime des catastrophes naturelles. La garantie joue uniquement si un arrêté interministériel paru au Journal Officiel constate l'état de catastrophe naturelle dans la commune où se trouve le bien endommagé. Le maire de la commune adresse une demande au préfet qui lui-même fait remonter la demande à une délégation interministérielle qui déclare si oui ou non cette commune est en état de catastrophe naturelle. Pour les sinistres ne relevant pas du régime, ils sont indemnisés au titre des garanties dommages.

La tarification de la garantie catastrophes naturelles est assise sur celle des garanties dommages.

Extrait du Code des Assurances L125-2

La garantie catastrophes naturelles « est couverte par une prime ou cotisation additionnelle, individualisée dans l'avis d'échéance du contrat visé à l'article L. 125-1 et calculée à partir d'un taux unique défini par arrêté pour chaque catégorie de contrat. Ce taux est appliqué au montant de la prime ou cotisation principale ou au montant des capitaux assurés, selon la catégorie de contrat. »

1.2.2 Tempête, grêle et neige (TGN)

Les dommages consécutifs aux tempêtes, à la neige et à la grêle ont été exclus du régime des catastrophes naturelles depuis 1990. Ils font l'objet d'une (ou de) garantie(s) dommages à part entière.

Les assureurs tarifient librement cette garantie selon leur connaissance du risque.

Chacune de ces catégories (Cat. Nat. et TGN) représente environ 10% de la cotisation dommages pour le marché français.

II. Tempête : Risque à enjeu

II.1. Typologie des risques d'assurance

La maîtrise des risques est au cœur du métier de l'assurance. Les assureurs proposent des assurances IARD (Automobile, Multirisques Habitation ou professionnels ...) pour lesquelles on distingue trois catégories de sinistres : les sinistres de fréquence (RC matérielle, vol, incendie, dégâts des eaux ...), d'intensité (RC corporelle, familiale, professionnelle ...) et cycliques/pluriannuels (naturels et techniques).

Les deux premières sont des risques relativement bien connus et relativement maîtrisés par la profession, ce qui est moins le cas pour la dernière catégorie, du fait de son caractère très aléatoire.

Les risques naturels peuvent être la tempête, l'inondation, la sécheresse, la grêle, les tremblements de terre. Les risques techniques sont quant à eux liés à l'intervention humaine et regroupent la RC catastrophique (exemple : type tunnel du Mont Blanc), la conflagration, le terrorisme.

II.2. Pourquoi maîtriser le risque tempête ?

Afin de maîtriser sa solvabilité et palier la méconnaissance de ces risques, la profession utilise différents leviers tels que les mécanismes de réassurance. Ainsi, l'assureur transfère une partie du risque aux réassureurs. Par exemple, pour les risques Attentat et « catastrophes naturelles » (inondation, sécheresse, tremblement de terre ...), les assureurs s'appuient sur la Caisse Centrale de Réassurance (CCR) qui possède la garantie de l'Etat avec une couverture illimitée. Pour aider les assureurs à choisir les traités de réassurance en fonction de leurs besoins, ils coopèrent avec des courtiers.

Actuellement, les traités dans le domaine des événements climatiques hors catastrophes naturelles ne contiennent pas de garantie illimitée. C'est pourquoi, **l'enjeu sur la solvabilité des sociétés non-vie porte sur ces événements tels que la tempête**, et la dérive du changement du climat nous incite à mieux appréhender ces risques climatiques.

II.3. Evaluation du risque tempête

La prime pure annuelle moyenne du risque tempête (PP) dépend à la fois de la probabilité de l'occurrence d'un évènement (proba) et aussi de la détermination de la prime pure (produit de la fréquence et du coût moyen) lors de sa survenance.

Pour modéliser l'aléa sur la probabilité de l'occurrence et l'aléa sur la charge lors de la survenance, il est indispensable d'avoir un système d'information décisionnel de qualité, avec des données centralisées et cohérentes. Elles doivent être partagées par tous les acteurs de l'entreprise (directions techniques, sinistres, solvabilité, réassurance ...).

II.3.1 Besoins lors de la survenance d'une tempête

La détermination de la prime pure lors de la survenance répond aux besoins de gestion des flux de sinistres et des arrêtés des comptes.

En fonction de l'ampleur de la tempête, COVEA met en place un Plan d'Intervention Exceptionnel (PIE). Il consiste à réorganiser la gestion des sinistres : transférer les appels sur des centres plus disponibles, déplacer des experts vers la région sinistrée, faciliter la déclaration des sinistres en guidant les assurés sur le site internet, en aménageant des locaux près des lieux sinistrés ... Afin d'évaluer l'importance de la tempête, les nombres de sinistres sont estimés en fonction notamment des vitesses de vent. Ces vitesses sont estimées et communiquées quelques jours avant la survenance. Ainsi, en fonction du nombre de sinistres prévu, le PIE est déclenché.

Dès lors que la tempête est survenue, les vitesses de vent sont relevées par station météorologique et lissées par code postal pour une estimation des nombres de dossiers plus précise et pour une évaluation des coûts de sinistres. Cette dernière est utilisée pour les arrêtés des comptes et est communiquée pour des besoins internes (tarification, financier ...) et externes (réassurance, médias ...).

II.3.2 Utilités de la prime pure annuelle moyenne

A chaque survenance de tempête, il est possible d'associer une probabilité d'occurrence encore appelée période de retour. A partir de la relation entre la charge de l'évènement et la période de retour, ainsi que d'hypothèses complémentaires, des tempêtes peuvent être simulées et une prime pure annuelle moyenne en est déduite.

Ces simulations permettent également de prévoir le ratio combiné et ainsi établir le tarif adapté en tenant compte des provisions pour égalisation et de la réassurance. Elles sont également exploitées pour définir les caractéristiques du programme de réassurance (priorité, portée ...) et pour calculer le besoin en capital et en marge de solvabilité S2.

II.4. Objet du mémoire

L'objectif de ce mémoire est de résoudre l'équation suivante :

$$PP = \sum_{\substack{\text{Ensemble} \\ \text{des} \\ \text{événements}}} \text{proba (fréq x CM)}$$

PP : Prime Pure annuelle moyenne

Fréq : Fréquence de sinistre lors de l'évènement

Proba : Probabilité d'occurrence de l'évènement

CM : Coût Moyen des sinistres lors de l'évènement

La modélisation sur l'aléa lors de la survenance de la tempête sera présentée lors du chapitre 2 « Modélisation lors de la survenance d'une tempête », puis une mesure de la volatilité de cet aléa lors du chapitre 3 « Incertitude autour de la moyenne » et enfin l'aléa sur la probabilité d'occurrence lors du chapitre 4 « Définition courbe sinistralité et impact tarifaire ».

III. Outils de marché : une solution ?

Avant de réaliser les différentes modélisations, cette partie a pour objet de décrire les outils de marché permettant de modéliser la tempête afin d'en appréhender les méthodes d'évaluation.

Les courtiers et les réassureurs s'appuient sur des logiciels de marché pour définir la prime de réassurance des risques climatiques en fonction des besoins de l'assureur (description des biens assurés, définition des traités de réassurance ...).

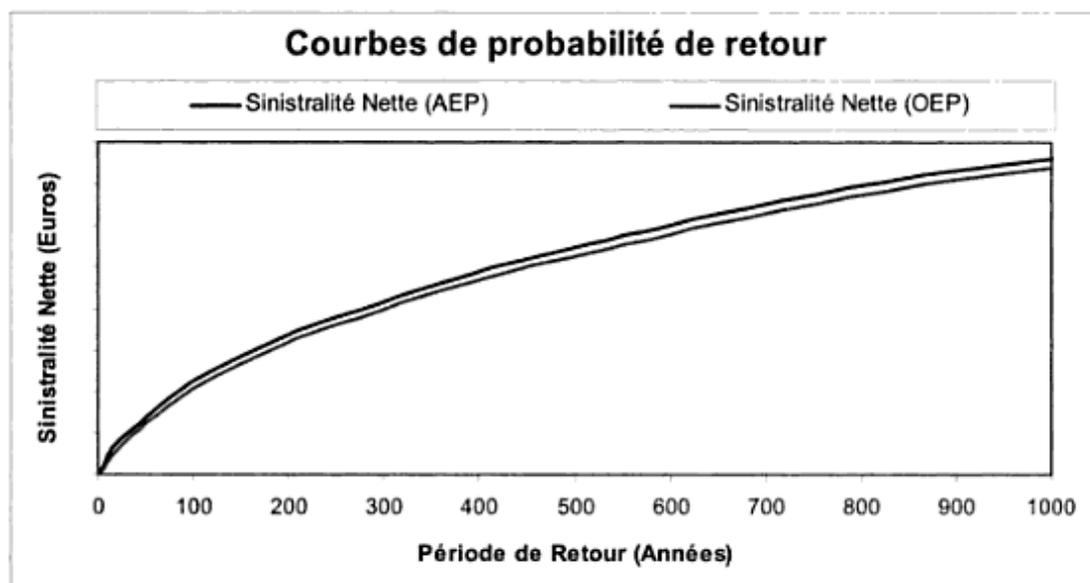
La description de ces outils servira de base pour les modélisations du risque tempête.

III.1. Courbes de charge et période de retour

Il existe sur le marché trois sociétés principales qui éditent chacune leur logiciel de modélisation du risque climatique : Applied Insurance Research (AIR), Risk Management Solutions (RMS) et ABS Consulting (EQECAT).

Ces 3 logiciels produisent des résultats sous forme de courbes :

- OEP (Occurrence Exceedance Probability) qui attribue un coût maximal d'un évènement sur une année en fonction d'une période de retour
- AEP (Aggregate Exceedance Probability) qui attribue un coût total des évènements sur une année en fonction d'une période de retour.



Source : « La titrisation du risque d'assurance » de Bauwens V. et de Walhin JF Larcier

On peut également assimiler une période de retour à une Value at Risk. Ainsi les assureurs doivent être en capacité de faire face à une Value at Risk de 99.5 pour répondre aux exigences de solvabilité 2, c'est-à-dire d'être capable d'indemniser un évènement catastrophique d'une période de retour de 200 ans ($1 / (1 - 99.5\%)$).

Selon les courbes de marché OEP, on pourrait supposer que l'évènement LOTHAR a une période de retour de 70 ans et MARTIN de 35 ans, loin d'un évènement deux centennal.

Il est à noter que cette période de retour est associée à deux probabilités : celle de la survenance de l'évènement (caractéristique météorologique) et celle de la distribution du coût. En effet, en simulant différentes trajectoires probables de la tempête LOTHAR, la trajectoire réalisée correspond au 3^{ème} scénario le plus onéreux (d'après une présentation sur « Les innovations face au développement des catastrophes non-vie » à Tours le 18/19 mars 2013).

Ces outils de modélisation pour le risque tempête, développés sur le marché américain, sont appliqués sur le marché français avec toutefois des calibrages mal maîtrisés, et sont exploités par le marché de la réassurance. Le logiciel le plus usité par la réassurance est RMS.

En somme, **les réassureurs**, et par voie de conséquence les assureurs, **sont dépendants de ces logiciels dont les contours des méthodes restent obscurs**. Les versions des logiciels remettent en cause les hypothèses de simulations définies lors de leurs versions précédentes, pouvant entraîner des variations de primes cédées importantes demandées par les réassureurs.

III.2. Présentation des différents modules

Dans cette partie, nous allons nous intéresser aux principes de fonctionnement de ces logiciels et y ajouter les recherches complémentaires.

Les logiciels de modélisation traitent un certain nombre de périls et s'articulent autour de 4 modules : aléa, enjeux, vulnérabilité et résultats. Chacun de ces modules est présenté afin d'en dégager les premiers éléments orientés sur le péril tempête : climatique, météorologique, critères influant la sinistralité...

Cette présentation servira de support pour les modélisations des aléas sur la survenance et sur l'occurrence d'un évènement.

III.2.1 Le module aléa : vitesse de vents et période de retour

L'objectif de ce module est de récolter les informations météorologiques.

III.2.1.1 Définition du vent

Avant toute chose, il est important de comprendre comment naît la tempête.

a) Vent : Anticyclones et dépressions

L'air est composé de molécules de gaz (essentiellement d'azote, d'eau et d'oxygène) qui se déplacent sans cesse. La pression est plus forte lorsqu'il y a plus de molécules de gaz. Les molécules se déplacent de l'endroit où il y a plus de molécules vers l'endroit où il y a moins de molécules afin de rétablir un équilibre.

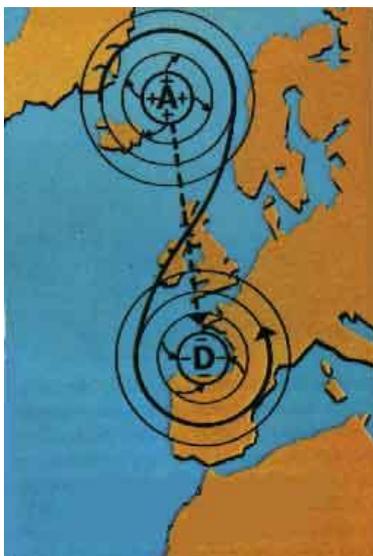
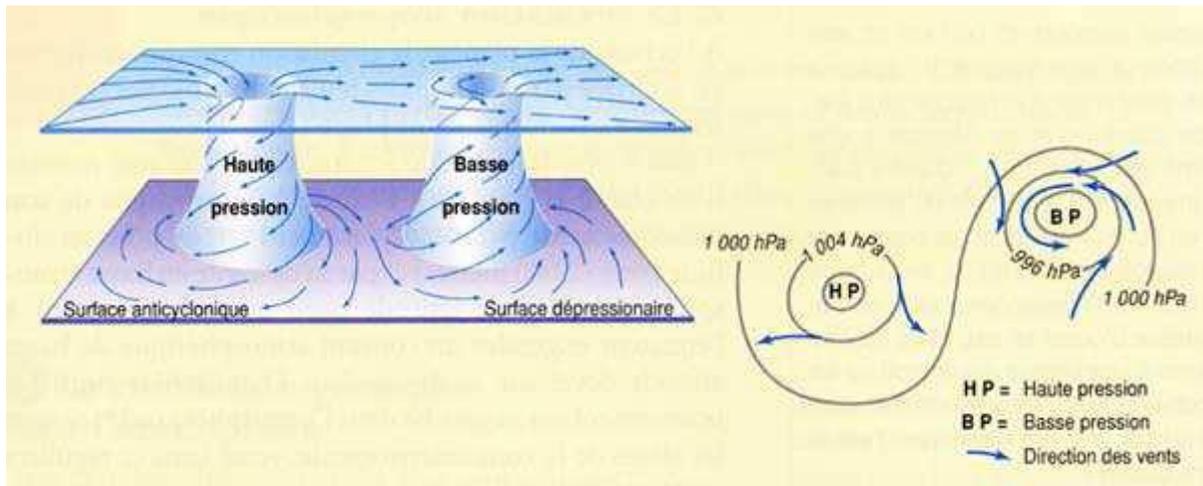
Les contrastes de températures entre les masses d'air chaud et froid combinés aux effets de la force de Coriolis provoquent la formation d'un courant aérien très puissant qui circule entre 8 et 12 km d'altitude à la limite supérieure de la troposphère. Ce courant aérien est appelé le courant-jet. Sa vitesse dépend de l'écart de températures entre les masses d'air. Elle est de l'ordre de 200 km/h, variant de 100 km/h à plus 400 km/h. La puissance de la tempête est conditionnée par l'intensité du courant-jet. Lors de la tempête de Martin en décembre 1999, la vitesse du courant-jet a atteint 529 km/h à 8128 m au-dessus de Brest.

Ainsi, le vent est le déplacement de masse d'air qui se dirige des hautes pressions (anticyclone) vers les basses pressions (dépression).

L'anticyclone se caractérise par une pression maximale à son centre (entre 1020 et 1050 hPa), diminuant vers la périphérie. Son mouvement descendant entraîne la disparition des nuages et empêche leur formation.

La dépression se caractérise par une pression minimale à son centre (inférieure à 1000 hPa), augmentant vers la périphérie. Son mouvement ascendant qui en résulte va entraîner la formation des nuages. En effet, en s'élevant la vapeur d'eau se condense et on assiste à la formation de nuages. Si le phénomène a lieu au-dessus d'un océan tropical, il peut, dans certains cas, engendrer un cyclone.

Il peut arriver qu'une ondulation du courant jet, se déplaçant majoritairement d'ouest en est, rattrape la petite dépression et se marie avec elle. Si cette configuration se maintient durablement alors la tempête naît et croît en puissance.



www.passion-meteo.net

Mais le vent ne souffle pas tout droit, il suit un chemin qui se développe en spirales à cause de grands mouvements d'air guidés par la force de Coriolis, due à la rotation de la Terre.

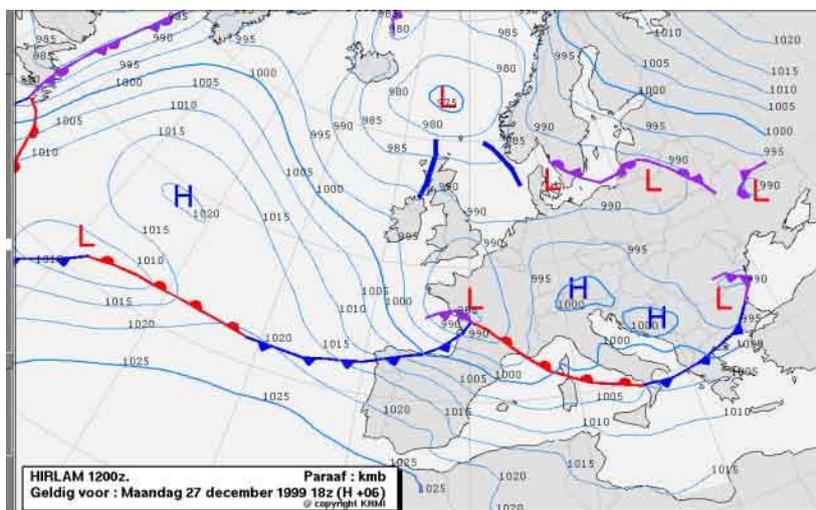
Le sens de la direction des vents est inversé dans l'hémisphère sud.

b) Force du vent et Isobares

Si l'on relève les valeurs de la pression atmosphérique en différents points du globe et que l'on relie entre eux les points de pression identique, on obtient une série de courbes, appelées isobares. L'ensemble des hautes et basses pressions dessine une sorte de carte du "relief" de l'atmosphère.

La force du vent est due à la différence de pression atmosphérique. Ainsi, plus les isobares sont resserrées, plus le vent est fort. En effet, le vent "coule" comme une rivière qui dévalerait la pente d'une montagne. Plus la pente est raide, plus l'eau coule vite. Il en est de même pour le vent. Plus la différence de pression est grande, plus le déplacement des masses d'air est important et donc plus le vent souffle fort.

Pour illustrer les propos précités, prenons pour exemple la carte isobare de la tempête MARTIN de décembre 1999. Nous voyons très bien les isobares resserrées sur les chaînes pyrénéennes avec un centre dépressionnaire de 985hPa situé au sud-est de la Bretagne.



www.passion-meteo.net

c) Vent de surface

La surface du sol a une influence importante sur les vents de surface jusqu'à une altitude de 100 mètres au-dessus du sol. La rugosité du terrain, les obstacles naturels ou artificiels et la rotation de la terre peuvent influencer la vitesse et la direction du vent.

On peut noter qu'il existe des lois de passage d'une vitesse de vent d'une altitude à une autre en fonction de la nature du terrain (Cisaillement).

$$\frac{V_1}{V_2} = \left(\frac{h_1}{h_2} \right)^\alpha$$

V_1 et V_2 : vitesses de vent horizontal (en m/s) aux hauteurs respectives h_1 et h_2 (en m)

L'exposant α caractérise le terrain comme dans le tableau ci-dessous

Nature du terrain	Inégalité du sol en cm	Exposant α
Plat : neige, glace, mer, marécages, herbes courtes	0 à 20	0,08 à 0,12
Mer		0,13
Peu accidenté : champs et pâturages, cultures	20 à 200	0,13 à 0,16
Rivage		0,16
Plaine		0,2
Accidenté : bois, zones peu habitées	1 000 à 1 500	0,20 à 0,23
Plaine boisée		0,24
Très accidenté : villes	1 000 à 4 000	0,25 à 0,4

III.2.1.2 Définition et qualification des tempêtes

Échelle des vitesses du vent et sa qualification à 10 m de hauteur

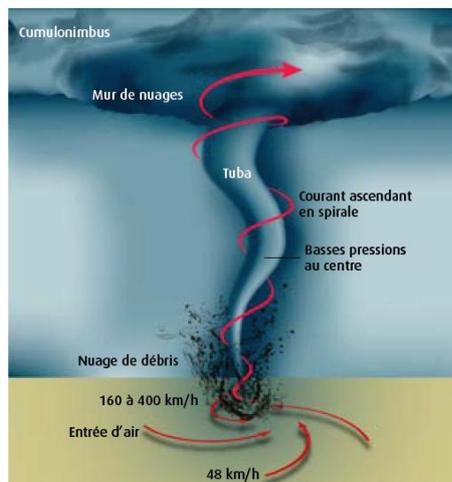
m/s	km/h	nœuds	Échelle de Beaufort	Force du vent et distance séparant 2 isobares	Vent	Impact à terre
0,0-0,4	<1	0,0-0,9	0			
0,4-1,8	1-5	0,9-3,5	1		Vent insensible et très faible	La fumée monte verticalement (les feuilles semblent immobiles)
1,8-3,6	6-11	3,5-7,0	2	600 km	Brise infime Brise légère (les moulins démarrent)	la fumée est légèrement déviée
3,6-5,8	12-19	7-11	3		Vent modéré (frais)	les petits rameaux remuent un peu
5,8-8,5	20-29	11-17	4	500 km	Vent léger Brise moyenne (bon frais)	les petites branches sont agitées les poussières et les papiers sont soulevés
8,5-11	30-38	17-22	5	400 km	Forte brise Grand frais Très forte brise	les branches oscillent et se balancent des vaguelettes sur les lacs
11-14	39-51	22-28	6	300 km	Vent fort	les feuilles sont arrachées, les grosses branches sont agitées
14-17	52-61	28-34	7	200 km	Vent très fort	les petites branches cassent, les arbres oscillent en entier
17-21	62-74	34-41	8		Vent impétueux	les branches moyennes se brisent, marche difficile face au vent
21-25	75-88	41-48	9	100 km	Fort coup de vent	début de dégâts sur les toitures (antennes, cheminées)
25-29	89-101	48-56	10		Tempête	les arbres se déracinent, des dégâts structurels importants
29-34	102-120	56-65	11		Tempête violente	très gros dégâts aux toitures, voitures ...
>34	>120	>65	12		Ouagan	Dégâts étendus dans les forêts et les villes

D'après l'échelle de Beaufort, on parle de tempête lorsque les vents dépassent les 89 km/h (soit le degré 10). Selon Météo France, la tempête est qualifiée comme telle, dès que 5% des stations météorologiques enregistrent un vent supérieur à 100km/h pendant 3 jours. Elle est considérée comme forte lorsque la part passe à 20% au lieu des 5%

La plupart des tempêtes qui touchent l'Europe se forment au-dessus de l'Atlantique, entre 35° et 70° de latitude Nord. Ces tempêtes surviennent principalement en hiver et en automne pour l'Europe (notamment entre novembre et février). On dénombre en moyenne 15 tempêtes par an.

Historiquement, les noms des tempêtes étaient donnés selon la journée où elles se produisaient ou toute autre association marquante et ce nom variait d'un pays à l'autre. Depuis les années 1990, les dépressions et les anticyclones reçoivent des noms féminins ou masculins. Pour les années paires, les dépressions reçoivent des noms féminins et les anticyclones des noms masculins et inversement pour les années impaires. C'est l'université de Berlin qui établit une liste de noms pour les tempêtes.

III.2.1.3 Cas particulier des tornades



La tornade est une manifestation particulière des tempêtes et des orages. Une tornade est une colonne d'air tournante très violente, issue d'un nuage instable qu'elle relie au sol. L'intensité des tornades est évaluée sur l'échelle dite de Fujita améliorée qui comporte six degrés (EF0 à EF5) et repose sur une appréciation de l'importance des effets de la tornade.

Source : Ministère de l'Écologie, du Développement durable et de l'Énergie

L'échelle Fujita

Classe	Vents maximum	Dégâts
F0	Vents légers de 64 à 116 km/h	Branches cassées, torsion de panneaux de signalisation.
F1	Vents modérés de 117 à 180 km/h	Tuiles arrachées, caravanes renversées.
F2	Vents forts de 181 à 252 km/h	Arbres et toitures arrachées.
F3	Vents très forts de 253 à 330 km/h	Murs arrachés, camions et trains renversés.
F4	Vents dévastateurs de 331 à 417 km/h	Soulèvement et déplacement de bâtiments sans fondation, de véhicules légers. Les objets jusqu'à 100 kg s'envolent.
F5	Vents incroyables de 418 à 509 km/h	Soulèvement et déplacement de bâtiments avec fondations, de camions, de trains. Les débris divers se transforment en projectiles d'une violence inouïe.

Source : www.meteofrance.fr/prevoir-le-temps/phenomenes-meteo/les-trombes-et-tornades

III.2.1.4 Méthode déterministe : Écoulement des vents

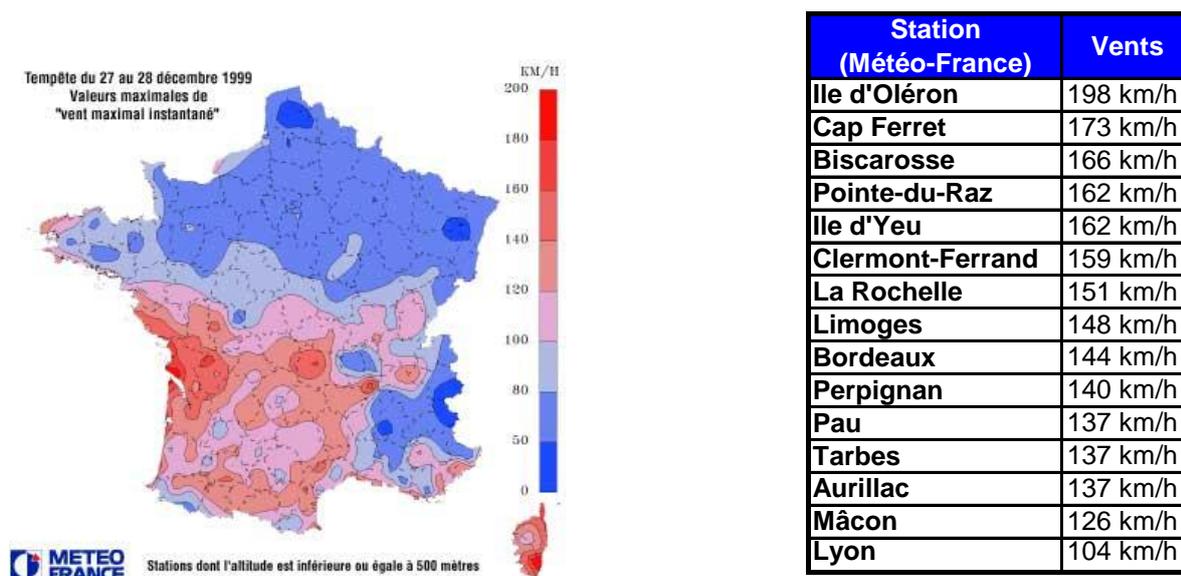
L'Organisation Météorologique Mondiale (OMM) coordonne un réseau mondial de stations d'observation en altitude de haute qualité, dont la plupart sont exploitées par les services météorologiques nationaux. Ces stations sont dotées d'instruments capables de mesurer des variables atmosphériques comme la pression, la température, l'humidité ainsi que la vitesse et la direction du vent dans la troposphère (jusqu'à 18 kilomètres au-dessus du sol) et dans la basse stratosphère (20-30 kilomètres au-dessus du sol). Ainsi, Météo France et

d'autres organismes météorologiques s'appuient sur un réseau de stations météorologiques qui respectent les règles définies par l'OMM.

Une des règles porte sur le capteur de vent. Il devrait être installé à une hauteur de 10 m dans un espace dégagé de tout obstacle, à une distance d'au moins 10 fois sa hauteur. Cette normalisation permet de comparer les mesures du vent d'un capteur à l'autre. Par exemple, la vitesse de vent en haut de la tour Eiffel (324 mètres) est bien plus élevée que celle relevée en haut des immeubles aux alentours au même moment.

OMM préconise aussi une mesure des rafales des vents supérieures à 3 secondes avec une fréquence d'échantillonnage de 4 Hz.

Météo France recueille les données de ces stations et les corrige (station endommagée, données incohérentes, normalisées ...), puis extrapole les vitesses de vents sur l'ensemble du territoire et aboutit alors au bout de 2 jours à une carte sur l'écoulement des vents.



Carte des vents de MARTIN

COVEA recueille également les relevés par station météorologique et réalise elle-même le lissage des vitesses de vents entre les stations. La présentation de ce lissage ne fait pas partie de ce mémoire car la méthodologie porte sur des phénomènes physiques et non sur des techniques statistiques.

Ainsi, la vitesse des vents par code postal est retraitée dans ce modèle aléa. Nous verrons que ces données sont alors associées aux portefeuilles et à la sinistralité du code postal issues

du module enjeu, et sont en entrée du module vulnérabilité pour établir les fonctions d'endommagement (fréquence, coût moyen ...).

III.2.1.5 Méthode Stochastique : Période de retour

Les outils de marché ont plusieurs méthodes de simulations basées sur :

- l'influence des paramètres météorologiques (pression, vitesse de vent) des tempêtes existantes, contrôlés par un modèle physique pour vérifier la pertinence des tempêtes simulées,
- un modèle dynamique générant des champs de vents probables, via la technologie Numérical Weather Prediction, basé sur les tempêtes historiques.

Ces méthodes permettent d'obtenir une période de retour sur une tempête simulée. Avec la mise en application de la méthode déterministe sur la vulnérabilité, un nombre de sinistre et une charge sont déterminés pour cette tempête simulée aboutissant à une distribution de charge OEP en simulant une infinité de tempêtes.

III.2.2 Le module enjeu : caractéristiques du risque assuré et du code postal

Cette partie développe les critères qui peuvent être discriminants sur l'évaluation lors de la survenance d'une tempête.

- Portefeuille de risques assurés

Le risque assuré est bien évidemment un critère discriminant. La typologie de sinistre diffère entre une maison, un appartement et une entreprise, ou bien entre un propriétaire et un locataire. Une étude qualitative COVEA a été menée afin d'illustrer par un portrait robot les types de dommages engendrés en fonction de la typologie du risque assuré.

Ainsi, les informations du portefeuille sont indispensables pour définir une mesure du risque. Pour alimenter les logiciels, les courtiers demandent certaines informations et notamment au travers de la notion de somme assurée. Elle peut être définie comme fonction d'une surface immobilière, valeur mobilière et d'un coût au m² ainsi que du coût lié à la perte d'exploitation pour les risques professionnels, dépendant donc de la caractéristique du portefeuille. Un assureur n'ayant que des locataires aura une somme assurée bien inférieure à celle d'un assureur n'ayant que des propriétaires.

- Modélisation géographique

Les réassureurs ont très longtemps utilisé les départements comme zones CRESTA (Catastrophe Risk Evaluating and Standardizing Target Accumulations). Ce découpage n'était pas assez fin pour l'ensemble des périls. Les codes postaux ont alors été utilisés, soit 6 600 codes postaux pour les 36 000 communes. Un travail de normalisation de ces codes est nécessaire (ex : cédex).

Pour le péril tempête, la précision du code postal permet de prendre en considération les spécificités locales.

Pour les autres périls, le maillage doit être encore plus fin. Les risques doivent être alors géocodés. C'est le cas par exemple du péril inondation. Il est en effet nécessaire de connaître le positionnement précis du risque assuré par rapport à une zone inondable, un cours d'eau ...

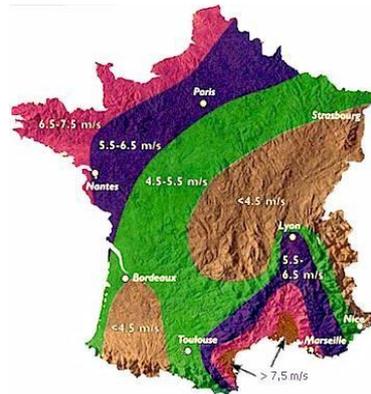
- Vulnérabilité des constructions

On note des vents dominants en France, illustrés sur cette carte des vents dominants annuels moyens avec la rose des vents.



www.alertes-meteo.com

La carte du potentiel éolien illustre également l'habitude du vent, et permet ainsi de découper le territoire selon les vitesses moyennes de vents observés.



Source : ADEME 2009

De plus, certaines régions sont plus fréquemment exposées aux tempêtes. Munich Re a défini des zones d'exposition aux tempêtes, sur la base de données NATHAN.

En fonction de la fréquence des vents et de leur intensité, des normes de construction sont définies, et ainsi la France est découpée en zones sur la résistance des bâtiments, définie par le CSTB.

- Densité au code postal

Il est important de noter les spécificités locales au travers des données géographiques.

En effet, les données topographiques influent sur la vitesse du vent qui dépend essentiellement de la nature du terrain au-dessus duquel se déplacent les masses d'air (cf. page 18)

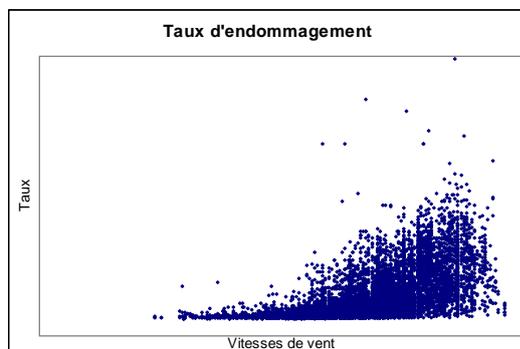
La rugosité du sol (relief) est certes prise en compte à travers les données des vitesses de vent mais elle est peut être réduite par la densité d'habitation. Effectivement, une étude américaine (Usbeck et al, 2010) a montré que dans une zone urbaine dense la vitesse de vent est moindre ce qui explique la fréquence de sinistre faible dans les agglomérations. Ainsi, les zones urbaines ou rurales peuvent être considérées comme un critère discriminant sur la sinistralité pour le péril tempête.

III.2.3 Le module vulnérabilité : modélisations des aléas

Ce module recueille les informations collectées dans les 2 modules précédents.

Les courbes d'endommagement ou de vulnérabilité (fréquence, coût moyen ...) sont déterminées en fonction de l'intensité du vent et des critères discriminants.

Voici un exemple avec les données COVEA :



Chaque point du graphique ci-dessous représente un code postal donné avec sa vitesse de vent lissée et son taux d'endommagement (rapport entre le nombre de sinistres et le nombre de contrats).

III.2.3.1 Aléa : charge lors de la survenance

Différentes méthodes statistiques sont alors utilisées afin d'ajuster les courbes à ce nuage de points. Le module consiste donc à associer un indicateur de sinistralité (nombre, coût moyen et/ou charge) à une vitesse de vent en vision moyenne. Les logiciels de marché utilisent principalement un taux de destruction étant le rapport entre une charge et les sommes assurés. Elles correspondent au produit des surfaces immobilières et mobilières et des valeurs au m².

C'est l'objet du chapitre 2 « Modélisation lors de la survenance d'une tempête » qui présentera une courbe sur les fréquences et une autre sur les coûts moyens en fonction de critères discriminants. La méthode utilisée est un modèle linéaire généralisé. Des courbes d'endommagement seront alors ajustées selon le portefeuille souscrit au moment de la survenance de la tempête.

III.2.3.2 Aléa : Probabilité de survenance

Ainsi, en appliquant les courbes de vulnérabilités en fréquence/coût moyen ou en charge (méthode déterministe) aux simulations des tempêtes associées à leur probabilité d'occurrence (méthode stochastique), on aboutit à la courbe de charge en fonction de période de retour.

La chapitre 4 « Définition courbe sinistralité et impact tarifaire » propose une méthode pour obtenir une courbe qui relie une charge des tempêtes et leur période de retour.

III.2.4 Le module résultat : mise en application

L'objectif de ce module est d'évaluer le risque de la tempête sur le portefeuille actuellement souscrit.

Il restitue les résultats des deux aléas à l'aide des modules précédents :

- Aléa lors de la survenance de la tempête :

Le module résultat permet de restituer une prévision sur le nombre de sinistres et sur la charge de la tempête tout d'abord au niveau des communes et enfin au niveau national. Pour ce faire, les vitesses de vents réelles par commune doivent être communiquées (module aléa), les caractéristiques du portefeuille de ces communes doivent être connues (module enjeu) et enfin les courbes d'endommagement selon des indicateurs de sinistralité doivent être calibrées (module vulnérabilité).

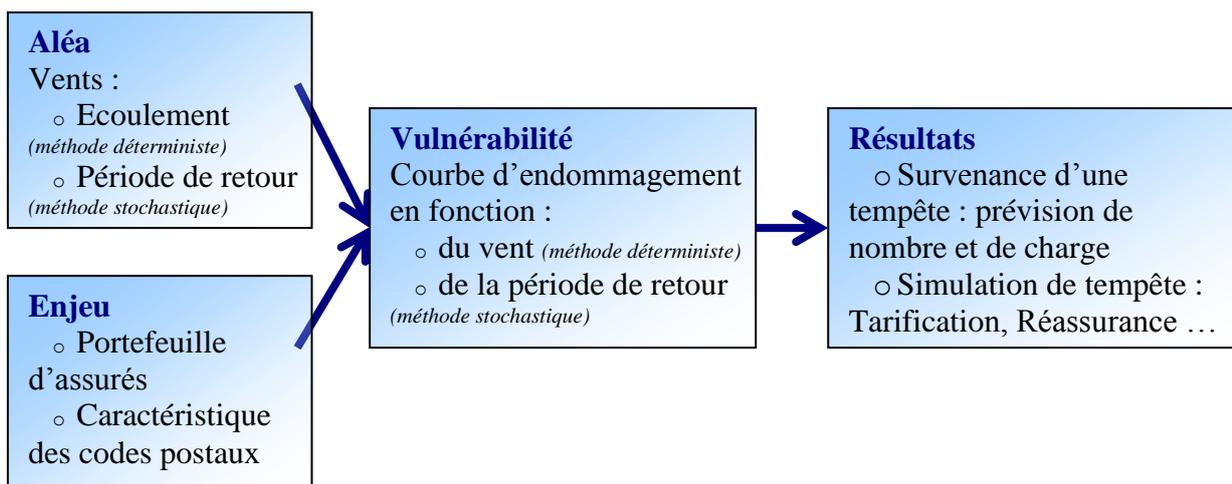
- Aléa sur la période de retour des tempêtes :

En modifiant les vitesses de vents des tempêtes simulées dans le module aléa tout en y associant une période de retour à chacune de ces tempêtes, une charge est alors estimée pour chacune de ces tempêtes dans le module résultat. Une courbe OEP peut alors être déterminée (relation entre charge de sinistres et période de retour) et alimentée le module vulnérabilité. En simulant des tempêtes à l'aide de cette courbe OEP, on peut ainsi évaluer la pertinence du tarif actuel en estimant la prime pure annuelle moyenne et pour établir un programme de réassurance.

IV. Synthèse

Ce chapitre a dressé l'intérêt d'étudier le risque tempête. Les sinistres de masse et les sinistres graves sont des risques relativement maîtrisés contrairement aux risques climatiques dont le caractère est très aléatoire. L'exclusion de certains événements climatiques (tempêtes, grêle, neige) du régime des catastrophes naturelles depuis 1990 nous oblige à bien les appréhender. Les traités de réassurance ne proposent pas de garantie illimitée pour ces périls contrairement aux catastrophes naturelles. L'enjeu majeur de la solvabilité des assureurs porte alors sur le risque tempête.

Celui-ci se caractérise par l'aléa sur l'estimation de la charge lors de la survenance d'une tempête et par l'aléa sur la période de retour de la survenance. Les logiciels de marché sont des outils dont les méthodes restent obscures et ne sont pas adaptés aux risques français. Néanmoins, le principe de leurs modules, schématisés ci-dessous, convient à évaluer le risque tempête.



La surveillance des phénomènes d'une tempête (déplacement des masses d'air chaud et froid, courant-jet ...) permet d'anticiper sa survenance. Quelques jours avant son arrivée, en fonction des vitesses de vents estimées et des informations actualisées des modules enjeu et vulnérabilité, des nombres de sinistres sont prévus pour connaître la pertinence du déclenchement d'un Plan d'Intervention Exceptionnel. Quelques jours après sa survenance, les vitesses sont réajustées selon les relevés des stations météorologiques et les charges de sinistres sont estimées pour provisionner l'évènement dans le cadre des arrêtés des comptes.

A chaque tempête simulée, des charges sont estimées ainsi que des périodes de retour. Ainsi, une courbe de charge de sinistres en fonction de période de retour est définie pour répondre à des besoins de tarification et de réassurance.

Chapitre 2. APPLICATION D'UN MODELE LORS DE LA SURVENANCE D'UNE TEMPETE

I. Objectif

Le modèle proposé dans ce second chapitre devra répondre au premier aléa sur l'estimation des nombres de sinistres et des coûts engendrés lors de la survenance d'une prochaine tempête en moyenne pour répondre aux besoins décrits précédemment. Ces estimations sont cantonnées sur la France métropolitaine et sur le portefeuille Multirisque Habitation MRH.

Le modèle va donc s'appuyer sur les principes des 4 modules. En effet, l'évaluation de la sinistralité passe par les éléments du module aléa sur les vitesses de vent et du module enjeu sur les caractéristiques du risque.

La surveillance quotidienne des conditions météorologiques (déplacement des masses d'air, courant-jet ...) permet d'anticiper l'arrivée d'un évènement tempétueux et d'estimer les vitesses de vent. Dès lors que la tempête est survenue, elles sont relevées par stations météorologiques et lissées sur tous les codes postaux de la France métropolitaine.

Les données sur les critères du portefeuille MRH peuvent être collectées pour estimer la sinistralité avant et après la tempête.

Les modélisations vont consister à trouver une relation entre l'intensité du vent et les critères influençant la sinistralité. C'est l'un des principes du module vulnérabilité.

Ainsi, cette partie a pour objet de rechercher les critères discriminants sur les nombres de sinistres et les coûts moyens puis d'établir des modèles linéaires généralisés sur les deux indicateurs. L'analyse va porter sur les données de 6 tempêtes françaises majeures de 1999 à nos jours.

II. Recherche des critères discriminants

II.1. Présentation des données

II.1.1 Tempêtes historiques

La modélisation de la sinistralité porte sur les 6 tempêtes historiques :

- Lothar et Martin en 1999
- Klaus et Quinten en 2009
- Xynthia en 2010
- Joachim en 2011

L'importance des tempêtes est illustrée dans le tableau ci-dessous en mettant LOTHAR en base 100.

Tempêtes	Dates	Trajectoire	Fréquence	Coût Moyen	Prime Pure
LOTHAR	25-26/12/1999	nord	100.0%	100.0%	100.0%
MARTIN	27-28/12/1999	sud ouest	67.8%	87.0%	59.0%
KLAUS	24/01/2009	sud ouest	28.8%	115.2%	33.2%
QUINTEN	8-9/02/09	centre ouest au nord est	6.8%	75.0%	5.1%
XYNTHIA	26/02/2010	centre ouest au nord est	15.3%	90.7%	13.8%
JOACHIM	15-16/12/2011	centre ouest au nord est	5.1%	55.3%	2.8%

Données COVEA

II.1.2 Extraction et présentation des critères

Les variables portefeuilles MRH et sinistres jugées explicatives des tempêtes historiques ont été extraites des différents systèmes d'informations décisionnels du groupe COVEA :

- portefeuilles : pour chaque code postal, le nombre de contrats actifs par type d'habitat (maison, mobilhome, appartement ...) et par qualité juridique (propriétaire, propriétaire non occupant, locataires ...).
- sinistres : pour chaque numéro de sinistre, le coût du dossier, le type d'habitat, la qualité juridique, la garantie sinistrée et le montant de la franchise.

La vision des données sinistres correspond à celle au moment de l'extraction. On considère que les dossiers des tempêtes historiques sont clos.

A la réception de ces données sinistres, il a été nécessaire de les retraiter pour qu'elles puissent être comparables et exploitables.

En effet, pour la tempête Klaus, le groupe COVEA a décidé de ne pas appliquer la franchise, cette dernière restant donc à la charge de l'assureur. Face à cette décision, le nombre de dossiers traités a donc augmenté d'une part (supplément correspondant aux dossiers dont les évaluations sont inférieures à la franchise) et d'autre part, le coût de chaque dossier a également progressé du montant de la franchise. Ainsi, les dossiers de la tempête KLAUS ont été « corrigés » de l'abandon de franchise.

Concernant l'évènement Xynthia, deux périls sont survenus : la tempête conjuguée à l'inondation. Cette dernière rentre dans le régime des catastrophes naturelles. C'est pourquoi, les sinistres inondations sont retraités pour ne tenir compte que des dommages liés au vent.

L'exhaustivité de ces informations n'a pu être collectée pour les 3 sociétés du groupe COVEA. Les sociétés 1 et 2 ont transmis les données concernant les tempêtes de 2009 et 2010. La société 3 a communiqué en plus celles de 1999 et 2011.

En complément des données portefeuille et sinistre, les données sur les vitesses de vent et sur les zones relatives aux normes de construction ont été collectées par code postal.

Les vitesses de vent relevées ont été analysées et lissées à chaque code postal. Les vitesses de vent trop faibles et trop élevées ont été écartées de la modélisation pour ne pas retenir des vitesses de vent avec de trop faibles volumétries. De plus, elles ont été arrondies avec un pas de 5 km/h afin de gommer les quelques imprécisions de mesure du vent.

COVEA a découpée la France métropolitaine en 4 zones selon la vulnérabilité des bâtiments en s'appuyant sur différentes sources et notamment aux zonages des normes de construction en vigueur.

En somme, à **chacune de ces tempêtes et à la maille du code postal** (agrégation des données sinistres), **sont associées différentes informations : Portefeuille, Sinistralité, Zone de norme de construction et Vitesses de vents.**

Définitions et présentations des critères :

▪ Type de portefeuille

Cette variable croise les critères :

- Type d'habitat :
 - maisons : « Maison, Mobil home, Château, Copropriété, Immeuble locatif, dépendances ... »
 - appartements : « Appartement »
- Qualité juridique
 - propriétaires : « Propriétaire occupant ou non, bailleur, syndic bénévole ... »
 - locataires : « Locataire »

Dans la base constituée, cette variable est nommée « type_port » et contient trois modalités :

- propriétaires de maison,
- propriétaires d'appartement,
- locataires (ce dernier regroupe les maisons et les appartements : pour des raisons de volumétrie et d'enjeu très faible).

▪ Société

Les modalités de la variable « societe » sont 1, 2 et 3, correspondant aux 3 enseignes du groupe COVEA

▪ Tempêtes

La variable « tempête » contient les 6 tempêtes historiques précédemment citées.

▪ Zones de normes de construction

Cette variable « norm_const » découpe la France en 4 zones. Elles sont numérotées de 1 à 4, de la norme la moins exigeante à la plus exigeante.

- Zones d'urbanisation

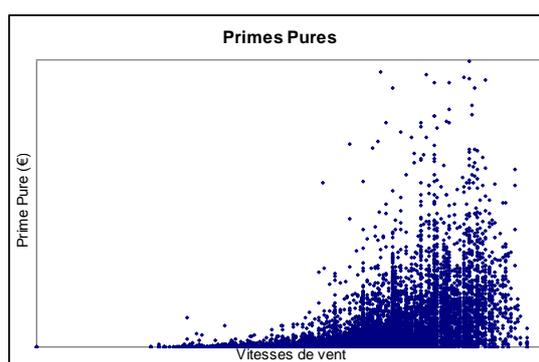
Ce critère « type_ville » permet de prendre en compte la densité de la ville, en distinguant la commune (de 0 à 5000 habitants) et la grande ville (plus de 30 000 habitants), en passant par la petite ville (de 5 000 à 10 000 habitants) et la moyenne ville (de 10 000 à 30 000 habitants).

Il a été décidé de ne pas faire subir aux coûts des sinistres Lothar et Martin d'inflation. En effet, il a été noté que pour des dommages identiques, le coût moyen est resté stable depuis 1999 et se justifie par une bonne maîtrise des coûts de sinistre depuis la création de COVEA AIS (direction de gestion des sinistres commune pour les 3 enseignes).

II.2. Profil des courbes d'endommagement

Dans cette partie, nous allons visualiser les profils de la sinistralité en fonction de la vitesse de vent. Ces profils sont déclinés et commentés selon les indicateurs de sinistralité traditionnels (fréquence et coût moyen ainsi que, de manière plus succincte, la prime pure) et selon les critères sélectionnés.

II.2.1 Courbe de prime pure



Ce graphique donne les primes pures de chaque code postal des 6 tempêtes historiques en fonction de la vitesse de vent. Il illustre des primes pures progressant globalement avec l'intensité du vent.

Notons S_{ij} le coût du $j^{\text{ème}}$ sinistre du $i^{\text{ème}}$ code postal, Y_i la charge pour le $i^{\text{ème}}$ code postal et N_i le nombre de sinistres pour ce code postal.

La charge pour le code postal i s'écrit :

$Y_i = S_{i1} + S_{i2} + \dots + S_{iN_i}$ avec le nombre N_i et les coûts S_{ij} 2 variables aléatoires.

La charge totale aléatoire est la somme des sinistres imputables à chaque code postal ($\sum_{i=1}^{na} Y_i$). On peut cesser de s'intéresser à l'identification des sinistres propres au code postal en posant :

$N = N_1 + N_2 + N_3 + \dots + N_{na}$ où na représente le nombre de code postaux.

On peut alors réécrire :

$$\sum_{i=1}^{na} Y_i = \sum_{i=1}^{na} \sum_{j=1}^{N_i} S_{i,j} = \sum_{k=1}^N Y_k$$

Sous réserve de faire deux hypothèses, on peut alors calculer l'espérance de la charge de sinistres :

- Indépendance et stationnarité des coûts de sinistres : les variables aléatoire S_{ij} sont indépendantes et identiquement distribuées (iid).
- Indépendance fréquence-coût : la distribution commune des S_{ij} ne dépend pas de la valeur prise par N_i .

L'espérance de la charge s'écrit :

$$E(S) = E(N) \cdot E(Y).$$

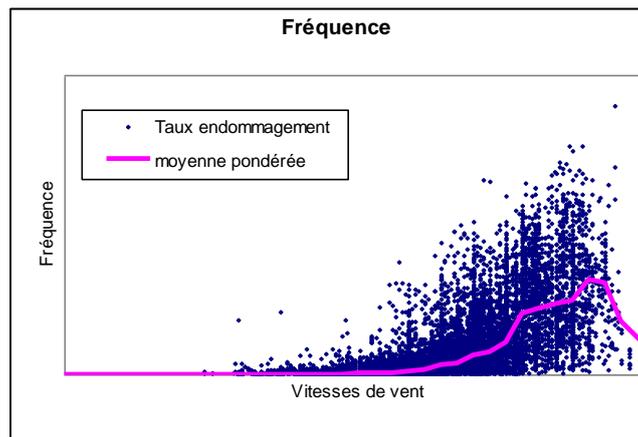
et devient :

$$E(S/X) = E(N/X) \cdot E(Y/X)$$

lorsque l'on introduit p variables explicatives $X = (X_1, \dots, X_p)$ avec S , N et Y des vecteurs aléatoires $i=1 \dots na$

Ainsi, l'estimation de la prime pure, charge rapportée au nombre de contrat en portefeuille, en vision moyenne $E(S/X)$ sera décomposée en une prévision sur les fréquences $E(N/X)$ et une autre sur les coûts moyens $E(Y/X)$.

II.2.2 Courbe de fréquence (N)



Le graphique ci-dessus montre les fréquences, encore appelées taux d'endommagement, de chaque code postal pour l'ensemble des 6 tempêtes en fonction de la vitesse de vent.

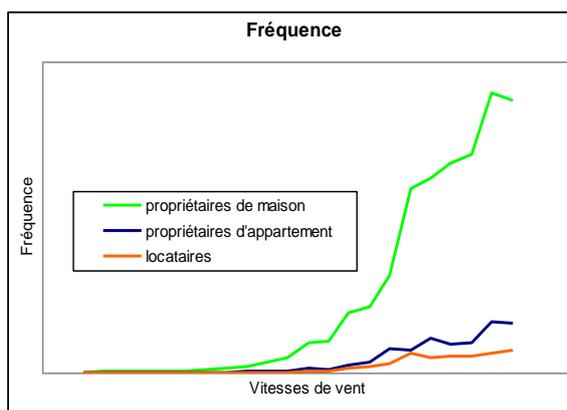
Le taux d'endommagement moyen est le rapport entre la somme des nombres de sinistres et le nombre de contrat pour une vitesse de vent donnée. La courbe de ce taux moyen est relativement stable proche de 0 jusqu'à une certaine vitesse de vent. Puis, elle croît de façon exponentielle pour les vitesses de vent de forte intensité.

Les parties suivantes décomposent cette courbe de vision moyenne en fonction de différents critères : type de portefeuille, société, tempêtes, zones de normes de construction et zones d'urbanisation.

Cette analyse descriptive univariée permet d'illustrer les propos du module enjeu, décrit dans le premier chapitre.

❖ Type de portefeuille

Les propriétaires de maison sont plus sensibles au risque tempête que :

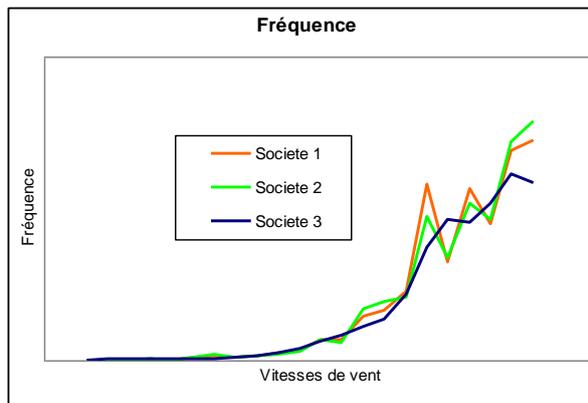


- les propriétaires d'appartement (du fait de la structure du bâtiment : pas de toiture ... ou impact uniquement pour l'assurance immeuble et pas pour l'assurance MRH)

- les locataires (concernés par les biens mobiliers et non les biens immobiliers qui sont à la charge des propriétaires du bien).

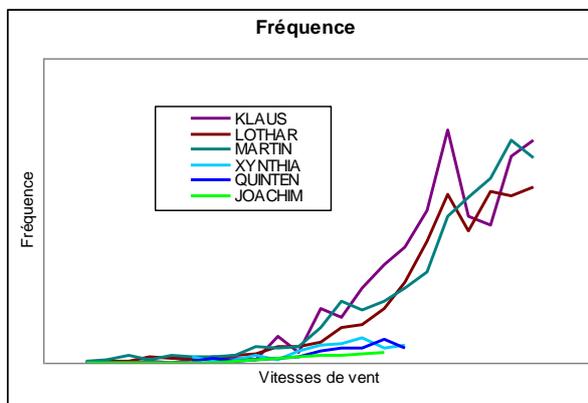
Les critères suivants sont analysés sur les propriétaires de maison où se trouve l'enjeu majeur.

❖ Société



Ce graphique permet de comparer les courbes des 3 sociétés et de montrer que ce critère ne semble pas être discriminant. En effet, il n'y a pas de raison particulière de dissocier les bâtiments des assurés selon la société et plus particulièrement pour une même zone de norme de construction.

❖ Tempêtes

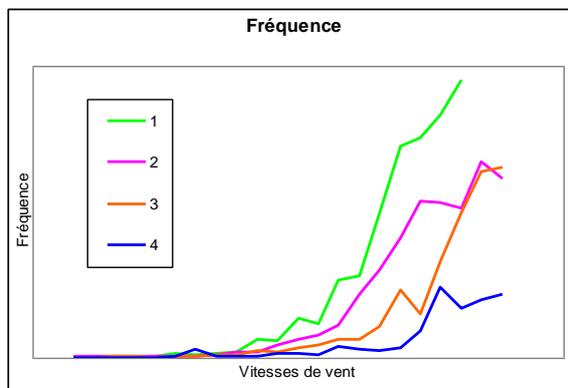


Les taux d'endommagement par tempête semblent distinguer celles de forte intensité et celles de faible intensité.

Les fréquences des tempêtes mineures (Joachim, Quinten et Xynthia) progressent peu tandis que celles des tempêtes majeures (Lothar, Martin et Klaus) croissent fortement à partir d'une certaine vitesse de vent.

On note par ailleurs qu'au sein de ces deux familles de tempêtes les taux moyens restent relativement proches.

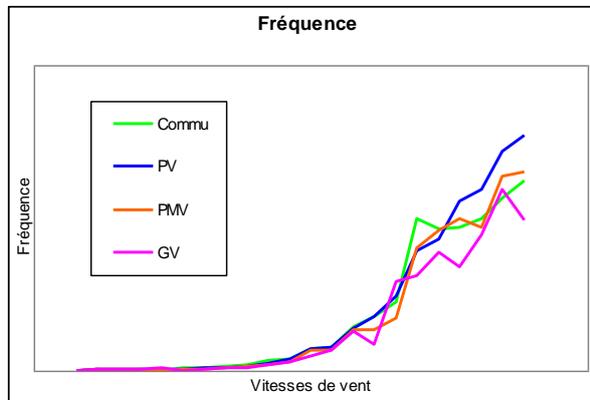
❖ Zones de normes de construction



Le taux d'endommagement est clairement différent selon la zone de normes de construction.

Moins les normes de construction sont exigeantes, plus les maisons sont endommagées par le vent et inversement.

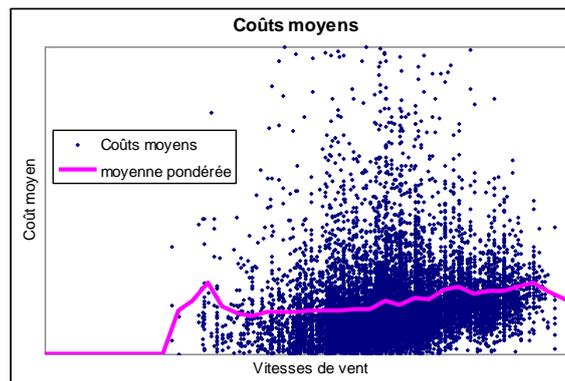
❖ Zones d'urbanisation



La densité de population joue peu sur le niveau de la fréquence. On distingue tout de même que les communes et les petites villes semblent plus impactées que les moyennes et grandes villes.

En effet, la concentration des bâtiments joue comme « un rempart » face aux rafales de vents.

II.2.3 Courbe de coût moyen (Y)

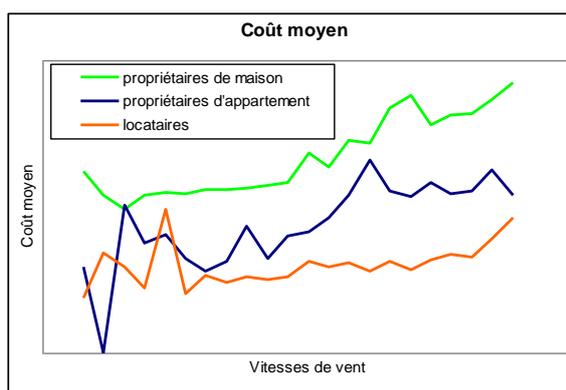


Le graphique ci-dessus montre tous les coûts moyens enregistrés au niveau de chaque code postal. Ils progressent également en fonction de la vitesse de vent. En effet, plus la vitesse de vent est élevée plus les dommages sont importants.

Ces dommages peuvent être de différentes natures en fonction de la vitesse de vent : des mobiliers extérieurs, des antennes, une tuile jusqu'aux chutes d'arbre, la couverture complète avec infiltration d'eau voir même la structure même du bâtiment assuré.

Comme pour les fréquences, les parties suivantes présentent une analyse descriptive univariée des courbes de coût moyen selon différents critères : type de portefeuille, société, tempêtes, zones de normes de construction et zones d'urbanisation.

❖ Type de portefeuille

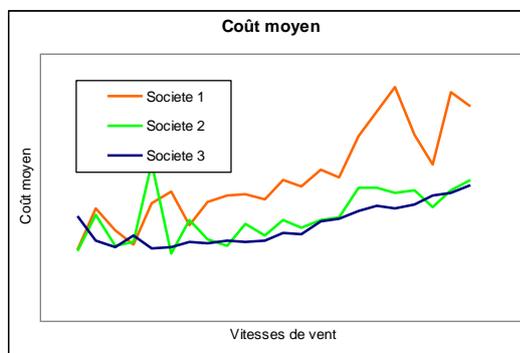


La segmentation du portefeuille est un critère important sur le coût moyen.

Les dommages sont différents selon la qualité juridique de l'assuré et selon le type d'habitat. Le dommage le plus fréquent et un des plus coûteux est la couverture et consécutivement l'infiltration par toiture. C'est pourquoi, le coût moyen diffère selon les trois modalités.

Les critères suivants sont analysés sur les propriétaires de maison où se trouve l'enjeu majeur.

❖ Société



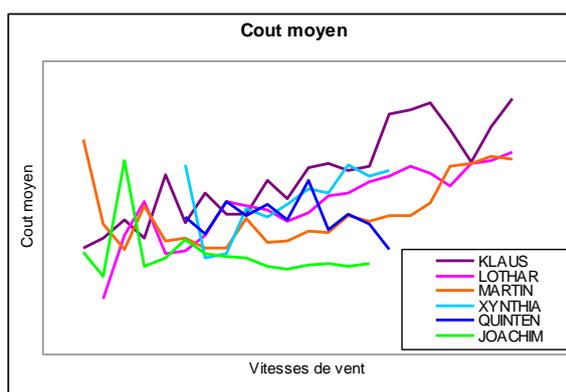
En analysant le coût moyen selon les sociétés, on note que leur coût moyen est différent car reflète le risque assuré.

En effet, les assurés de la société 1 ont des capitaux immobiliers (et mobiliers) plus importants que les deux autres sociétés. D'où des coûts plus élevés et ce, malgré un niveau de franchise clairement plus important pour la société 1.

Les coûts moyens de la société 2 sont très proches de ceux de la société 3.

Ce constat se justifie également par les valeurs des sommes assurées transmises aux courtiers pour alimenter les logiciels marché.

❖ Tempêtes

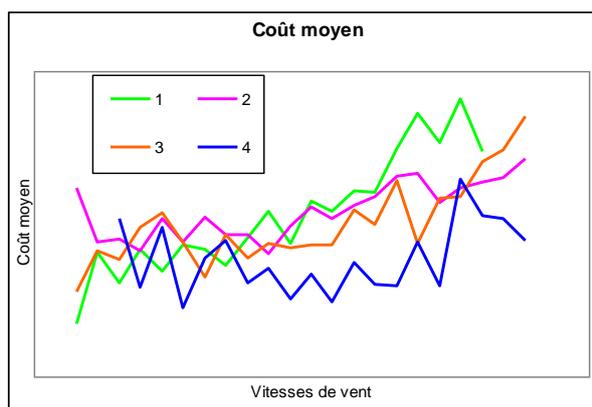


Globalement, les coûts moyens par tempête augmentent en fonction de la vitesse de vent.

En comparant les tempêtes entre elles, certaines semblent plus coûteuses Klaus en tête suivie par Xynthia et Lothar. Joachim est clairement en deçà et Martin semble plus faible.

Il faut tout de même être prudent sur l'interprétation des résultats puisque Lothar, Martin et Joachim concernent uniquement la société 3 et n'intègrent donc pas les sinistres de la société 1 qui ont des coûts moyens plus élevés.

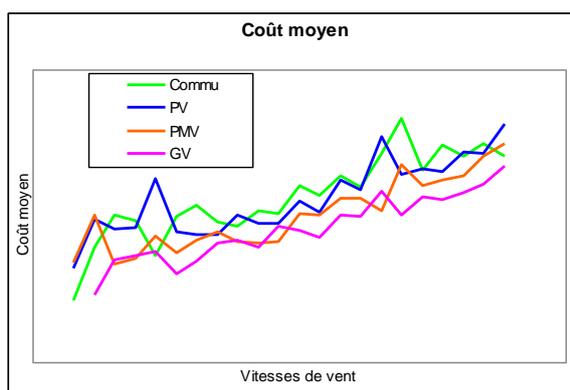
❖ Zones de normes de construction



Jusqu'à une certaine vitesse de vent, les coûts moyens par zone de normes de construction paraissent proches et stables. Puis, ils se distinguent en progressant plus ou moins rapidement en fonction de la norme de construction.

Dès lors que la structure des bâtiments est touchée, les coûts sont plus importants et ce graphique sur les normes de construction le fait bien apparaître.

❖ Zones d'urbanisation



La densité de population influe sur le niveau du coût moyen. Cela se justifie par la structure des bâtiments différents entre les zones urbaines et rurales.

II.3. Corrélation entre les variables

Cette partie vise à analyser les corrélations entre les variables décrites précédemment. En effet, c'est une des hypothèses pour calculer l'espérance de la charge.

Les statistiques pour la corrélation entre les variables deux par deux sont nombreuses. Celle choisie est la corrélation de V de Cramer.

Cet indicateur est compris entre 0% et 100% et est défini comme suit :

$$V = \sqrt{\frac{\chi^2}{n * \min(L-1, C-1)}}$$

$$\chi^2 = \sum_{k=1}^K \frac{(o_k - e_k)^2}{e_k} \quad \text{Ecart entre les effectifs observés et théoriques}$$

$$o_k \text{ valeur observée ; } e_k = \frac{n_l * n_c}{n} \text{ valeur théorique en cas de dépendance}$$

n nombre d'observation total ; L, C nombre de modalité des 2 variables

Le tableau ci-dessous donne la corrélation de V de Cramer entre les variables étudiées :

	enseigne	tempete	type_port	Norm_const	type_ville	vent
enseigne	100%					
tempete	51%	100%				
type_port	6%	9%	100%			
Norm_const	10%	25%	2%	100%		
type_ville	2%	6%	7%	9%	100%	
vent	19%	33%	8%	37%	12%	100%

Les variables tempete, vent et Norm_const semblent corrélées. La caractéristique de la tempête dépend des vitesses de vent, mais ces variables ne peuvent pas être retirées du modèle puisque pour certaines vitesses de vent, les fréquences divergent entre les tempêtes.

La proportion de sinistres dans chaque zone de normes de construction est relativement équivalente quelle que soit la tempête. C'est pourquoi le coefficient de corrélation ressort entre les variables tempête (et vent) et normes de construction. Néanmoins, le risque n'est pas le même selon les zones de construction.

De plus, il est à noter que la variable « tempête » ne pourra pas être retenue dans le modèle final puisque les tempêtes futures ne correspondront pas aux tempêtes historiques.

La corrélation forte observée entre les variables enseigne et tempête se justifie par le fait que les données Lothar, Martin et Joachim sont exclusivement de la société 1, entraînant par la même occasion celle avec le vent.

III. Modélisation des tempêtes historiques

III.1. Méthode d'estimation de la sinistralité

Les analyses descriptives univariées montrent un profil haussier en fonction de l'intensité du vent. Ces profils sont identiques d'une variable à l'autre et d'une modalité à l'autre mais avec des progressions plus ou moins importantes.

L'objectif est d'appliquer un modèle d'ajustement de la distribution de sinistres en fonction des critères discriminants.

C'est pourquoi, les modèles linéaires généralisés ont été choisis pour estimer la sinistralité. En effet, ils répondent aux deux aspects puisqu'ils permettent d'établir un lien entre la variable à expliquer et les variables explicatives et entre les modalités. Cette méthode est très utilisée dans les études actuarielles et notamment en tarification.

III.2. Rappel sur la théorie des modèles linéaires généralisés

Ce paragraphe rappelle l'aspect théorique des modélisations en faisant l'analogie entre les régressions classiques et les modèles linéaires généralisés (GLM pour Generalized Linear Model).

La présentation théorique du GLM permettra de poser les hypothèses nécessaires du modèle, d'appréhender les estimations des paramètres, de tester leur significativité et d'interpréter les résultats.

III.2.1 Régressions Classiques

III.2.1.1 Formule générique

Le modèle linéaire s'écrit sous la forme :

$$Y = X\beta + \varepsilon$$

tel que :

- Y : vecteur de la variable à expliquer
- β : vecteur des paramètres du modèle
- X : matrice des variables explicatives
- ε : matrice des erreurs

L'espérance du modèle linéaire est $E(Y) = \mu = X\beta$

Les valeurs Y_i du vecteur aléatoire Y sont indépendantes et

$$Y \sim \mathcal{N}(X\beta, \sigma^2 I_n)$$

$$\text{avec } \varepsilon \sim \mathcal{N}(0, \sigma^2 I_n)$$

III.2.1.2 Estimations

Deux méthodes sont utilisées pour estimer les paramètres β :

- maximum de vraisemblance
- moindres carrés

La solution de ces méthodes est :

$$\beta = ({}^tXX)^{-1} {}^tXT$$

III.2.1.3 Qualité du modèle

a) Indicateur de la qualité du modèle

On souhaite savoir si les variations de la variable à expliquer se justifient par le modèle.

Deux indicateurs permettent d'appréhender la justesse du modèle : la somme des écarts au carré et le coefficient de détermination.

- Tableau de l'analyse de la variance (ANOVA) :

La méthode ANOVA est une analyse de la variance.

La somme des écarts au carré s'écrit de la manière suivante :

$$\sum_{i=1}^N (Y_i - \bar{Y})^2 = \sum_{i=1}^N (\hat{Y}_i - \bar{Y})^2 + \sum_{i=1}^N (Y_i - \hat{Y}_i)^2$$

tel que \bar{Y} moyenne des observations, \hat{Y} valeurs estimées

Source	Somme des écarts au carré	Degré de liberté	Carré moyen
Modèle	$\sum_{i=1}^N (\hat{Y}_i - \bar{Y})^2$	p	$\sum_{i=1}^N (\hat{Y}_i - \bar{Y})^2 / p$
Résiduelle	$\sum_{i=1}^N (Y_i - \hat{Y}_i)^2$	N-p-1	$\sum_{i=1}^N (Y_i - \hat{Y}_i)^2 / N-p-1$
Totale	$\sum_{i=1}^N (Y_i - \bar{Y})^2$	N-1	

- Le coefficient de détermination permet de savoir si le modèle est adapté :

$$R^2 = \frac{\sum_{i=1}^N (\hat{Y}_i - \bar{Y})^2}{\sum_{i=1}^N (Y_i - \hat{Y}_i)^2}$$

Plus ce ratio est proche de 100%, plus les variables explicatives justifient les variations de la variable à expliquer.

b) Test de régression

Le test de régression est $H_0 : \beta_1 = \beta_2 = \beta_3 = \dots = \beta_p = 0$ contre la non régression $H_1 : \ll$ au moins un β est différent de 0 \gg .

C'est le test de Fisher qui permet de valider l'hypothèse. Il consiste à comparer la valeur de

$$F = \left(\frac{\sum_{i=1}^N (\hat{Y}_i - \bar{Y})^2}{p} \right) / \left(\frac{\sum_{i=1}^N (Y_i - \hat{Y}_i)^2}{(N-p-1)} \right)$$

et celle lue dans la table Fisher (P, N-P-1) ddl au seuil α .

c) Analyse des résidus

Les résidus correspondent aux écarts entre la modélisation et les observations et sont définis comme :

$$\hat{e}_i = Y_i - \mu$$

Il est nécessaire de vérifier certaines hypothèses (la normalité et l'indépendance des résidus) et d'analyser le nuage des résidus standardisés sur les valeurs prédites et chacune des variables explicatives.

III.2.2 Modèle Linéaire Généralisé

III.2.2.1 Formule générique

Le modèle linéaire généralisé s'écrit sous la forme :

$$Y = E(Y) + \varepsilon$$

mais $E(Y)$ n'est pas forcément égale à $X\beta$.

$$E(Y) = \mu \quad g(\mu) = {}^tX\beta + \xi = \eta$$

avec ξ « offset » ou variable de contrainte

g est appelé fonction de lien telle que $E(Y) = g^{-1}(X\beta + \xi)$. Il faut donc que la fonction de lien soit inversible.

Distribution de la variable explicative Y_i

La distribution d'une variable aléatoire Y_i appartient à la famille exponentielle si la fonction de densité peut être écrite sous la forme :

$$f(y_i; \phi, \theta_i) = \exp \left[\frac{y_i \theta_i - b(\theta_i)}{a(\phi)} + c(y_i, \phi) \right]$$

θ_i et ϕ sont appelés respectivement paramètre naturel (ou canonique) et paramètre de dispersion. Les fonctions a , b et c sont spécifiées en fonction du type de la loi exponentielle.

Ce paramètre de dispersion est également un paramètre d'échelle ou « nuisance parameter » et permet de gérer les lois de la famille exponentielle à 2 paramètres.

III.2.2.2 Estimations

Les estimations du modèle linéaire généralisé sont fondées sur l'étude de la log-vraisemblance (L), appelé score :

$$\frac{\partial \ln L(y_i; \phi; \theta_i)}{\partial \theta_i} = \frac{y_i - b'(\theta_i)}{a(\phi)}$$

La dérive du score devient :

$$\frac{\partial^2 \ln L(y_i; \phi; \theta_i)}{\partial \theta_i^2} = \frac{-b''(\theta_i)}{a(\phi)}$$

○ Moments de la variable explicative Y

Les propriétés suivantes permettent de déterminer l'espérance et la variance aléatoire de Y :

$$E\left(\frac{\partial \ln L(y_i; \phi; \theta_i)}{\partial \theta_i}\right) = 0$$

$$-E\left(\frac{\partial^2 \ln L(y_i; \phi; \theta_i)}{\partial \theta_i^2}\right) = E\left(\frac{\partial \ln L(y_i; \phi; \theta_i)}{\partial \theta_i}\right)^2$$

Ainsi :

$$E(Y) = b'(\theta)$$

$$V(Y) = b''(\theta) * a(\phi)$$

Le paramètre θ influe à la fois sur la moyenne et sur la variance.

Le paramètre ϕ contrôle la variance.

En somme,

$$E(Y) = g^{-1}(X\beta + \xi) = b'(\theta_i)$$

$$V(Y) = b''(\theta_i) * a(\phi)$$

○ Paramètres β

Les paramètres sont estimés par la méthode du maximum de vraisemblance, c'est-à-dire en annulant le score :

$$\frac{\partial \ln L(y_i; \phi; \theta_i)}{\partial \beta_i} = 0$$

$$\frac{\partial \ln L(y_i; \phi; \theta_i)}{\partial \theta_i} \frac{\partial \theta_i}{\partial \mu_i} \frac{\partial \mu_i}{\partial \eta_i} \frac{\partial \eta_i}{\partial \beta_i} = 0$$

$$\text{avec } \frac{\partial \ln L(y_i; \phi; \theta_i)}{\partial \theta_i} = \frac{y_i - b'(\theta_i)}{a(\phi)} \quad / \quad \frac{\partial \theta_i}{\partial \mu_i} = \frac{1}{b''(\theta_i)} \quad / \quad \frac{\partial \eta_i}{\partial \beta_i} = x_{ij}$$

$$\sum_{i=1}^N \frac{(y_i - \mu_i)x_{ij}}{V(Y_i)} \frac{\partial \mu_i}{\partial \eta_i} = 0 \quad j = 1, \dots, p$$

Cette équation est résolue de façon itérative, soit par la méthode de Newton-Raphson (via la matrice Hessienne) soit par la méthode de scoring de Fisher (via l'opposé de l'espérance de la matrice Hessienne, appelée aussi matrice d'information).

$$\beta^{r+1} = \beta^r - H(\beta^r) \frac{\partial \ln L(y_i; \phi; \theta_i)}{\partial \beta_i}$$

$$\beta^{r+1} = \beta^r + E(H(\beta^r)) \frac{\partial \ln L(y_i; \phi; \theta_i)}{\partial \beta_i}$$

III.2.2.3 Qualité du modèle

a) Test de régression

- Test de la déviance

Le test d'ajustement est une comparaison entre la vraisemblance du modèle étudié et celle du modèle saturé (modèle ayant autant de paramètres que d'observations et ajustant alors parfaitement).

La déviance est définie comme :

$$D(y, \bar{y}) = 2 (\ln L(y, y) - \ln L(\bar{y}, y))$$

Il est démontré qu'asymptotiquement $D(y, \bar{y})$ suit une loi χ^2 à $N-p$ degrés de liberté. Ainsi, le test de la déviance permet de valider H_0 : « le modèle à p paramètres est adéquat » ou valider H_1 : « le modèle à p paramètres n'est pas adéquat » selon le niveau de significativité $\alpha\%$.

- Test du Khi-deux de Pearson

$$\chi^2 = \sum_{i=1}^N \frac{(y_i - \mu_i)^2}{V(\mu_i)}$$

Le χ^2 de Pearson suit un χ^2 de $N-p$ degrés de liberté.

Ces deux tests ont des résultats très proches mais le second est très instable dès lors qu'au moins une des variables explicatives est continue et que le nombre d'observations n'est pas suffisant.

- Test d'hypothèse sur les paramètres

Le test consiste à vérifier si d paramètres sont significativement différents de 0 sur les p paramètres, c'est-à-dire :

$$H_0 : \beta = \beta_0 = (\beta_0, \dots, \beta_d)' \text{ contre } H_1 : \beta = \beta_1 = (\beta_0, \dots, \beta_p)'$$

Ainsi, $D_0 - D_1 = 2 * (\ln L(y, \hat{\beta}_1) - \ln L(y, \hat{\beta}_0)) > 0$ et $D_0 - D_1 \sim \chi^2_{p-d}$

On rejette H_0 au profit de H_1 lorsque $D_0 - D_1 > \chi^2_{p-d}$ au risque α

- Test sur les variables (Test typeIII sous SAS)

Ce test consiste à comparer le modèle contenant toutes les variables à celui excluant la variable testée.

b) Indicateur de la qualité du modèle

- Tableau de l'analyse de la déviance

L'analyse de la déviance est proche de celle de l'ANOVA : l'une étudie la déviance, l'autre la variance.

Source	Somme des écarts au carré	Degré de liberté	Carré moyen
Modèle	$D(\hat{y}, \bar{y}) = 2 (\ln L(\hat{y}, y) - \ln L(\bar{y}, y))$	p	$D(\hat{y}, \bar{y}) / p$
Résiduelle	$D(y, \hat{y}) = 2 (\ln L(y, y) - \ln L(\hat{y}, y))$	N-p-1	$D(y, \hat{y}) / N-p-1$
Totale	$D(y, \bar{y}) = 2 (\ln L(y, y) - \ln L(\bar{y}, y))$	N-1	

- Analyse de Khi-deux de Pearson

La même analyse est réalisée sur le Khi-deux de Pearson en remplaçant la formule de la déviance par celle du Khi-deux de Pearson avec les mêmes degrés de liberté.

c) Analyse des résidus

Les résidus correspondent aux écarts entre la modélisation et les observations et ces résidus bruts sont définis comme :

$$\hat{e}_i = Y_i - \mu$$

Afin de donner une stabilité au modèle, il est nécessaire d'étudier les résidus de déviance et de Pearson et d'écarter en conséquence les observations mal calibrées.

Les résidus de déviance r^d et de Pearson r^p sont définis respectivement comme :

$$r_i^d = \text{signe}(y_i - \mu_i) \sqrt{d_i} \quad \text{avec } D = \sum_{i=1}^n d_i \rightarrow \sum_{i=1}^n (r_i^d)^2 = D$$

$$r_i^p = \frac{y_i - \hat{\mu}_i}{\sqrt{V(\hat{\mu}_i)}} \rightarrow \sum_{i=1}^n (r_i^p)^2 = \chi^2$$

Les lois des résidus standardisés de Pearson et de déviance sont inconnues et dissymétriques. En effet, la loi de la variable explicative Y peut être différente de la loi Normale et la fonction de lien de la fonction identité.

Dans le cas contraire, la réalisation d'un modèle linéaire généralisé en choisissant la loi Normale et une fonction de lien « identité » revient à réaliser un modèle linéaire classique et ainsi les résidus sont normaux.

Gauss M. Cordeiro et Alexandre B.SIMAS ont, dans leur article « The distribution of Pearson residuals in generalized linear models », montré une formule asymptotique de densité des résidus de Pearson dans le cas du GLM et défini une distribution de ces résidus corrigés.

Anscombe a également proposé de faire opérer une transformation préalable (t) afin de construire des résidus suivant une loi normale :

$$r_i^a = \frac{t(y_i) - t(\hat{\mu}_i)}{t'(y_i) \sqrt{V(\mu_i)}}$$

Ce résidu n'est pas fourni sous SAS.

L'analyse des résidus standardisés consiste à étudier les nuages des résidus standardisés sur les valeurs prédites et chacune des variables explicatives.

III.2.2.4 Lois et fonctions de lien usuelles

Le tableau ci-dessous donne les valeurs des paramètres pour établir les fonctions de répartition de 3 lois usuelles.

Lois	$a(\phi)$	$b(\theta)$	$c(y_i, \phi)$
Normal $\mathcal{N}(\mu, \sigma)$	$\phi = \sigma^2$	$\frac{\theta^2}{2} = \frac{\mu^2}{2}$	$-y_i/2\sigma^2 - \ln\sqrt{2\pi\sigma^2}$
Poisson $\mathcal{P}(\lambda)$	$\phi = 1$	$e^\theta = \lambda$	$-\ln(y!)$
Gamma $\gamma(\alpha, \lambda)$	$\phi = 1/\alpha$	$-\ln(-\theta) = -\ln(\frac{\alpha}{\beta})$	$(\alpha-1)\ln y + \alpha \ln \alpha - \ln \Gamma(\alpha)$

Les fonctions de lien usuelles sont :

Nom de lien	Fonction de lien
Identité	$g(\mu) = \mu$
Log	$g(\mu) = \log(\mu)$
Logit	$g(\mu) = \log\left(\frac{\mu}{1-\mu}\right)$
Réciproque	$g(\mu) = \frac{1}{\mu}$

Lorsque la fonction de lien est logarithmique, le modèle est dit multiplicatif car

$$g(\mu_i) = \beta_1 x_{i,1} + \dots + \beta_p x_{i,p} + \xi \text{ avec } g = \ln$$

$$\mu_i = \exp(\beta_1 x_{i,1}) * \dots * \exp(\beta_p x_{i,p}) * \exp(\xi)$$

III.3. Modélisation des nombres et des coûts de sinistres

Les parties précédentes ont montré une analyse descriptive des données d'entrée du modèle et ont présenté l'aspect théorique du GLM. La modélisation de cette partie a été réalisée sous le logiciel SAS avec la procédure GENMOD.

Pour rappel, ces modèles devront répondre aux besoins lors de la survenance d'une tempête :

- le nombre de sinistre : pour gérer les flux de sinistres
- les coûts de sinistres : pour provisionner l'évènement et pour communiquer le chiffre pour des besoins internes et externes.

Un premier modèle portera sur les fréquences pour déterminer les nombres de sinistres. Un second sur les coûts moyens pour déterminer les charges de sinistres à l'aide du modèle précédent.

Ces modèles vont s'articuler sur une description des hypothèses (loi, fonction de lien, contrainte) puis sur l'analyse d'un modèle initial avec toutes les variables, pour percevoir les significativités et les interprétations des paramètres (et donc des variables et de leurs modalités). Cette analyse amènera à retenir, à croiser et à lisser les paramètres. Enfin, un backtesting permettra d'étudier les écarts entre le prévisionnel et le réel et donc la pertinence du modèle.

III.3.1 Modélisations des Fréquences

III.3.1.1 *Hypothèses du modèle*

Le modèle sur les nombres de sinistres au code postal est un modèle linéaire généralisé poissonnien et multiplicatif avec une fonction de lien logarithmique. Il a pour contrainte le nombre de polices afin de modéliser la fréquence de sinistres.

Rappel sur la loi de Poisson de paramètre λ

$$\forall k \in \mathbb{N}, \mathcal{P}(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}$$
$$E[X] = Var[X] = \lambda$$

Propriété

Si X_1, \dots, X_n sont des variables aléatoires indépendantes de lois poisson de paramètres

$$\lambda_1, \dots, \lambda_n \text{ alors } \sum_{i=1}^N X_i \text{ suit une loi } \mathcal{P}\left(\sum_{i=1}^N \lambda_i\right)$$

Le nombre de sinistres déclarés pour un code postal i (N_i) dépend du nombre de contrat noté $port_i$ (offset noté $\exp(\xi)$) dans la partie 55).

Le nombre de sinistre est une variable aléatoire discrète et de comptage. C'est pourquoi, la loi retenue est la loi de poisson.

On suppose alors que le modèle de nombre poissonnien a pour moyenne :

$$E(N_i / X) = port_i * \exp(\beta'X_i)$$

Le modèle de nombre de sinistres s'écrit de la manière suivante :

$$N_i \sim \mathcal{P}(port_i \cdot \exp(\beta'X_i))$$

III.3.1.2 Significativité des critères

a) Discrimination et interprétation

Au regard de la présentation du module enjeu et des analyses descriptives de chacun des critères, le premier modèle consiste à intégrer l'ensemble des variables explicatives à priori ainsi que la totalité de leurs modalités afin de connaître la significativité de chacune d'elle.

Ce modèle se décompose de la manière suivante :

$$N_i \sim \mathcal{P}[port_i \cdot \exp(1.\beta_1 + societe.\beta_{2,3,4} + tempete.\beta_{5,\dots,10} + type_port.\beta_{11,12,13} + norme_const.\beta_{14,15,16,17} + type_ville.\beta_{18,19,20,21} + vent \beta_{22,\dots})]$$

Les modalités des variables ont été décrites pages 42 et 43

- Qualité du modèle

Le premier modèle complet permet d'obtenir les valeurs suivantes aux tests de régression :

Critères d'évaluation de l'adéquation			
Critère	DDL	Valeur	Valeur/DDL
Deviance	12E4	169130.7897	1.4269
Pearson Chi-Square	12E4	325629.7981	2.7472

Ces indicateurs montrent que le modèle est relativement de bonne qualité.

- Significativité et détermination de paramètres :
 - des variables

Statistique LR pour Analyse de Type 3

Source	DDL Num.	DDL Res.	Valeur F	Pr > F	Khi-2	Pr > Khi-2
Vent	17	118531	5409.64	<.0001	91963.8	<.0001
Type_port	2	118531	24430.0	<.0001	48860.0	<.0001
tempete	5	118531	1746.81	<.0001	8734.04	<.0001
Norm_const	3	118531	2088.02	<.0001	6264.05	<.0001
Type_ville	3	118531	705.46	<.0001	2116.37	<.0001
societe	2	118531	332.40	<.0001	664.80	<.0001

Le test de type III (défini p. 58) indique que toutes les variables sont significatives dans l'estimation des nombres de sinistres, puisque Pr > Khi-2 prend la valeur < 0.0001.

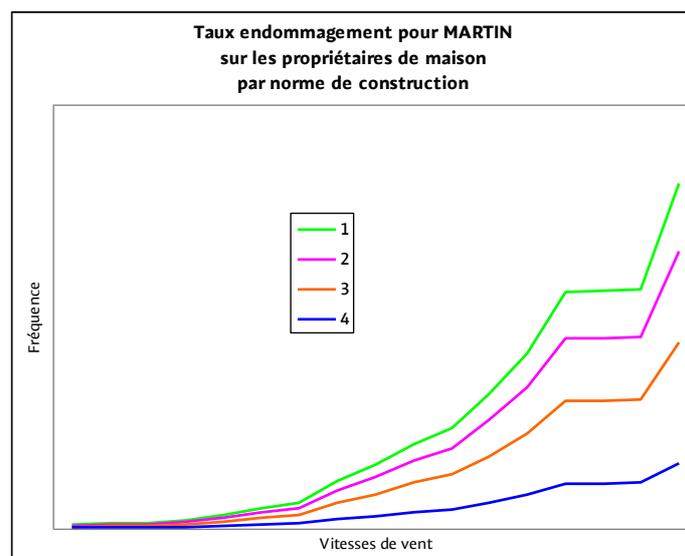
- des modalités

Variabiles	Modalités	exp(coeff)	Pr > Khi-2
societe	societe 2	81.7%	<.0001
societe	societe 3	107.6%	<.0001
societe	societe 1	100.0%	.
tempete	JOACHIM	74.1%	<.0001
tempete	KLAUS	249.1%	<.0001
tempete	LOTHAR	189.1%	<.0001
tempete	MARTIN	220.8%	<.0001
tempete	QUINTEN	65.2%	<.0001
tempete	XYNTHIA	100.0%	.
Type_port	locataire	8.2%	<.0001
Type_port	prop_app	18.5%	<.0001
Type_port	prop_mais	100.0%	.
norme_const	1	525.1%	<.0001
norme_const	2	421.2%	<.0001
norme_const	3	282.6%	<.0001
norme_const	4	100.0%	.
type_ville	Commu	104.2%	<.0001
type_ville	GV	64.5%	<.0001
type_ville	MV	85.4%	<.0001
type_ville	PV	100.0%	.

Les paramètres, présentés ci-contre, ont été déterminés par le maximum de vraisemblance et convertis en exponentielle (fonction de lien logarithmique), afin de comparer les modalités d'une variable avec une modalité de référence en base 100.

Le test de Wald nous montre que chacune de ces modalités est également significative.

Les coefficients des paramètres des vitesses sont illustrés graphiquement ci-dessous. Chacun de ces paramètres est également significatif.



- Interprétations des paramètres

Le tableau précédent donne les coefficients issus du modèle linéaire généralisé et permet de montrer l'importance des modalités au sein d'une variable explicative.

Ces coefficients traduisent bien les analyses descriptives que nous avons illustrées précédemment :

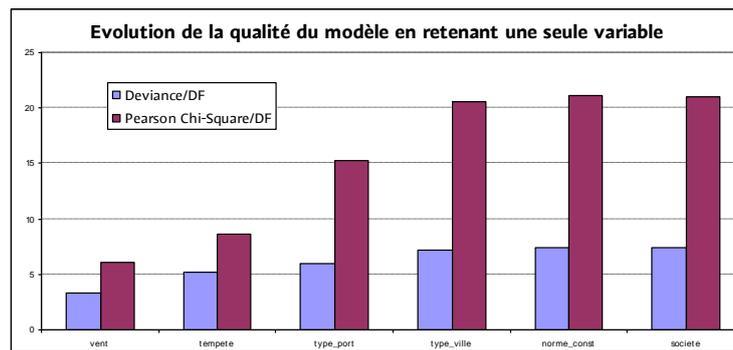
- Société : On note des écarts entre les sociétés. Les fréquences de la société 2 sont au moins 20% en deçà des autres sociétés toutes choses égales par ailleurs. Après échange avec les métiers sinistres, nous n'avons pas identifié les raisons de cet écart.
- Tempête : On retrouve deux familles de tempêtes, celle de forte intensité (Lothar, Martin et Klaus) qui ont des coefficients proches de 200% et celles de faible intensité (Joachim, Quinten et Xynthia).
- Type portefeuille : Les propriétaires de maison sont largement plus exposés. Les propriétaires d'appartement et les locataires sont nettement moins influents sur les fréquences, les coefficients sont respectivement de 18% et de 8%.
- Normes de construction : L'exigence des normes de construction influe sur la fréquence. Les coefficients sont décroissants en lien avec l'exigence.
- Type de ville : L'urbanisation des villes joue également sur la fréquence de sinistre, puisque les grandes villes sont moins impactées que les plus petites. Les coefficients sont faibles pour les villes de plus de 10 000 habitants.
- Vitesse de vent : Illustrée dans le graphique, le modèle retranscrit que plus la vitesse de vent est élevée, plus les biens sont exposés.

b) Influence des variables explicatives

La méthode forward et le test de wald permettent de connaître les variables les plus discriminantes.

o Méthode forward

Le principe consiste à introduire au fur et à mesure les variables explicatives dans le modèle. La première étape est de déterminer la variable qui minimise la déviance ou le khi-deux de Pearson. Ainsi, la variable la plus influente est la vitesse de vent, comme l'illustre le graphique ci-dessous.



Les étapes suivantes consistent, sur le même principe, à ajouter aux variables déjà sélectionnées, la variable la plus significative, jusqu'à la dernière.

	vent	+ type_port	+ tempete	+ norme_const	+ type_ville	+ societe
Deviance/DF	3.31	1.94	1.63	1.49	1.44	1.43
Pearson Chi-Square/DF	6.06	4.37	3.27	3.16	2.77	2.75

La deuxième variable la plus pertinente introduite dans le modèle après celle du vent est le type de portefeuille, qui permet d'améliorer sensiblement la qualité du modèle. L'introduction des autres variables minimise les indicateurs de qualité de manière plus marginale.

o Analyse du test de wald (type III)

Ce test consiste à comparer le modèle (rapport de vraisemblance) contenant toutes les variables à celui sans la variable à tester.

Statistique LR pour Analyse de Type 3

Source	DDL Num.	DDL Res.	Valeur F	Pr > F	Khi-2	Pr > Khi-2
Vent	17	118531	5409.64	<.0001	91963.8	<.0001
Type_port	2	118531	24430.0	<.0001	48860.0	<.0001
tempete	5	118531	1746.81	<.0001	8734.04	<.0001
Norm_const	3	118531	2088.02	<.0001	6264.05	<.0001
Type_ville	3	118531	705.46	<.0001	2116.37	<.0001
societe	2	118531	332.40	<.0001	664.80	<.0001

Le classement des valeurs du khi-deux par ordre décroissant confirme le même ordre de significativité des variables dans le modèle de nombres de sinistres.

III.3.1.3 Présentation du modèle final

Pour obtenir un modèle qui puisse s'adapter à la survenance d'une tempête future, il est nécessaire de retravailler les variables et leurs modalités. Cela passe par différentes étapes qui visent à retirer les variables, à les croiser entre elles, à regrouper des modalités ainsi qu'à lisser les coefficients des paramètres. Différentes possibilités auraient pu être retenues. La présentation suivante est axée uniquement sur les étapes qui ont conduit au modèle final.

a) Présentation des étapes :

- 2^{ème} modèle : exclusion de la société et regroupement des tempêtes par famille

Certes, la société semble ressortir comme variable discriminante mais il n'y a pas de raison de distinguer les biens assurés des 3 sociétés. De plus, l'analyse précédente sur la significativité a montré que c'était la variable qui influait le moins. C'est pourquoi, cette variable n'est pas retenue.

Au regard des résultats sur la notion de tempêtes, les 6 tempêtes ont été regroupées en deux catégories : tempêtes majeures et mineures. Ce regroupement a été nécessaire pour deux raisons. Premièrement, une tempête future ne sera pas identique aux tempêtes modélisées, elle aura ses propres caractéristiques. Deuxièmement, on a constaté qu'en fonction de la puissance de la tempête, pour une même vitesse de vent donnée, la fréquence est bien différente.

- 3^{ème} modèle : interaction entre les familles de tempêtes et le vent

Dorénavant, le modèle suppose un lien entre les familles de tempêtes et la vitesse de vent. On entend ici une influence différente de la vitesse de vent sur la fréquence selon la famille de la tempête. C'est-à-dire que les courbes de fréquence en fonction de la vitesse de vent ont un profil différent selon la famille de tempête.

Cette interaction permet de minimiser les écarts entre les tempêtes mais de manière marginale tout de même.

- 4ème modèle : retrait des observations aberrantes

Le 4^{ème} modèle consiste à retirer du modèle 3 les observations « aberrantes » dont les résidus de Pearson standardisés sont extrêmes (0.1%), c'est-à-dire lorsque le quantile est inférieur à 0.05% ou supérieur à 99.95%.

Ce retrait permet de rendre le modèle plus pérenne dans le temps et donc de ne pas être influencé par les valeurs extrêmes.

b) Qualité des différents modèles :

	1er modèle	2ème modèle	3ème modèle	4ème modèle
Déviance / ddl	1.4269	1.4654	1.4571	1.3428
Pearson Chi-Square / ddl	2.7472	2.8798	2.8897	2.3589

Le retrait de la variable « societe » et le regroupement des tempêtes (modèle 2) ne dégrade pas vraiment la qualité du modèle.

Ajouter l'interaction (modèle 3) influe peu sur la qualité du modèle. Par contre, le retrait des observations « aberrantes » (modèle 4) a plus de conséquences sur la qualité du modèle : baisse manifeste de la déviance et du khi-deux de Pearson.

c) Etape complémentaire : lissage et projection des coefficients des vitesses de vent

- Ajustement des paramètres du modèle par une loi

Le 4^{ème} modèle a déterminé des coefficients pour chaque vitesse de vent en s'appuyant sur des fréquences empiriques. Ces coefficients ont une progression exponentielle, à l'image de du graphique du premier modèle.

L'objectif de cet ajustement est de :

- lisser les coefficients de la variable « vent » sur les vitesses de vent observées
- projeter ces coefficients sur les vitesses de vent extrêmes

En effet, la recherche d'un lissage des fréquences empiriques est justifiée par le fait que la courbe des fréquences présente de nombreuses irrégularités en fonction de la vitesse de vent. Or, on ne peut supposer que ces variations assez abruptes ne sont pas dues à des variations de l'incidence réelle, mais à une insuffisance de données.

Il existe différentes techniques pour lisser les fluctuations de l'échantillonnage afin d'obtenir une courbe en fonction de la vitesse de vent : modélisation paramétrique, non paramétriques ou relationnels. La technique retenue est la première.

La méthodologie consiste, tout d'abord, à convertir les coefficients en fréquences pour les tempêtes mineures et majeures et ensuite, ces fréquences sont alors ajustées par une loi progressant de façon exponentielle. Elles sont prolongées sur les vitesses de vent plus importantes en ayant comme contrainte d'atteindre une fréquence de 100% pour une vitesse maximum de 350 km/h.

L'énergie d'un vent est proportionnelle au carré de sa vitesse. Par exemple, un vent de 200 km/h exerce une force quatre fois supérieure à celle d'un vent de 100 km/h. Ce propos s'illustre par les coefficients des paramètres des vitesses de vent observés (et des analyses descriptives). Les échelles de Beaufort et de Fujita (cf p 30 et 31) donnent la relation entre les vitesses de vents et les dommages occasionnés. C'est pourquoi, la croissance des fréquences est jugée exponentielle pour atteindre rapidement une très forte proportion d'assurés touchés puis l'atteint de l'exhaustivité des assurés plus progressivement.

Différentes lois ont été testées : Lognormale, Weibull, Makeham, MBBEFD et Bêta.

L'estimation des fréquences de chacune de ces lois a été réalisée par la méthode des moindres carrés. Les estimations des fréquences issues des lois ont été comparées graphiquement avec les fréquences réelles. La loi Bêta ajuste le mieux la réalité.

Loi Bêta de paramètres (a, b), notée β eta(a, b)

La fonction de densité est $f(x) = \frac{1}{\beta(a,b)} x^{a-1} (1-x)^{b-1}$ si $x \in]0,1[$ (0 sinon)

où $\beta(a, b)$ désigne la fonction Bêta définie par :

$$\beta(a, b) = \beta(b, a) = \int_0^1 x^{a-1} (1-x)^{b-1} dx = \frac{\Gamma(a) * \Gamma(b)}{\Gamma(a+b)}$$

$$E[X] = \frac{a}{a+b}$$

$$Var[X] = \frac{a*b}{(a+b+1)(a+b)^2}$$

Propriétés

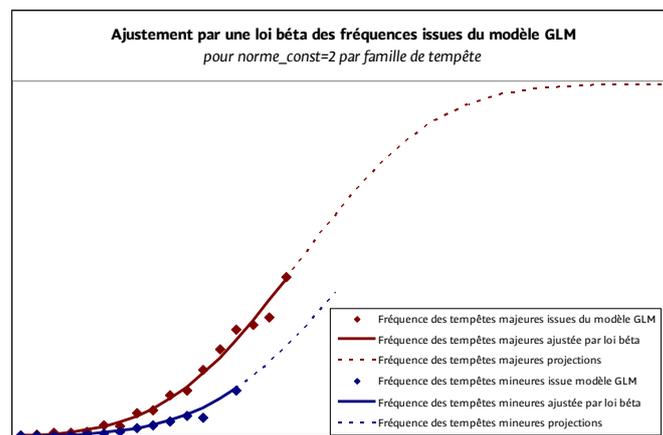
- Soit $F_{a,b}$ la fonction de répartition de la loi β eta(a, b) et $F_{a,b}^{-1}$ la fonction réciproque de $F_{a,b}$ alors, elles vérifient les relations suivantes :

$$\forall x \in]0, 1[, F_{a,b}(x) = 1 - F_{b,a}(1 - x)$$

$$\forall x \in]0, 1[, F_{a,b}^{-1}(x) + F_{a,b}^{-1}(1 - x) = 1.$$

- Si X et Y sont deux variables aléatoires indépendantes de lois gamma respectives :

$$\gamma(a1, b) \text{ et } \gamma(a2, b), \text{ alors } \frac{X}{X+Y} \text{ suit une loi } \beta \text{ eta}(a1, a2).$$



Les points du graphique donnent les fréquences issues du modèle GLM par famille de tempête. Les courbes pleines ajustent ces fréquences par famille selon la loi bêta. Elles sont prolongées en pointillés pour les vitesses de vent pas encore observées.

Dès lors, ces fréquences modélisées sont retranscrites en coefficient pour la variable « vent ». Cette variable est alors introduite comme contrainte dans le modèle linéaire généralisé.

d) Analyse du modèle linéaire généralisé retenu

$$N_i \sim \mathcal{P}[port_i \cdot (famille_tempete.vent)_i \cdot \exp(1.\beta_1 + type_port.\beta_{11,12,13} + norme_const.\beta_{14,15,16,17} + type_ville.\beta_{18,19,20,21})]$$

Ce modèle, intégrant comme contrainte la variable vent, a pour but de réajuster les paramètres des autres variables sur les données sans les points aberrants définis lors du modèle 4.

- Qualité du modèle

Critères d'évaluation de l'adéquation

Critère	DDL	Valeur	Valeur/DDL
Deviance	12E4	162612.1263	1.3732
Pearson Chi-Square	12E4	315318.4103	2.6627

L'ajustement des coefficients de la variable « vent » dégrade peu la qualité du modèle.

- Significativité et détermination de paramètres :

- des variables

Variables	Pr > Khi-2
Type_port	<.0001
norme_const	<.0001
type_ville	<.0001

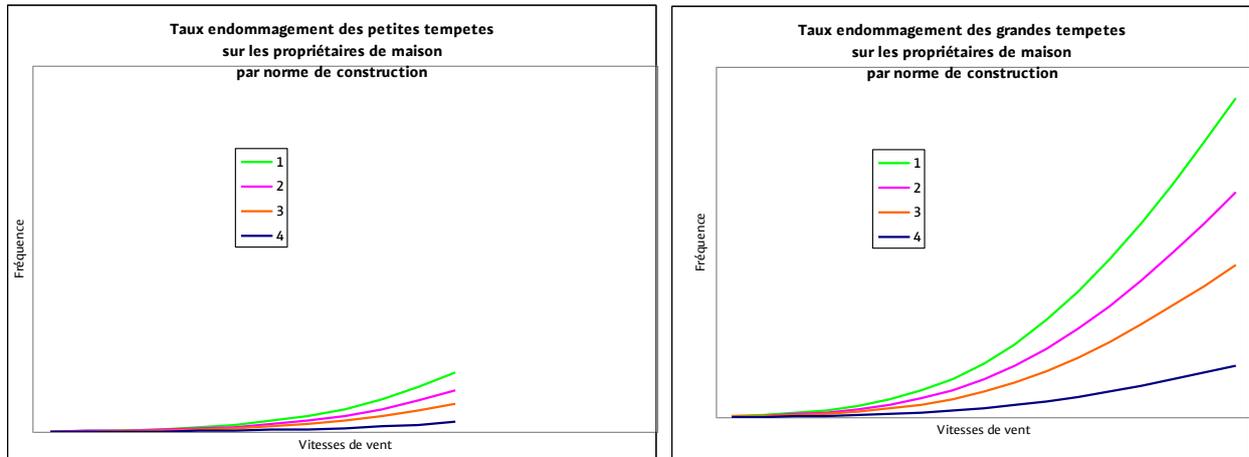
D'après le test de typeIII, les variables restantes sont encore significatives.

- des modalités

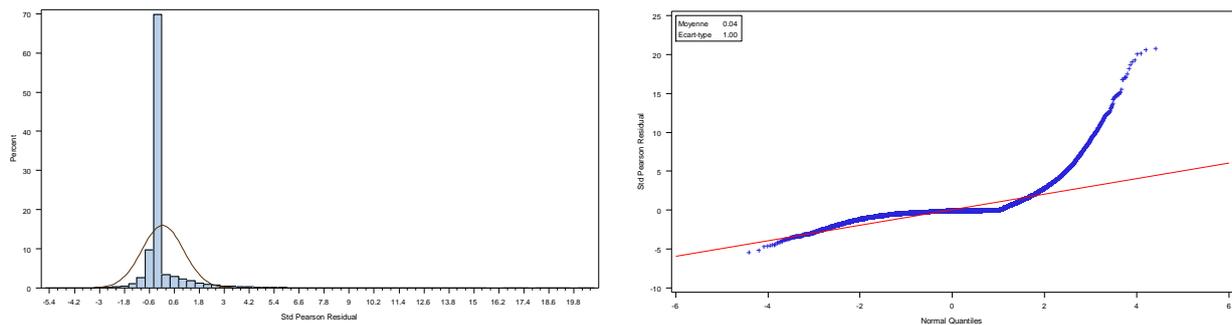
Variables	Modalités	exp(coeff)	Pr > Khi-2
Type_port	locataire	8.6%	<.0001
Type_port	prop_app	18.5%	<.0001
Type_port	prop_mais	100.0%	.
norme_const	1	616.4%	<.0001
norme_const	2	440.1%	<.0001
norme_const	3	298.3%	<.0001
norme_const	4	100.0%	.
type_ville	Commu	105.1%	<.0001
type_ville	GV	64.7%	<.0001
type_ville	MV	87.4%	<.0001
type_ville	PV	100.0%	.

Les paramètres diffèrent peu et les modalités sont toujours significatives.

En synthèse, les courbes de fréquences retenues en fonction de la vitesse de vent pour les propriétaires de maison et selon les zones de normes de construction ont le profil suivant :



○ **ANALYSES DES RESIDUS**



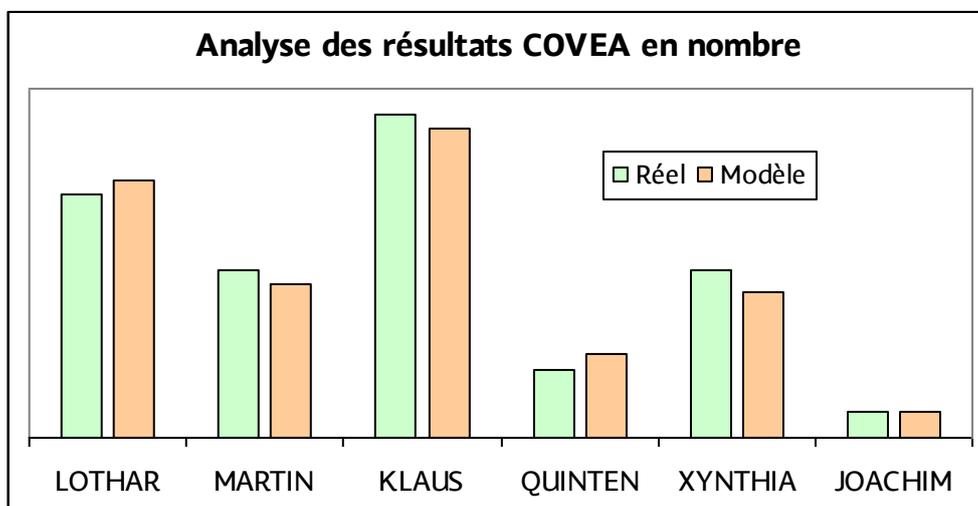
La distribution des résidus de Pearson standardisés n'est pas normale. Lors de la présentation de la partie théorique, cette remarque a été abordée et ne remet pas en cause les hypothèses du modèle linéaire généralisé.

Néanmoins, la distribution des résidus illustre un nombre de très faibles écarts entre les nombres réels et prévisionnel mais également des prévisions bien en deçà de la réalité.

Afin de donner encore plus robustesse au modèle, les points aberrants pourront être analysés. Cette analyse permettra de connaître leur caractéristique : zone géographique à affiner, autres critères discriminants à intégrer ... Cette étude ne sera pas abordée dans ce mémoire.

III.3.1.4 Analyses des résultats

Ce graphique compare les nombres de sinistres réels de chacune des tempêtes avec ceux prévisionnels en vision moyenne issus du modèle final.



Il montre la pertinence du modèle retenu. En effet, les erreurs de prévisions sont contenues.

III.3.2 Modélisations des Coûts Moyens (Y)

La même démarche d'analyse sur les nombres de sinistres a été appliquée sur celle des coûts moyens.

Le modèle de coût moyen des sinistres à un code postal est

- un modèle linéaire généralisé gamma
- multiplicatif avec une fonction de lien logarithmique.

Loi Gamma de paramètres (α, β) , notée $\gamma(\alpha, \beta)$

La fonction de densité est $f(x) = \frac{\beta}{\Gamma(\alpha)} (\beta x)^{\alpha-1} e^{-\beta x}$ si $x > 0$ (0 sinon)

$$E[X] = \frac{\alpha}{\beta} \quad \text{Var}[X] = \frac{\alpha}{\beta^2}$$

La fonction caractéristique est $E[e^{itX}] = \frac{1}{(1 - \frac{it}{\beta})^\alpha}$

Quelques propriétés

$\text{Exp}(\lambda) \equiv \gamma(1, \lambda)$.

Si X suit une loi $\gamma(\alpha, \beta)$, alors $\forall a > 0$, aX suit une loi $\gamma(\alpha, \frac{\beta}{a})$.

Si X_1, \dots, X_n sont des variables aléatoires indépendantes de lois respectives:

$\gamma(\alpha_1, b), \dots, \gamma(\alpha_n, b)$, alors $\sum_{i=1}^n X_i$ suit une loi $\gamma(\sum_{i=1}^n \alpha_i, b)$.

Soit Y_i le coût moyen pour un code postal i est défini comme le coût total des sinistres déclarés pondéré par le nombre de sinistres déclarés au sein du même code postal i .

Le coût moyen de sinistre est une variable aléatoire continue. C'est pourquoi, la loi retenue est la loi gamma qui est généralement utilisée pour cet indicateur.

On suppose alors que le modèle Gamma des coûts moyens a pour moyenne :

$$E(Y_i / X) = \exp(\beta^t X_i)$$

Le modèle des coûts moyens s'écrit de la manière suivante :

$$Y_i \sim \mathcal{G}(\exp(\beta^t x_i))$$

III.3.2.1 Significativité des critères

a) Discrimination et interprétation

Ce premier modèle consiste à intégrer l'ensemble des variables explicatives à priori ainsi que la totalité de leurs modalités afin d'en connaître la significativité de chacune d'elle.

Ce modèle s'écrit de la manière suivante :

$$Y_i \sim \gamma [\exp(1.\beta_1 + \text{societe}.\beta_{2,3,4} + \text{tempete}.\beta_{5,\dots,10} + \text{type_port}.\beta_{11,12,13} + \text{norme_const}.\beta_{14,15,16,17} + \text{type_ville}.\beta_{18,19,20,21} + \text{vent} \beta_{22,\dots})]$$

On retrouve ici les mêmes modalités que les nombres et décrites p. 42 et 43.

- Qualité du modèle

Critères d'évaluation de l'adéquation

Critère	DDL	Valeur	Valeur/DDL
Deviance	28E3	55541.8055	1.9933
Pearson Chi-Square	28E3	96080.1220	3.4482

Ces indicateurs montrent une qualité du modèle moins bonne que celle des nombres de sinistres.

- Significativité :

- des variables

Statistique LR pour Analyse de Type 3

Source	DDL Num.	DDL Res.	Valeur F	Pr > F	Khi-2	Pr > Khi-2
societe	2	27864	458.38	<.0001	916.76	<.0001
tempete	5	27864	56.80	<.0001	283.98	<.0001
Type_port	2	27864	771.86	<.0001	1543.71	<.0001
Norm_const	3	27864	34.25	<.0001	102.76	<.0001
Type_ville	3	27864	92.30	<.0001	276.90	<.0001
Vent	17	27864	39.07	<.0001	664.21	<.0001

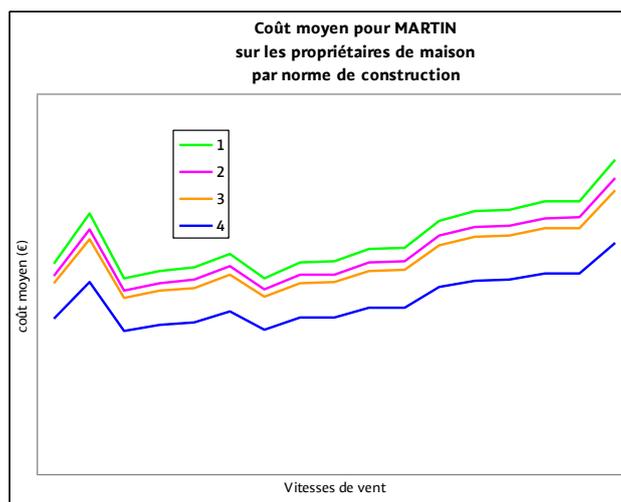
Ce test permet de montrer que l'ensemble des variables est significatif pour l'estimation des coûts moyens de sinistres.

- des modalités

Variables	Modalités	exp(coeff)	Pr > Khi-2
societe	societe 2	71.0%	<.0001
societe	societe 3	70.2%	<.0001
societe	societe 1	100.0%	.
tempete	JOACHIM	87.5%	<.0001
tempete	KLAUS	106.9%	<.0001
tempete	LOTHAR	102.4%	0.1329
tempete	MARTIN	89.0%	<.0001
tempete	QUINTEN	86.6%	<.0001
tempete	XYNTHIA	100.0%	.
Type_port	locataire	44.0%	<.0001
Type_port	prop_app	78.8%	<.0001
Type_port	prop_mais	100.0%	.
norme_const	1	136.0%	<.0001
norme_const	2	127.9%	<.0001
norme_const	3	122.7%	<.0001
norme_const	4	100.0%	.
type_ville	Commu	105.1%	<.0001
type_ville	GV	87.2%	<.0001
type_ville	MV	92.3%	<.0001
type_ville	PV	100.0%	.

Les paramètres, présentés ci-contre, ont été déterminés par maximum de vraisemblance et convertis en exponentielle (fonction de lien logarithmique), afin de comparer les modalités d'une variable avec une modalité de référence en base 100. La majorité des modalités est également significative.

Les coefficients des paramètres des vitesses de vent sont illustrés sur le graphique ci-dessous. Chacun des paramètres est significatif.



- Interprétations

Le tableau ci-dessus donne les coefficients issus du modèle linéaire généralisé et permet de montrer l'importance des modalités au sein d'une variable explicative.

Ces coefficients traduisent bien les analyses descriptives que nous avons illustrées précédemment :

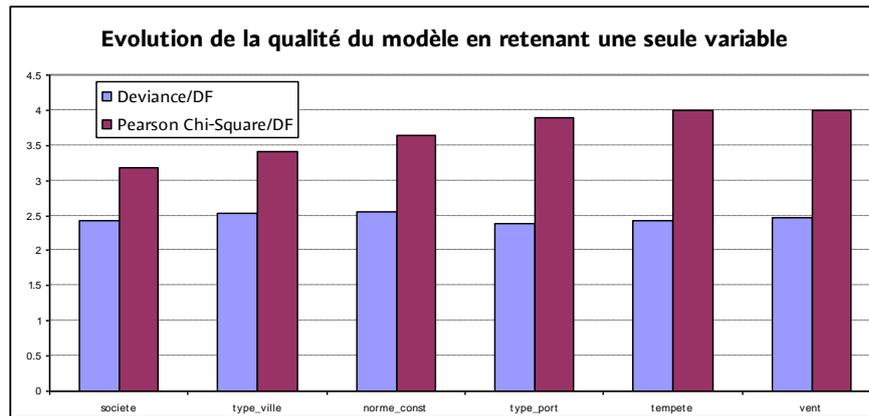
- Société : Les coefficients sont similaires entre les sociétés 2 et 3 et en deçà de 30% des coefficients de la société 1, ce qui se justifie par les souscriptions MRH plus importants (capitaux assurés plus élevés).
- Tempête : On retrouve le même ordre d'importance entre les tempêtes : Lothar, Klaus et Xynthia puis les autres.
- Type portefeuille : Les propriétaires ont des coûts plus importants que les locataires, et plus particulièrement ceux qui possèdent une maison.
- Norme de construction : L'exigence des normes de construction influe sur le coût. Comme les fréquences, la décroissance des coefficients est en lien avec l'exigence des normes.
- Type de ville : L'urbanisation des villes joue également sur le coût. En effet, la structure des bâtiments est différente.
- Vitesse de vent : Plus la vitesse de vent est élevée plus les coûts sont importants au-delà d'une certaine vitesse et stable en deçà.

b) Influence des variables explicatives

Reprenons la même méthodologie que pour les nombres de sinistres.

○ Méthode forward

L'introduction de la première variable montre que la variable la plus influente est la société. Néanmoins, le graphique ci-dessous montre que la variable societe ne ressort pas de manière significative.



L'introduction des autres variables permet de minimiser les indicateurs de qualité mais de manière plus marginale.

	societe	+ type_port	+ vent	+ tempete	+ norme_const	+ type_ville
Deviance/DF	2.47	2.24	2.08	2.04	2.01	1.99
Pearson Chi-Square/DF	4.01	3.39	3.67	3.72	3.47	3.45

○ Analyse du test de wald (type III)

Source	Statistique LR pour Analyse de Type 3					
	DDL Num.	DDL Res.	Valeur F	Pr > F	Khi-2	Pr > Khi-2
societe	2	27864	458.38	<.0001	916.76	<.0001
tempete	5	27864	56.80	<.0001	283.98	<.0001
Type_port	2	27864	771.86	<.0001	1543.71	<.0001
Norme_const	3	27864	34.25	<.0001	102.76	<.0001
Type_ville	3	27864	92.30	<.0001	276.90	<.0001
vent	17	27864	39.07	<.0001	664.21	<.0001

Cette sortie confirme globalement le même ordre de significativité des variables dans le modèle des coûts moyens.

III.3.2.2 Présentation du modèle final

a) Présentation des étapes

- 2ème modèle : regroupement des sociétés et des tempêtes.

La variable « societe » est conservée pour distinguer les sociétés.

Le regroupement des sociétés 2 et 3 estime un coût moyen 32% en deçà de celle de la société 1.

Pour les mêmes raisons que les nombres, la variable tempête ne peut être retenue en soi. Mais il est difficile de regrouper les tempêtes selon un critère. Elles ont été regroupées selon les mêmes catégories que les nombres de sinistres afin d'avoir la même continuité : tempêtes majeures et mineures.

Le regroupement des tempêtes semble également opportun avec les tempêtes majeures qui ont un coût moyen des indemnisations supérieur à celui des tempêtes mineures de plus de 4%.

- 3ème modèle : regroupement des vitesses de vent

On note que les coefficients issus du modèle sur les vents faibles sont très proches et progressent ensuite. Face à cette stabilité (coefficients proches), l'objectif est de rechercher jusqu'à quel seuil de vitesse de vent, il est nécessaire de les regrouper en une seule modalité. Pour ce faire, un modèle linéaire généralisé a été réalisé par zone de norme de construction.

L'analyse de ce modèle a permis d'établir la vitesse de vent seuil. Elle est jugée identique quelle que soit la norme de construction.

Les vitesses de vent faible ont donc été regroupées dans ce 3^{ème} modèle linéaire généralisé.

- 4ème modèle : interactions des tempêtes et vitesse de vent

Ce modèle consiste à déconnecter le lien des tempêtes mineures et majeures en fonction des vitesses de vent, comme pour les fréquences.

- 5ème modèle : suppression de la famille de tempêtes

Au regard des paramètres de l'interaction, seuls les paramètres au-delà de la vitesse seuil diffèrent entre les familles des tempêtes. Or, les vitesses de vent au-delà de la vitesse seuil ne sont pas nombreuses pour les tempêtes mineures. C'est pourquoi, la famille de tempêtes est finalement écartée du modèle.

- 6ème modèle : retrait des observations aberrantes

Du modèle précédent, les observations qui ont des résidus de Pearson standardisés extrêmes (0.1%) sont écartées de ce nouveau modèle.

Ce retrait permet de rendre le modèle plus pérenne dans le temps et donc de ne pas être influencé par les valeurs extrêmes.

b) Qualité des différents modèles :

	1er modèle	2ème modèle	3ème modèle	4ème modèle	5ème modèle	6ème modèle
Déviante / ddl	1.9933	2.0306	2.0378	2.0330	2.0394	1.9577
Pearson Chi-Square / ddl	3.4482	3.4604	3.5415	3.5066	3.5128	2.8539

Les regroupements des sociétés et des vitesses de vent influent un peu sur la qualité du modèle. La suppression des tempêtes dans le modèle n'a que peu d'effet sur la déviance et Pearson.

Par contre, le retrait des observations a permis de faire progresser la qualité du modèle.

c) Etape complémentaire : lissage et projection des coefficients des vitesses de vent

- Ajustement des paramètres du modèle par une loi

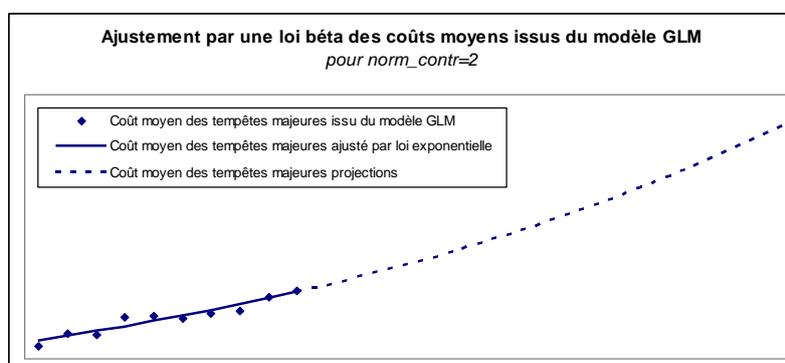
Tout comme les fréquences, on note que les coûts moyens de sinistres progressaient avec l'intensité du vent.

L'objectif de ce nouveau modèle est identique à celui des nombres :

- Lisser les coefficients de la variable « vent » sur les vitesses de vent observées
- Projeter ces coefficients sur les vitesses de vent extrêmes

La progression des coûts moyens est jugée exponentielle au-delà du seuil. Elle se justifie par le fait que plus l'intensité du vent, plus l'énergie déployée est grande, plus les dommages impactent la structure du bâtiment, eu égard les conséquences du passage des tornades (échelle de Beaufort et de Fujita). C'est pourquoi, la méthodologie consiste, tout d'abord, à convertir les coefficients en coûts moyens et ensuite, ces coûts moyens sont alors ajustés par une tendance exponentielle. Cette tendance est alors prolongée sur les vitesses de vent plus importantes sans contrainte particulière.

L'ajustement sur les coefficients au-delà du seuil est réalisé en lissage linéaire sur les logarithmes des coûts moyens proposés par le modèle GLM. Cet ajustement est ensuite passé en exponentiel.



Dès lors que les coûts moyens modélisés sont établis, ils sont convertis en coefficient pour la variable « vent ». Celle-ci est introduite comme contrainte du modèle.

d) Analyse du modèle linéaire généralisé retenu

Ce modèle s'écrit de la manière suivante :

$$Y_i \sim \gamma [(gr_vent)_i \cdot \exp(1.\beta_1 + gr_societe.\beta_{3,4} + type_port.\beta_{11,12,13} + norme_const.\beta_{14,15,16,17} + type_ville.\beta_{18,19,20,21})]$$

- Qualité du modèle :

Critères d'évaluation de l'adéquation

Critère	DDL	Valeur	Valeur/DDL
Deviance	28E3	54738.6725	1.9648
Pearson Chi-Square	28E3	79877.9458	2.8672

Comme pour les nombres, l'ajustement des coefficients de la variable « vent » influe peu sur la qualité du modèle.

- Qualité des variables :

- des variables

Variables	Pr > Khi-2
gr_societe	<.0001
Type_port	<.0001
norme_const	<.0001
type_ville	<.0001

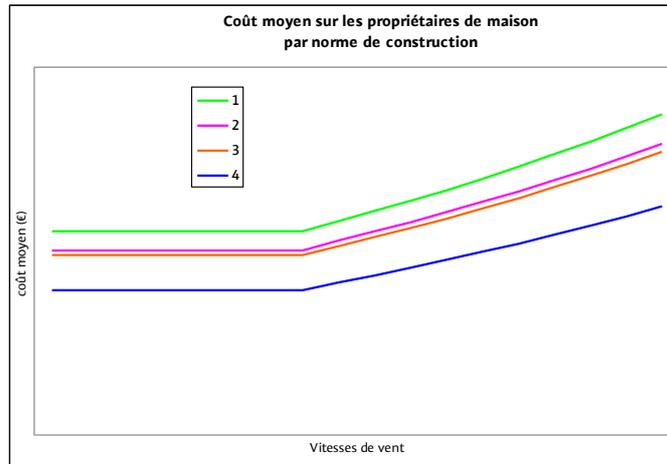
Ce test permet de montrer que l'ensemble des variables est significatif dans l'estimation des coûts moyens de sinistres.

- des modalités

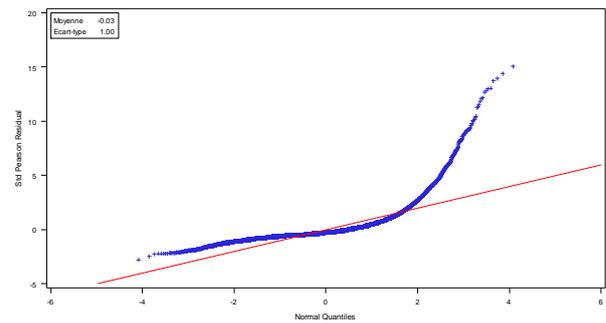
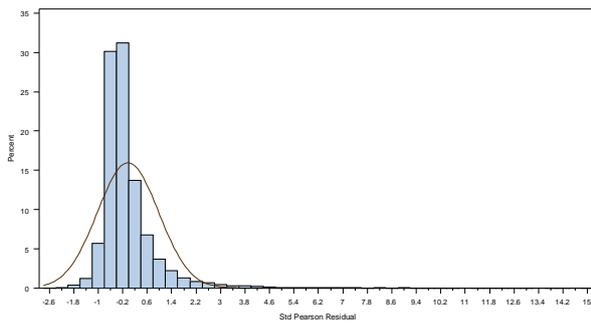
Variables	Modalités	exp(coeff)	Pr > Khi-2
gr_societe	societe 23	68.0%	<.0001
gr_societe	societe 1	100.0%	<.0001
Type_port	locataire	43.5%	.
Type_port	prop_app	78.2%	<.0001
Type_port	prop_mais	100.0%	<.0001
Type_ville	Commu	105.5%	<.0001
Type_ville	GV	88.4%	<.0001
Type_ville	MV	93.3%	<.0001
Type_ville	PV	100.0%	.
norme_const	1	140.4%	<.0001
norme_const	2	127.4%	<.0001
norme_const	3	124.0%	<.0001
norme_const	4	100.0%	.

Les valeurs des paramètres ont très peu varié et les modalités sont significatives.

En synthèse, les courbes de coût moyen retenues en fonction de la vitesse de vent pour les propriétaires de maison et selon les zones de normes de construction ont le profil suivant :

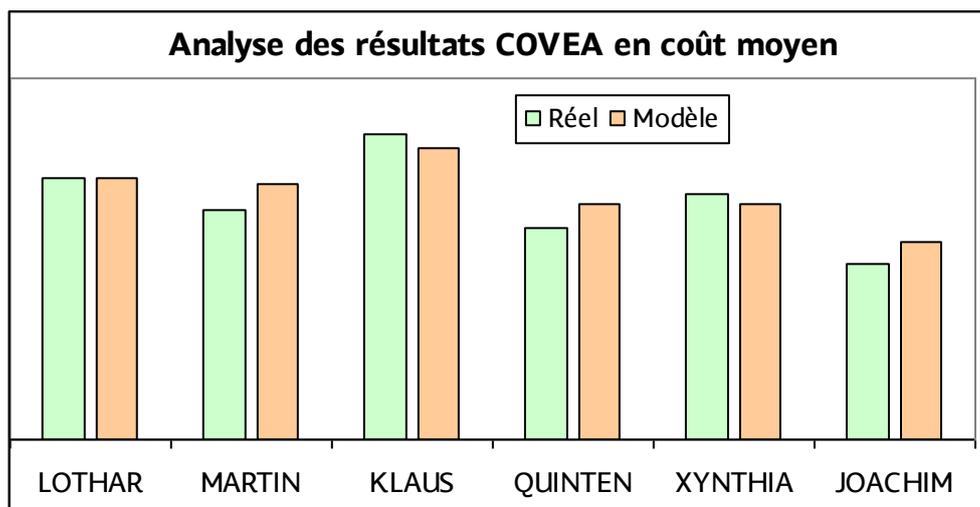


○ **ANALYSES DES RESIDUS**



La distribution des résidus de Pearson standardisés ressemble à celle des nombres de sinistres. Ainsi, les écarts se caractérisent par des coûts moyens réels très élevés.

III.3.2.3 Analyses des résultats

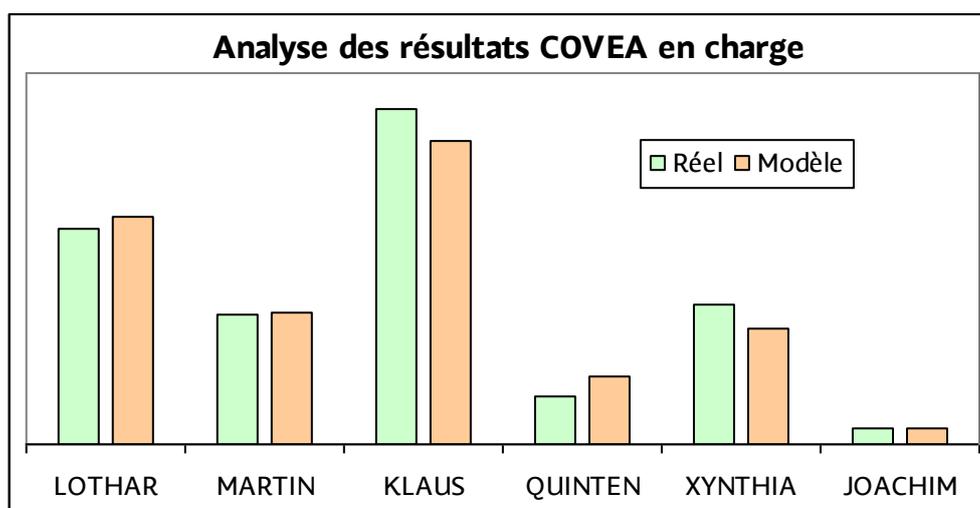


Ce graphique montre les résultats par tempête et nous permet de montrer la pertinence des résultats prévisionnels par tempête.

III.3.3 Résultats des modélisations en terme de charge

Lors de la page 44, l'espérance de la charge a été définie par $E(S/X) = E(N/X) \cdot E(Y/X)$ lorsque l'on introduit p variables explicatives $X = (X_1, \dots, X_p)$.

Les espérances des nombres de sinistres et des coûts moyens ont été présentées précédemment et permettent donc de déduire l'espérance de la charge par tempête.



Ce graphique donne la résultante en terme de charge et illustre de bonnes adéquations.

IV. Synthèse

L'objectif de ce chapitre était de présenter une solution pour répondre au premier aléa concernant l'estimation de la sinistralité lors de la survenance d'une tempête.

Pour cela, l'analyse de la sinistralité de 6 tempêtes historiques nous a permis de visualiser les critères qui pouvaient être discriminants à priori. En effet, les fréquences et les coûts moyens, indicateurs matérialisant la charge de sinistre, sont dépendants de l'intensité du vent, de la structure de portefeuille (qualité juridique, type d'habitat, géolocalisation), de la zone de construction où l'habitat se situe ...

La méthode utilisée pour modéliser ces indicateurs de sinistralité en fonction de ces critères est le modèle linéaire généralisé. Dans un premier temps, des hypothèses ont été posées pour chaque modèle : de loi, indépendance entre variables ... Dans un second temps, les variables discriminantes ont été sélectionnées en étudiant leur pertinence ainsi que celle de leur modalité.

Les paramètres sur les vitesses de vent correspondent à une moyenne des observations et présentent des irrégularités de fréquence en fonction de l'intensité du vent. C'est pourquoi, ces paramètres ont été ajustés selon une loi, permettant de projeter pour des vitesses de vent pas encore enregistrées ou pas suffisamment robustes.

Les sélections et les regroupements de modalités ont été réalisés dans l'optique d'une application d'une prochaine tempête, ceci en jugeant de la qualité des modèles.

Ainsi, lors de l'arrivée d'une prochaine tempête, il est nécessaire de prévoir les vitesses de vent à la commune (module aléa) et de connaître les caractéristiques de notre portefeuille du moment (module enjeu). En appliquant les modèles de fréquences et de coûts moyens (module vulnérabilité), un nombre de sinistre est estimé ainsi qu'une charge. Ces estimations répondent aux besoins de gestion sinistre et des arrêtés des comptes.

Chapitre 3. INCERTITUDE AUTOUR DE LA MOYENNE

I. Objectif

Le modèle linéaire généralisé a permis d'estimer des indicateurs de sinistralité (fréquence, coût moyen et charge) satisfaisant en vision moyenne pour chacune des tempêtes. Mais les critères discriminants choisis ne permettent pas de prévoir avec exactitude les valeurs réelles des tempêtes historiques. Les écarts entre les valeurs réelles et estimées se justifient par : une modélisation qui ne se fait pas au code postal mais en vision agrégée, des critères plus ou moins fiables, des informations discriminantes manquantes (durées et trajectoires des rafales, interpolation des vitesses de vent, zonages des normes plus fins...)... Ces écarts sont matérialisés par les résidus.

L'objectif de ce chapitre est de déterminer l'incertitude autour de la moyenne sur les indicateurs de sinistralité. On parle ici de l'incertitude autour de la moyenne qui n'est pas à confondre avec un intervalle de confiance sur les estimations. Pour ce faire, il serait nécessaire de donner une distribution à chaque code postal et de trouver une corrélation entre eux pour les agréger au niveau national. L'analyse des résidus entre les codes postaux les plus proches permettrait d'établir cette corrélation. Or, la notion de distance entre les codes postaux n'était pas disponible.

Afin de mesurer l'incertitude des estimations autour de la moyenne, deux méthodes seront analysées : celle sur les écarts à la prédiction (MSEP) et celle sur les bootstrap. Ce sont des méthodes très utilisées pour calculer les incertitudes dans le cadre du provisionnement. La méthode des MSEP utilise des produits de matrices. Pour le provisionnement, les dimensions des matrices correspondent aux nombres de survenances étudiées et donc très limitées. Or, pour ce mémoire, le nombre d'observations implique des dimensions de matrice trop conséquentes et le calcul matriciel n'a pas abouti à cause de capacités informatiques limitées. Néanmoins, les résultats de la méthode bootstrap sont équivalents à ceux de la méthode MSEP. Les résultats sur l'incertitude seront donc présentés pour le bootstrap.

II. Erreur quadratique moyenne de prédiction

Avant d'exploiter la méthode des MSEP sur les différents indicateurs de sinistralité, nous allons définir cette méthode.

II.1. Définition

L'erreur quadratique moyenne de prédiction, « mean square error of prediction » en anglais est appelé MSEP de part ces sigles anglais. Par définition, le MSEP est la moyenne des différences au carré entre la variable aléatoire et son prédicteur, est égale à :

$$\text{MSEP}(\hat{Z}) = E[(Z - \hat{Z})^2]$$

Où Z est la variable aléatoire de prédiction et \hat{Z} est la valeur prédite par le modèle.

Cette formule peut également s'écrire sous la forme :

$$\text{MSEP}(\hat{Z}) = s^2(Z) + s^2(\hat{Z})$$

$$\begin{aligned} \text{En effet } E[(Z - \hat{Z})^2] &= E[(Z - E(Z) + E(Z) - \hat{Z})^2] \\ &= E[(Z - E(Z))^2] + E[(\hat{Z} - E(Z))^2] - 2 E[(Z - E(Z)) \cdot (\hat{Z} - E(Z))] \end{aligned}$$

En remplaçant $\hat{Z} - E(Z)$ par $\hat{Z} - E(\hat{Z})$, on obtient :

$$E[(Z - \hat{Z})^2] = E[(Z - E(Z))^2] + E[(\hat{Z} - E(\hat{Z}))^2] - 2 E[(Z - E(Z)) \cdot (\hat{Z} - E(\hat{Z}))]$$

En supposant l'indépendance entre les erreurs des valeurs $(Z - E(Z))$ et $(\hat{Z} - E(\hat{Z}))$, la formule devient :

$$E[(Z - \hat{Z})^2] \approx V(Z) + V(\hat{Z})$$

La valeur du MSEP se déduit donc des valeurs de la volatilité de la variable aléatoire et des valeurs de volatilité de la prédiction.

Les paragraphes suivants vont définir la volatilité des indicateurs de sinistralité.

II.2. Volatilité des nombres et des coûts moyens de sinistres

II.2.1 Volatilité des variables aléatoires

Lors du chapitre précédent, nous avons décomposé la variable aléatoire de la charge en deux variables aléatoires sur les nombres de sinistres et sur les coûts moyens. Ces derniers ont été estimés par des modèles linéaires généralisés et donc les valeurs prédites sont fonction des variables explicatives X et des paramètres $\hat{\beta}$.

Cette volatilité va donc être décrite dans les paragraphes suivants pour les nombres de sinistres puis pour les coûts moyens.

II.2.1.1 Nombre de sinistres N

Pour rappel, le modèle multiplicatif sur les nombres de sinistres est poissonnien (page 86) tel que :

$$N_i \sim \mathcal{P}(\text{port}_i \cdot \exp(\beta'x_i))$$

Les paramètres pour obtenir la fonction de densité de la loi de poisson à partir de la distribution de la famille exponentielle sont :

$$a(\phi) = 1 \quad b(\theta) = e^\theta = \lambda \quad c(y, \phi) = -\ln(y!)$$

$$\theta = \ln \lambda$$

La distribution de la famille exponentielle s'écrit

$$f(y_i; \phi, \theta_i) = \exp \left[\frac{y_i \theta_i - b(\theta_i)}{a(\phi)} + c(y_i, \phi) \right]$$

Vérification :

Ainsi, $f(y_i; \phi, \theta_i) = \exp(y_i \ln \lambda - \lambda - \ln(y!)) = \lambda^{y_i} \cdot \exp(-\lambda) / \exp(\ln(y!))$

On obtient alors la loi de densité de poisson

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

A partir de la formule définie p 56, voici les moments qui découlent de la variable aléatoire des nombres de sinistres :

$$E(N) = b'(\theta) = e^\theta = \lambda$$

$$V(N) = b''(\theta) * a(\theta) = e^\theta = \lambda$$

Les moments de la loi de poisson sont bien égaux.

Il est possible d'accepter une dispersion différente de l'espérance en définissant :

$$a(\phi) = \phi \neq 1$$

Si $a(\phi) > 1$ alors la variance est supérieure à l'espérance et on parle de Poisson sur-dispersé. Dans le cas inverse, la loi est appelée Poisson sous-dispersé.

Par voie de conséquence, les moments de la variable aléatoire des nombres de sinistres sachant les variables explicatives deviennent :

$$E(N / X) = b'(\theta) = e^\theta = \lambda$$

$$V(N / X) = b''(\theta) * a(\theta) = e^\theta * \phi = \lambda \phi = \phi * E(N / X)$$

La volatilité de la variable aléatoire des nombres de sinistres N s'écrit alors :

$$s^2(N / \mathbf{X}) = \phi \mathbf{E}(N / \mathbf{X})$$

II.2.1.2 Le coût moyen Y

Quant aux coûts moyens, le modèle est multiplicatif gamma (page 73) tel que :

$$Y_i \sim \mathcal{G}(\exp(\beta'x_i))$$

Pour la loi gamma $Z \sim \gamma(\alpha, \beta)$, les paramètres sont alors

$$a(\phi) = \phi \quad b(\theta) = -\ln(-\theta) \quad c(y, \phi) = \ln\left(\frac{y}{\phi}\right) / \phi - \ln y - \ln \Gamma(1 / \phi)$$

$$\phi = \frac{1}{\alpha} \quad \theta = -\frac{\beta}{\alpha}$$

Vérification :

$$\begin{aligned} \text{Ainsi, } f(y_i; \phi, \theta_i) &= \exp(-\alpha y_i \frac{\beta}{\alpha}) \cdot \exp(\alpha \ln(\frac{\beta}{\alpha})) \cdot \exp(\alpha \ln(\alpha y_i) - \ln y_i - \ln \Gamma(\alpha)) \\ &= [\exp(-y_i \beta) \cdot (\frac{\beta}{\alpha})^\alpha \cdot (\alpha y_i)^\alpha] / (y_i * \Gamma(\alpha)) \\ &= [\exp(-y_i \beta) \cdot (\frac{\beta}{\alpha})^\alpha \cdot (\alpha y_i)^\alpha] / (y_i * \Gamma(\alpha)) \end{aligned}$$

On obtient alors la loi de densité gamma

$$f(y) = \frac{\beta}{\Gamma(\alpha)} (\beta y)^{\alpha-1} e^{-\beta y} \text{ si } x > 0 \text{ (0 sinon)}$$

Les moments de la variable aléatoire des coûts moyens sachant les variables explicatives deviennent :

$$\begin{aligned} E(Y/X) &= b'(\theta) = (-\ln(-\theta))' = -\frac{1}{\theta} = \frac{\alpha}{\beta} \\ V(Y/X) &= b''(\theta) \cdot a(\theta) = (-\frac{1}{\theta})' \cdot \phi = \frac{\alpha^2}{\beta^2} \cdot \frac{1}{\alpha} = \frac{\alpha}{\beta^2} = \phi \cdot E^2(Y/X) \end{aligned}$$

La volatilité de la variable aléatoire des coûts moyens de sinistres Y s'écrit alors :

$$s^2(Y/X) = \phi E^2(Y/X)$$

II.2.2 Volatilité des prédictions

Dorénavant, intéressons-nous à la volatilité des valeurs prédites. Dans le cas générique des modèles linéaires généralisés avec une fonction de lien est logarithmique, la formule s'écrit :

$$\text{Ln}(\hat{Z}) = \beta'X = \hat{\mu}$$

$$\hat{Z} = \exp(\beta'X) = \exp(\hat{\mu})$$

Se pose la question de la variance de \hat{Z} , autrement dit $V(\exp(\hat{\mu}))$.

D'après la formule de Taylor, $f(x) = (x - x_0) \cdot f'(x_0) + \varepsilon$

Alors $V(f(x)) = V [x \cdot f'(x_0) - x_0 \cdot f'(x_0) + \varepsilon]$

et $V(f(x)) = f'(x_0)^2 \cdot V(x) + f'(x_0) \cdot \text{cov}(x, \varepsilon) + V(\varepsilon)$

et peut se simplifier par $V(f(x)) \approx f'(x_0)^2 \cdot V(x)$

Par voie de conséquence,

$$V(\hat{Z}_{ij}) = \left(\frac{\partial \hat{Z}_{ij}}{\partial \eta_{ij}} \right)^2 V(\hat{\mu}_{ij})$$

Puisque $\hat{Z} = \exp(\hat{\mu}_{ij})$, on obtient alors :

$$s^2(\hat{Z}_{ij}) = \hat{Z}_{ij}^2 s^2(\hat{\mu}_{ij})$$

Par conséquent, les volatilités des prédictions des nombres de sinistres (N) et des coûts moyens de sinistres (Y) s'écrivent alors :

$$s^2(\hat{N}_{ij} / \mathbf{X}) = \hat{N}_{ij}^2 s^2(\hat{\mu}_{ij})$$

$$s^2(\hat{Y}_{ij} / \mathbf{X}) = \hat{Y}_{ij}^2 s^2(\hat{\mu}_{ij})$$

II.3. Volatilité des charges de sinistres

La variable aléatoire charge se décompose comme le produit de la variable aléatoire du nombre de sinistres et de la variable aléatoire du coût moyen.

L'erreur quadratique moyenne de prédiction de la charge s'écrit :

$$\mathbf{MSEP}(\hat{S}) = s^2(\mathbf{S}) + s^2(\hat{S})$$

II.3.1 Volatilité de la variable aléatoire

L'espérance de la charge s'écrit :

$$E\left(\sum_{i=1}^N Y_i\right) = E\left(E\left(\sum_{i=1}^N Y_i / N\right)\right) = E\left(N E(Y / N)\right) = E\left(N E(Y)\right)$$

$$E(S) = E(N) E(Y)$$

La variance de la charge s'écrit :

$$V\left(\sum_{i=1}^N Y_i\right) = E\left(\sum_{i=1}^N Y_i^2\right) - E^2\left(\sum_{i=1}^N Y_i\right)$$

Or

$$E\left(\sum_{i=1}^N Y_i^2\right) = E\left(E\left(\sum_{i=1}^N Y_i^2 / N\right)\right) = E\left(E\left(V\left(\sum_{i=1}^N Y_i\right) + E^2\left(\sum_{i=1}^N Y_i\right) / N\right)\right)$$

$$= E\left(N E(V(Y) + N^2 E^2(Y))\right)$$

$$E(S^2) = E(N) \cdot V(Y) + E(N^2) \cdot E^2(Y)$$

Ainsi

$$V\left(\sum_{i=1}^N Y_i\right) = E(N) \cdot V(Y) + E(N^2) \cdot E^2(Y) - E^2(N) E^2(Y)$$

$$= E(N) \cdot V(Y) + V(N) \cdot E^2(Y)$$

La variance de la variable aléatoire de la charge sachant les variables explicatives s'écrit :

$$V(S/X) = E(N/X) \cdot V(Y/X) + V(N/X) \cdot E^2(Y/X)$$

Des volatilités des variables aléatoires des fréquences et des coûts moyens précisées précédemment, on en détermine :

$$s^2(S) = E(N/X) \cdot E^2(Y/X) \cdot \phi_{CM} + \phi_{NB} E(N/X) \cdot E^2(Y/X)$$

$$s^2(S) = E(N/X) \cdot E^2(Y/X) (\phi_{CM} + \phi_{NB})$$

$$\boxed{s^2(S) = \hat{S} \cdot \hat{Y} \cdot (\phi_{CM} + \phi_{NB})}$$

Tel que $\hat{Y} = E(Y/X)$ $\hat{S} = E(S/X) = E(N/X) \cdot E(Y/X)$

II.3.2 Volatilité des prédictions

$$\begin{aligned} V(\hat{S}) &= E(\hat{S}^2) - E^2(\hat{S}) \\ &= E(\hat{N}^2 \cdot \hat{Y}^2) - E^2(\hat{N} \cdot \hat{Y}) \end{aligned}$$

Tel que $\hat{N} = E(N/X)$

Par indépendance des variables aléatoires des nombres de sinistres (N) et des coûts moyens (Y), la variance devient :

$$\begin{aligned} V(\hat{S}) &= E(\hat{N}^2) \cdot E(\hat{Y}^2) - \hat{N}^2 \cdot \hat{Y}^2 \\ &= [E^2(\hat{N}) + V(\hat{N})] \cdot [E^2(\hat{Y}) + V(\hat{Y})] - \hat{N}^2 \cdot \hat{Y}^2 \end{aligned}$$

Des volatilités des prédicteurs des fréquences et des coûts moyens précisées précédemment, on en détermine :

$$\begin{aligned} V(\hat{S}) &= (\hat{N}^2 + \hat{N}^2 s^2(\hat{\mu}_{NB})) \cdot (\hat{Y}^2 + \hat{Y}^2 s^2(\hat{\mu}_{CM})) - \hat{N}^2 \cdot \hat{Y}^2 \\ &= \hat{N}^2 \cdot \hat{Y}^2 [(1 + s^2(\hat{\mu}_{NB})) \cdot (1 + s^2(\hat{\mu}_{CM})) - 1] \end{aligned}$$

$$\boxed{s^2(\hat{S}) = \hat{S}^2 [(s^2(\hat{\mu}_{NB}) + s^2(\hat{\mu}_{CM}) + s^2(\hat{\mu}_{NB}) \cdot s^2(\hat{\mu}_{CM}))]}$$

II.4. MSEP des indicateurs de sinistralité et sécurisations

En somme, les MSEP s'écrivent pour les :

- nombres de sinistres :

$$\text{MSEP}(N_{ij}/X) = \phi_{NB} E(N_{ij}/X) + \hat{N}_{ij}^2 s^2(\hat{\mu}_{ij})$$

$$\text{MSEP}(N_{ij}/X) = \hat{N}_{ij} (\phi_{NB} + \hat{N}_{ij} s^2(\hat{\mu}_{ij}))$$

- coûts moyens de sinistres :

$$\text{MSEP}(Y_{ij}/X) = E^2(Y_{ij}/X) \cdot \phi_{CM} + \hat{Y}_{ij}^2 s^2(\hat{\mu}_{ij})$$

$$\text{MSEP}(Y_{ij}/X) = \hat{Y}_{ij}^2 (\phi_{CM} + s^2(\hat{\mu}_{ij}))$$

- charges de sinistres

$$\text{MSEP (S/X)} = \hat{S} \cdot \hat{Y} \cdot (\phi_{CM} + \phi_{NB}) + \hat{S}^2 [(\mathbf{s}^2(\hat{\mu}_{NB}) + \mathbf{s}^2(\hat{\mu}_{CM}) + \mathbf{s}^2(\hat{\mu}_{NB}) \cdot \mathbf{s}^2(\hat{\mu}_{CM}))]$$

La valeur centrale des estimations des indicateurs de sinistralité peut alors être sécurisée selon un niveau de risque α % :

$$Z = E(Z/X) + U^{-1}(1-\alpha) \text{MSEP (Z/X)}$$

Où $U^{-1}(1-\alpha)$ est la fonction quantile de la loi Normale centrée réduite au risque α et Z est la variable aléatoire.

II.5. Limite de la méthode MSEP

Le calcul matriciel a été réalisé sous le logiciel SAS. Il demande une capacité informatique trop conséquente. Le message donné par SAS est « Unable to allocate sufficient memory ». Cette méthode n'a pu aboutir à de résultats chiffrés.

III. Bootstrap

III.1. Méthode

La méthode du bootstrap, proposée par Bradley Efron (1979), permet de connaître les inférences. Le bootstrap est une méthode de monte carlo basée sur les données observées. Cette technique permet de donner donc une espérance, une médiane et dans le cas qui nous intéresse la volatilité (variance) autour de cette moyenne.

Il existe la méthode de bootstrap paramétrique et celle non paramétrique. La première est basée sur la connaissance de la distribution à priori et des paramètres du modèle. La seconde ne fait aucune hypothèse sur la loi mais elle consiste à exploiter un échantillon bootstrap composé d'un rééchantillonnage avec remise de l'échantillon initial. Les propriétés des échantillons bootstrap sont alors analysées.

Afin de mesurer l'incertitude autour de la moyenne, la méthode du bootstrap non paramétrique a été choisie. Soit, $Z = (Z_1, \dots, Z_n)$ l'échantillon initial et $Z^*(1), \dots, Z^*(B)$ les échantillons bootstrap simulés sur Z . Ainsi, un estimateur aléatoire $\theta(Z^*(1), \dots, Z^*(B))$ est fondé par un échantillon aléatoire. La procédure de bootstrap consiste en un certain nombre de répétitions.

III.2. Bootstrap des résidus de Pearson standardisés

III.2.1 Principe général

Nous allons ainsi évaluer la précision de notre estimateur $\hat{\theta}$ à partir de l'échantillon des données $Z = (Z_1, \dots, Z_n)$.

Notre estimateur correspond ici à l'estimation de la moyenne donnée par le modèle linéaire généralisé. Ce sont les résidus de Pearson standardisés qui sont alors bootstrapés afin

d'établir un nouveau jeu de données, sur lequel le modèle linéaire généralisé est alors réutilisé.

Notons :

- i la $i^{\text{ème}}$ observation
- Résidu de Pearson est défini comme $\varepsilon_i = \frac{Z_i - \hat{\mu}_i}{\sqrt{V(\hat{\mu}_i)}}$ tel que précisé page 59
- Echantillon de résidus de Pearson bootstrapés $\varepsilon^* = (\varepsilon^*_1, \dots, \varepsilon^*_n)$ obtenu par rééchantillonnage avec remise n fois des données $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)$
- Echantillon des données de bootstrap déduites $Z^* = (Z^*_1, \dots, Z^*_n)$ telles que $Z^*_i = \hat{\mu}_i + \varepsilon^*_i \times \sqrt{V(\hat{\mu}_i)}$
- θ le paramètre d'intérêt et $\hat{\theta} = s(Z)$ l'estimateur de θ basé sur l'échantillon.

III.2.2 Etapas de construction

Les modèles linéaires généralisés réalisés sur les nombres de sinistres et sur les coûts moyens du chapitre 2 ont permis de donner une estimation de la moyenne sur les deux indicateurs et des résidus de Pearson associés.

Afin de donner la volatilité autour de ces moyennes, les résidus sont bootstrapés dans une première simulation. Un nouveau jeu de données est alors déduit sur les nombres de sinistres et sur les coûts moyens tel que :

- $N^*_i = E(N_i / X) + \varepsilon^*_{nbi} \times \sqrt{V(N_i / X)}$
- $Y^*_i = E(Y_i / X) + \varepsilon^*_{cmi} \times \sqrt{V(Y_i / X)}$

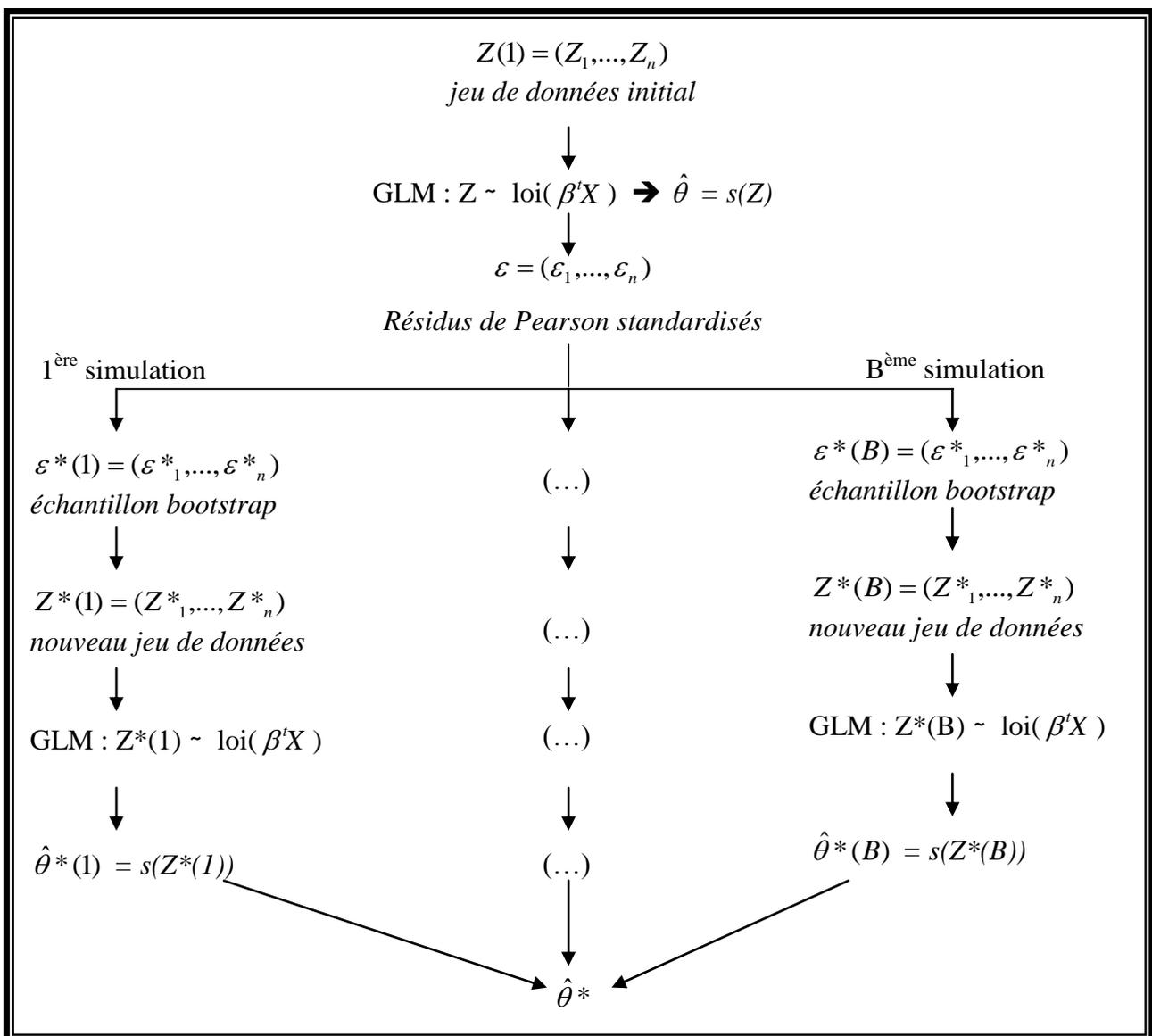
Or, en remplaçant les variances des variables aléatoires d'après les formules pages 87 et 88, on obtient :

- $N^*_i = E(N_i / X) + \varepsilon^*_{nbi} \times \sqrt{\phi_{nb} \cdot E(N_i / X)}$
- $Y^*_i = E(Y_i / X) + \varepsilon^*_{cmi} \times \sqrt{\phi_{cm} \cdot E^2(Y_i / X)}$

Avec ces nouveaux jeux de données, les modèles linéaires généralisés donnent des nouvelles estimations en moyenne sur les nombres de sinistres, sur les coûts moyens et par déduction sur les charges au niveau de chaque observation.

Cette procédure est répétée un grand nombre de fois pour donner une distribution autour des moyennes des 3 indicateurs.

Voici schématiquement les étapes de construction des échantillons bootstrap pour constituer la distribution de $\hat{\theta}$:

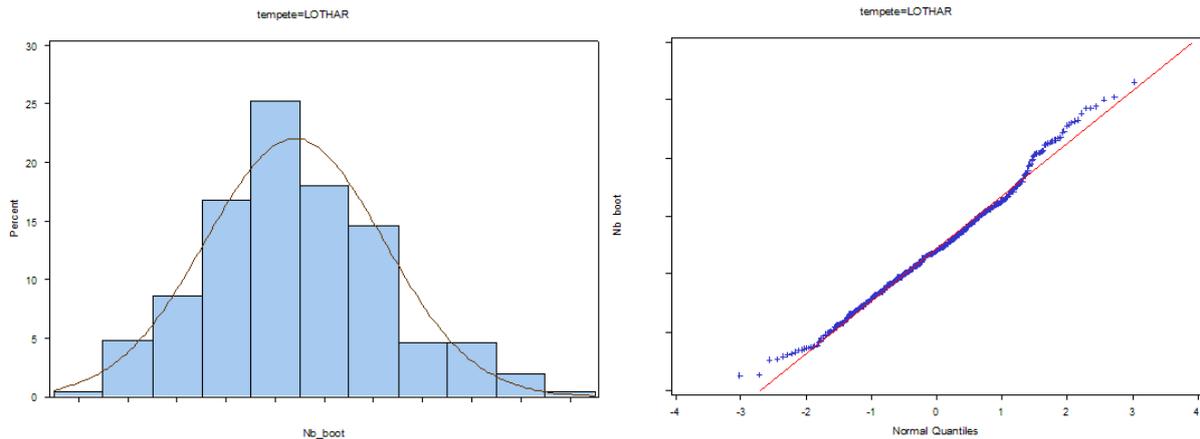


III.3. Mise en application et résultats

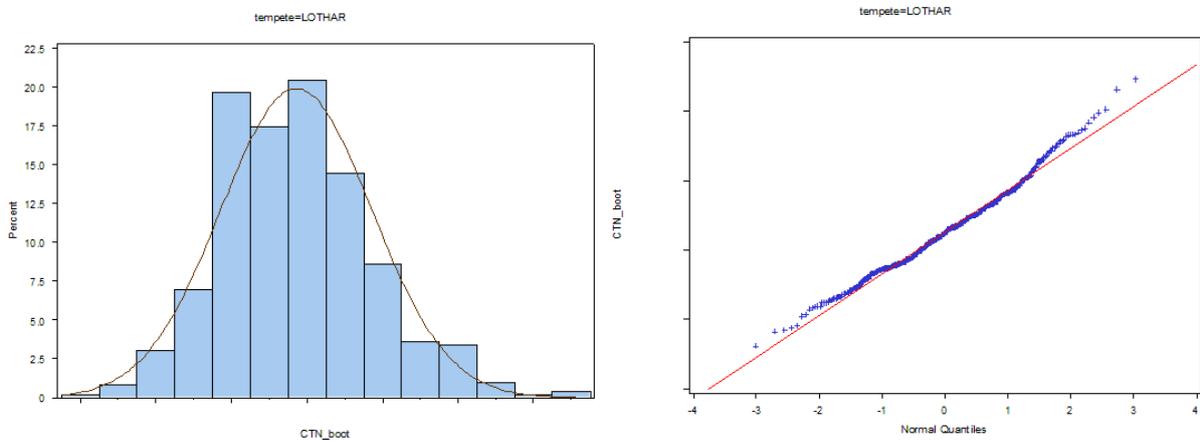
Après la Bème simulation, on aboutit à une distribution des 3 indicateurs de sinistralité : nombres, coûts moyens et charges de sinistres.

Voici un exemple sur les distributions de la tempête LOTHAR

Distribution sur les nombres de sinistres bootstrapés.



Distribution sur les charges bootstrapées.



On note d'une part, pour chaque tempête, que les distributions sont normales que ce soient pour les nombres de sinistres, pour les coûts moyens ou pour les charges.

D'autre part, on note également que le coefficient de variation, rapport entre l'écart type et la moyenne, est relativement constante pour les 6 tempêtes.

Ainsi, le passage de la moyenne à un niveau de quantile est relativement équivalent entre les tempêtes. Le coefficient de passage de la moyenne au quantile 95 pour les nombres de sinistres est de 1.1%, pour les coûts moyen de 2.9% et pour les charges de 3.5%.

III.4. Limite de la méthode du bootstrap

Nous avons montré précédemment que les observations bootstrapées Z^*i sont déterminées par résidus de Pearson. Dans le cas où $\hat{\mu}_i < \varepsilon^*_i \times \sqrt{V(\hat{\mu}_i)}$, les observations bootstrapées Z^*i sont alors négatives (plus fréquent sur les nombres de sinistres). Les modèles linéaires généralisés ne gèrent pas les valeurs négatives lorsque les lois utilisées sont Poisson ou Gamma, et les observations avec les valeurs négatives sont alors écartées du modèle.

Les valeurs prédites par les modèles sont dans ce cas là alors légèrement biaisées mais les impacts ne sont pas significatifs.

IV. Synthèse

Afin d'établir une volatilité autour de la moyenne, deux approches ont été proposées : MSEP et bootstrap.

La première consiste à réaliser des calculs matriciels en posant des hypothèses de loi des indicateurs de sinistralité des modèles linéaires généralisés. Or le calcul matriciel nécessite une grande capacité informatique.

La seconde méthode, basée sur le rééchantillonnage des résidus de pearson des modèles linéaires généralisés, permet d'analyser la distribution de la moyenne. La limite de cette méthode porte sur les valeurs reconstituées négatives.

Chapitre 4. DEFINITION DE LA DISTRIBUTION DE LA SINISTRALITE ET IMPACT TARIFAIRE

I. Objectif

L'objectif principal de ce chapitre est de connaître le tarif nécessaire pour équilibrer le risque événement climatique pour les branches commercialisées par MAAF. En effet, nous verrons qu'il est nécessaire d'élargir le périmètre tempête MRH aux autres branches et aux autres événements climatiques pour tenir compte des mécanismes de protection que sont les provisions pour égalisation et la réassurance.

Dans un premier temps, il est nécessaire d'évaluer le coût annuel moyen d'une tempête. Pour ce faire, une courbe est réalisée donnant une relation entre une charge et une période de retour d'une tempête. Dans un second temps, des hypothèses complémentaires sont données pour les autres périls constituant les événements climatiques (grêle, inondation, neige ...). Ainsi, une charge annuelle peut être établie en brute de cession puis en nette et les incidences sur les provisions pour égalisation peuvent être mesurées.

En ramenant les cotisations brutes et nettes de cession ainsi que les frais afférents aux garanties climatiques, on évalue ainsi le ratio combiné et la suffisance du tarif.

II. Prime pure annuelle moyenne du risque tempête

La détermination de la prime pure, charge rapportée à un contrat, annuelle moyenne va s'articuler en deux parties.

La première partie consiste à réaliser, sur le portefeuille au 31.12.2012, une courbe des coûts de sinistres de tempêtes annuelles en fonction des périodes de retour pour les tempêtes exceptionnelles.

Avec des hypothèses sur les tempêtes récurrentes (donc de plus faibles intensités) et avec des simulations sur les tempêtes exceptionnelles par année, des coûts de tempêtes sont alors estimés et une charge moyenne annuelle est ainsi déterminée dans la seconde partie.

II.1. Méthode de la constitution de la courbe de sinistralité en fonction des périodes de retour

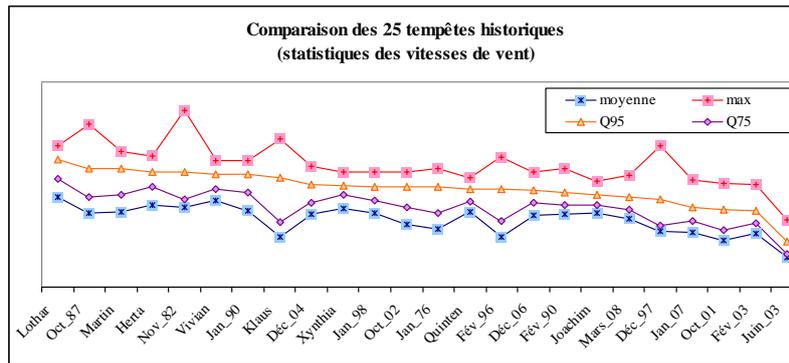
L'ajustement des courbes d'endommagement, exposées dans le chapitre 2 « Modélisation lors de la survenance d'une tempête », permet d'établir le coût des tempêtes avec les caractéristiques du portefeuille actuel, appelé coût as if. Cette démarche est la première étape pour apprécier le risque tempête, correspondant au premier aléa « évaluation lors de la survenance de la tempête ».

La seconde est d'associer une probabilité, c'est-à-dire une période de retour, à chacune de ces tempêtes afin d'ajuster une relation entre la charge d'une tempête et sa période de retour pour l'établissement de la courbe de charge de sinistres en fonction de période de retour, correspondant au deuxième aléa « sur la probabilité de la survenance ».

II.1.1 Coût as if sur le portefeuille MRH

II.1.1.1 Analyse des vitesses de vent de 25 tempêtes

Covéa a acheté les vitesses de vents de 25 tempêtes historiques et lissé ces vitesses au niveau du code postal. L'analyse statistique de ces vitesses de vent par tempête permet de comparer ces 25 tempêtes et d'en donner l'ordre d'importance de ces tempêtes.

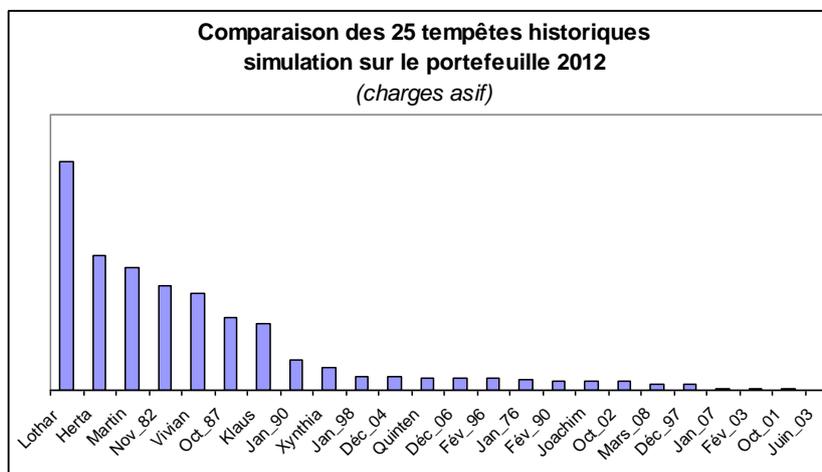


Ce graphique permet également de juger les tempêtes par famille. Les tempêtes ayant un quantile 95 de plus de 130 km/h sont considérées comme majeures (de Lothar jusqu'à Klaus). Les autres tempêtes sont considérées comme mineures. Cette classification est essentielle pour le modèle d'estimation des nombres de sinistres et donc de l'évaluation des charges de nos 25 tempêtes.

II.1.1.2 Estimation des coûts as if des 25 tempêtes

Les modèles linéaires généralisés du chapitre 2 ont été élaborés avec les sinistres et le portefeuille de l'époque. Ils ont également déterminés les paramètres des facteurs discriminants.

L'idée est d'actualiser les coûts au regard des risques souscrits au 31.12.2012 (appelés coûts as if). Le principe est donc d'estimer les nombres et les coûts moyens à partir des paramètres des modèles linéaires généralisés, des données portefeuilles MRH à la situation au 31.12.2012 et de la vitesse de vent au niveau de chaque code postal. Autrement dit, on simule les 25 tempêtes à fin d'année 2012 et on établit les 25 coûts as if, illustrés sur le graphique ci-dessous.



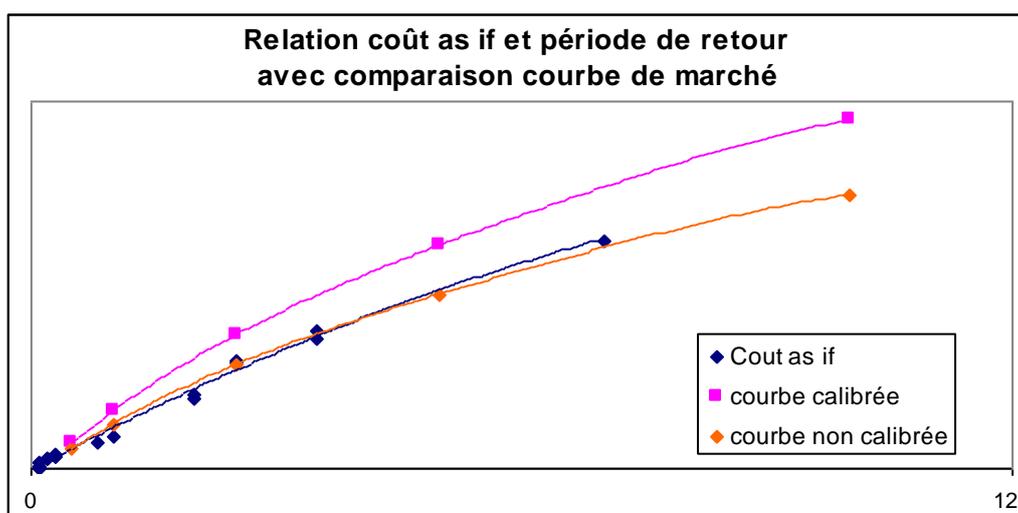
L'ordre d'importance des tempêtes sur les charges est relativement similaire à celui sur les vitesses de vents. Les différences se justifient par des trajectoires différentes entre les tempêtes et donc portent sur des caractéristiques de codes postaux divergentes (zones de norme de construction, nombre de propriétaires ...) et par son étendu sur le territoire français.

II.1.2 Ajustement des coûts as if en fonction des périodes de retour

II.1.2.1 Logiciels marché : source et comparaison

A l'aide des logiciels RMS, EQECAT ou AIR, les courtiers peuvent estimer la période de retour de certaines tempêtes exceptionnelles. Ces estimations sont variables selon les logiciels, selon leur version et selon les courtiers. Néanmoins, elles nous permettent d'apprécier le niveau de la période de retour et d'y associer la charge as if de nos tempêtes.

Pour les tempêtes restantes, des périodes de retour sont alors affectées à chaque tempête au regard de leur charge.



Les charges as if des tempêtes exceptionnelles et le choix sur leur période de retour sont proches des estimations du logiciel RMS sans calibrage. Une courbe avait été calibrée par rapport à la courbe du logiciel car il avait été estimé que les charges de certaines tempêtes étaient sous-estimées. Ce benchmark permet de valider la pertinence à la fois sur les modèles linéaires généralisés (vision moyenne des charges) et à la fois sur les périodes de retour associées et de juger que la courbe calibrée semble surestimer notre risque.

II.1.2.2 Ajustement d'une courbe

a) Hypothèses du profil et du prolongement de la courbe

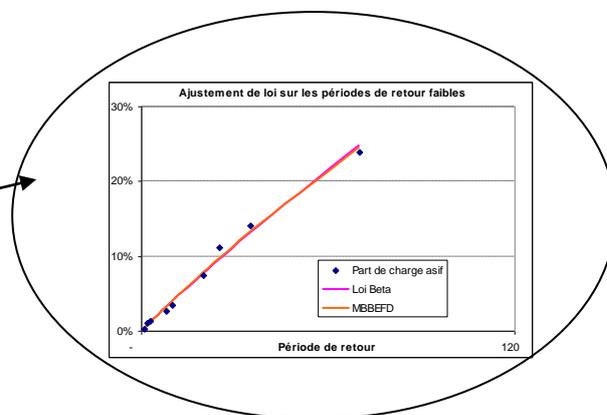
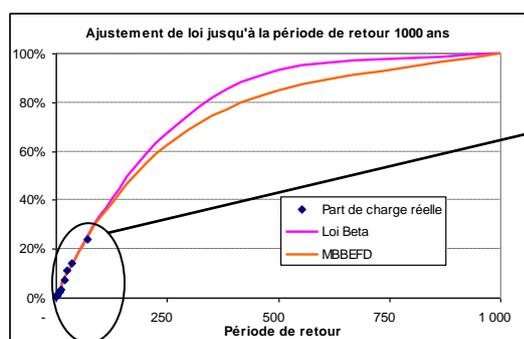
La courbe du « coût as if » du graphique précédent a un profil logarithmique. Ce profil est à l'image des courbes proposées par les courbes de marché.

Se pose alors la question sur un calibrage sur les périodes de retour au-delà de celui de LOTHAR, période de retour la plus élevée de nos 25 tempêtes historiques. L'idée vise à s'appuyer sur les estimations de nos courtiers en établissant une charge maximale cible, celle proposée sur la période de retour de 1000 ans.

b) Méthodologie de la constitution de la courbe

Dans un premier temps, une moyenne des charges as if de nos 25 tempêtes sont déterminées par période de retour. En effet, plusieurs tempêtes peuvent avoir une même période de retour. Dans un second temps, ces charges moyennes sont rapportées à la valeur maximale cible. Ce rapport est donc compris entre 0% et 100%, ce dernier correspondant à la charge maximale cible de la période de retour 1000 ans. Comme les courbes de marché, cette progression est jugée logarithmique. C'est pourquoi, ce rapport est ajusté selon une loi Bêta ou MBBEFD. Cet ajustement est réalisé par la méthode des moindres carrés pour établir les paramètres des lois.

c) Résultat et interprétation



On note que pour les tempêtes historiques jusqu'à LOTHAR, les ajustements des deux lois sont très proches puis la loi Bêta diverge en s'approchant plus rapidement du taux de 100%, c'est-à-dire de la valeur cible de la période de retour 1000 ans.

Pour sa vision légèrement prudentielle, c'est la loi Bêta qui a été retenue pour déterminer le rapport en fonction de la période de retour. Néanmoins, il est à noter qu'au-delà Lothar la probabilité de survenance est faible de 1.4% (1/70 ans). Ensuite ces taux issus de cet ajustement sont convertis en charges.

II.2. Distinction des tempêtes récurrentes et exceptionnelles.

La courbe de dommages selon les périodes de retour n'est pas adaptée pour les tempêtes récurrentes. En effet, elle débute par une période de retour d'un an. Qu'en est-il pour tous les coups de vent locaux ou/et modérés partout en France engendrant des dommages ?

Ainsi seule la courbe au-delà 5 ans est sélectionnée pour les tempêtes exceptionnelles. En deçà, les primes pures historiques hors tempêtes exceptionnelles seront analysées à part.

II.2.1 Analyse des primes pures des tempêtes récurrentes MRH

Les coûts des tempêtes récurrentes sont observés sur les exercices de survenances passés. C'est pourquoi, les primes pures des garanties tempêtes ont été retraitées des tempêtes exceptionnelles sur la période d'observation (Lothar, Martin, Klaus, Quinten et Xynthia).

L'hypothèse de la loi de la prime pure des tempêtes récurrentes est jugée normale. L'analyse permet d'établir les moments des primes pures comme la moyenne et l'écart type. Elle ne montre pas de tendance d'une survenance à l'autre. C'est pourquoi, les paramètres de la loi correspondent aux valeurs empiriques.

II.2.2 Simulation des tempêtes exceptionnelles MRH

Ce paragraphe vise à estimer la prime pure annuelle moyenne pour les produits MRH en simulant des tempêtes exceptionnelles annuelles. Cette simulation des tempêtes consiste à réaliser un tirage dans la courbe de charges de sinistres avec une hypothèse sur les nombres d'évènements.

II.2.2.1 Tirage dans la distribution de charge

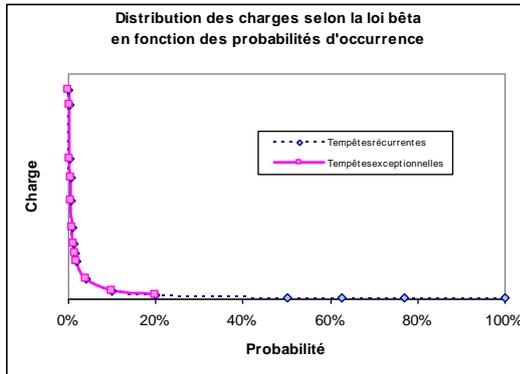
La courbe de dommages donne une relation entre la charge d'un évènement et sa période de retour. Cette période de retour (PR) peut être traduite en probabilité de survenance α :

$$PR = \frac{1}{1 - \alpha}$$

Par exemple, une période de retour de 2 ans donne une probabilité de survenance de 50% et une probabilité de 99,5% correspond à une période de retour de 200 ans.

Afin de définir la charge pour un évènement, on réalise alors un tirage uniforme.

$$\alpha \sim U(0,1)$$



Le graphique ci-contre est la traduction de la courbe issue de la loi bêta, en convertissant des périodes de retour en probabilité.

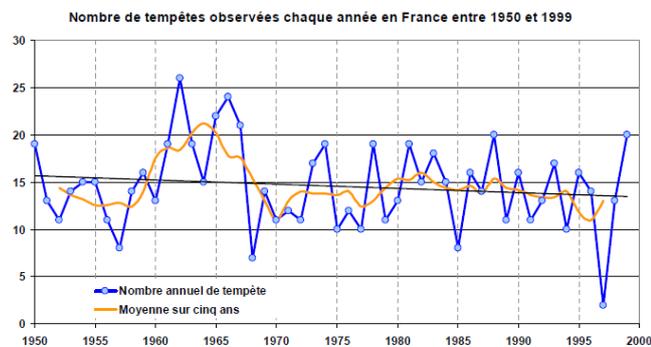
La fonction en deçà de la probabilité d'occurrence de 20%, correspondant à la période de retour au-delà de 5ans, est à assimiler aux tempêtes exceptionnelles.

Ainsi, lorsque le tirage uniforme donne une probabilité supérieure à 20%, aucune charge n'est retenue via la courbe de charge de sinistres puisqu'elle est jugée être prise en compte dans la prime pure des tempêtes récurrentes.

II.2.2.2 Tirage du nombre d'évènement

Pour rappel (p.30), Météo France qualifie un évènement tempétueux comme tempête lorsque 5% des stations météorologiques enregistrent une vitesse de vent maximal supérieur à 100 km/h.

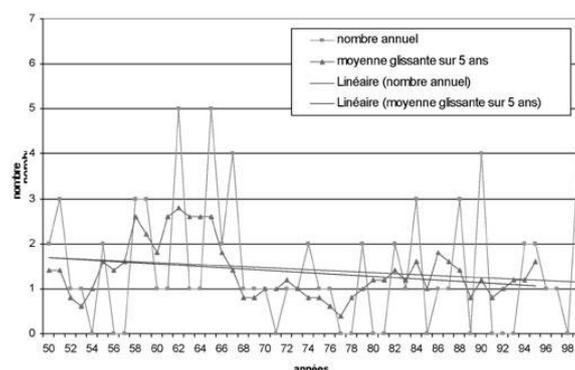
On dénombre en moyenne 15 tempêtes par an avec une variabilité très forte entre 2 et 26. Néanmoins, on ne note pas de tendance significative d'évolution de l'intensité des tempêtes en France entre 1950 et 2000, ce qui par ailleurs remet en cause les dérives du changement du climat.



http://www.cnrs.fr/cw/dossiers/dosclim1/biblio/pigb15/08_evolution.htm

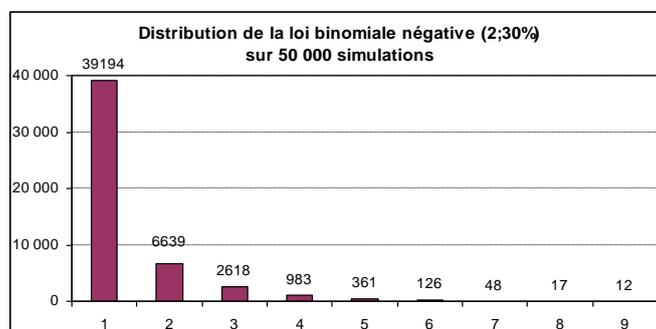
Les tempêtes sont qualifiées de « fortes » lorsque la part passe de 5% à 20% selon la définition de Météo France. Elle représente en moyenne une tempête sur 10.

Sur nos tempêtes historiques ayant une période de retour d'un an, on remarque que 20% des communes ont une vitesse de vent très proche de 100 km/h, correspondant au critère des fortes tempêtes. Par voie de conséquence, les tempêtes ayant une période de retour d'au moins un an sont considérées comme « fortes ».



http://www.cnrs.fr/cw/dossiers/dosclim1/biblio/pigb15/08_evolution.htm

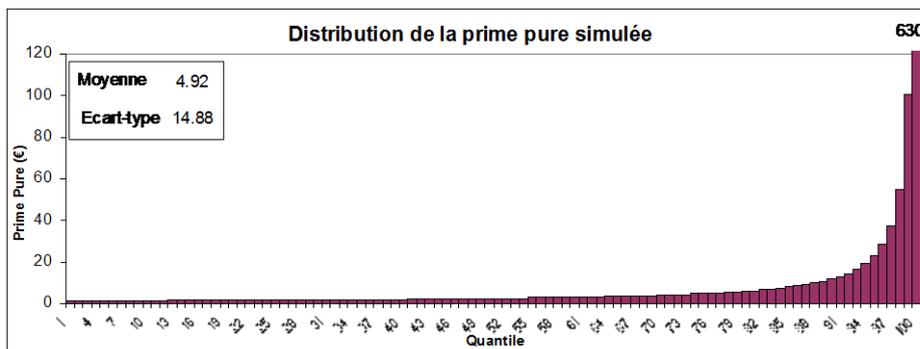
Ainsi, le nombre d'évènement à tirer dans la courbe de charge de sinistres doit à minima être d'un évènement par an. On suppose que le nombre d'évènement dans l'année suit une loi Binomiale Négative (2, 30%) avec un minimum d'un sinistre par an, et pour tenir compte du phénomène de répétition potentielle dans l'année (exemples : en 1999 avec LOTHAR et MARTIN, en 2009 avec KLAUS et QUINTEN) Le paramètre à 2 a été privilégié à 1 pour obtenir une moyenne plus proche de 1.4. Le second paramètre a été fixé à 30% pour éviter une queue trop lourde. En effet, à titre d'exemple, la part des évènements de plus de 3 par an passe de 3% à 9% pour un paramètre de 40%.



Cette hypothèse sur les nombres d'évènements donne une probabilité d'une tempête par an à 78%, 2 évènements à 13% et 3 à 5%.

II.2.2.3 Détermination de la prime pure annuelle brute de réassurance

A partir des hypothèses sur les tempêtes récurrentes (loi normale de la prime pure) et exceptionnelles (loi bêta de la courbe de charge de sinistres) ainsi que sur les nombres d'évènements (loi binomiale), des scénarii de tempêtes sont alors simulés. Ainsi, une charge annuelle en est déterminée ainsi qu'une distribution.



Ce graphique illustre la distribution de la charge ramené à un contrat.

La prime pure tempête moyenne annuelle s'établit alors à 4.92€ (correspondant au quantile 75). Pour comparaison, la prime pure issue de la courbe de marché retenue avec un calibrage s'établit à 5.82€, soit une différence de 18%.

II.2.2.4 Sensibilité des hypothèses

La construction de la courbe de charge de sinistres a nécessité un ensemble d'hypothèses. Ce paragraphe vise donc à connaître la sensibilité de chacune d'elles sur la prime pure.

Pour ce faire, la courbe de charge de sinistres a donc été reconstruite à chaque changement d'hypothèse.

a) Sensibilité des charges as if moyenne

La moyenne de la charge issue des modèles linéaires généralisés est aussi sujette à une volatilité. L'étude du chapitre 3 nous conduirait à augmenter la moyenne, et donc nos charges des tempêtes, de 3.5% pour la sécuriser à 95%.

Cette sécurisation nous amènerait à une prime pure annuelle en progression de 2.5%, soit 5.04€ : +3.5% sur les charges des tempêtes exceptionnelles diluées par celles des tempêtes récurrentes.

b) Sensibilité sur la loi et le montant maximal cible

La fonction entre les charges et les périodes de retour a été ajustée selon deux hypothèses :

- Etablissement d'un montant à atteindre pour la période de retour 1000 ans
- Choix de la loi Bêta au détriment de celle de la MBBEFD

		Sensibilité de la prime pure selon certaines hypothèses		
		Montant pour la période de retour 1000 ans		
Lois		500 000	650 000	700 000
Bêta		4.61 €	4.92 €	5.01 €
MBBEFD		4.66 €	4.90 €	4.95 €

On note que la différence de loi est moins notable que le choix du montant cible. Ces montants cibles, proposés ci-dessus (0.5M€, 0.65M€ et 0.7M€), correspondent aux différentes valeurs des courbes de marché calibrée ou non selon les logiciels.

c) Sensibilité sur la période de retour

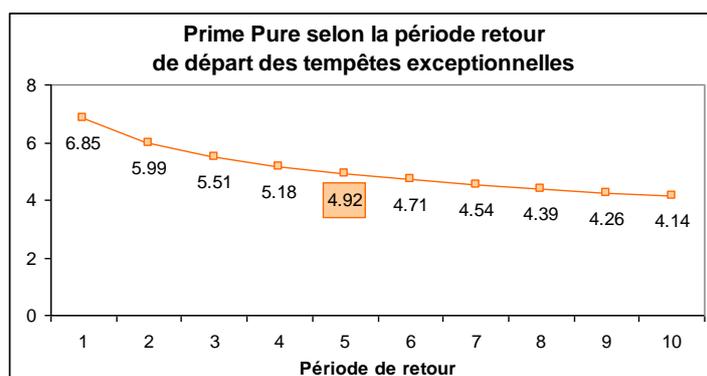
L'étude de sensibilité la plus marquée porte sur la période de retour. En effet, une variation de 1% sur la période de retour entraîne une variation de près de 1% sur la prime pure annuelle moyenne.

Sensibilité de la prime pure selon repositionnement de la période de retour

Période de retour	Prime Pure (en €)
-10%	5.38 (+9.4%)
-5%	5.16 (+4.7%)
0%	4.92
+5%	4.63 (-5.9%)
+10%	4.46 (-9.4%)

d) Sensibilité sur la troncature de courbe de dommages

On peut considérer que les tempêtes récurrentes correspondent à des tempêtes de période de retour entre un an et 10 ans. C'est pourquoi, la courbe de charge de sinistres au-delà une période de retour de 5 ans a été utilisée pour les tempêtes exceptionnelles. Les primes pures varieraient de plus ou moins 10% si on influe sur les périodes de retour acceptables à savoir entre 3 et 7 ans entre les tempêtes récurrentes et exceptionnelles.



III. Impact tarifaire

Afin de se protéger des événements climatiques majeurs, l'assureur utilise des mécanismes de protection : la réassurance pour protéger les fonds propres et les provisions pour égalisation pour lisser le résultat. Ces événements climatiques englobent certes les tempêtes MRH mais également ceux des autres produits ainsi que d'autres événements climatiques.

Après une présentation des mécanismes de protection, des hypothèses complémentaires seront exposées afin de mesurer les incidences sur ces mécanismes et sur la tarification.

III.1. Mécanismes de protection

III.1.1 les traités de réassurance

La réassurance est une opération juridique qui se matérialise par un contrat passé entre une entreprise d'assurance (cédante) et une entreprise de réassurance (cessionnaire).

Ce contrat consiste à un échange entre les parties :

- versement d'une prime payée par la cédante à un ou plusieurs réassureurs, appelée prime cédée
- en contrepartie, indemnisation d'une partie des sinistres subis (ou de perte).

L'objectif de la réassurance est de baisser la volatilité du résultat et de protéger les fonds propres de l'entreprise d'assurance, en transférant tout ou partie du risque.

Les traités de réassurance sont de différentes natures

III.1.1.1 Réassurance proportionnelle

- Quote-part (QP)

L'assureur cède au réassureur un pourcentage des primes d'un portefeuille donné, et en contrepartie le réassureur verse la même proportion des sinistres.

Ce type de traité de réassurance est notamment utilisé pour le risque catastrophes naturelles.

- Excédent de capitaux ou de pleins (XP)

A chaque police, il est calculé un taux de cession lorsque le capital assuré dépasse un seuil (le plein de conservation). Ce taux correspond au rapport entre le montant du capital assuré déduction faite du plein sur le montant total du capital assuré.

Il est exploité pour les risques avec des capitaux comme l'individuel accident et l'assurance décès.

III.1.1.2 Réassurance non proportionnelle

- Excédent de sinistres par risque ou/et par évènement (XS)

Le réassureur n'intervient que pour les sinistres dépassant un certain seuil, appelé priorité (correspondant à la franchise en assurance), jusqu'à un autre seuil, appelé plafond. La portée est l'intervention maximale du réassureur et donc l'écart entre le plafond et la priorité.

Ce traité s'écrit communément « portée XS priorité ». Par exemple, un traité 30 XS 50 signifie que la priorité est de 50M€, la portée 30M€ et le plafond de 80M€.

Afin de limiter le risque de défaut des réassureurs, les traités sont répartis entre différents acteurs.

Il est souvent utilisé pour les risques de RC (Responsabilité Civile) avec un plafond illimité et pour les évènements climatiques avec un plafond limité.

Ce traité comporte certaines clauses :

⇒ clause obligatoire

Au sein du contrat, il est défini la priorité et la portée avec la précision si le traité intervient par risque ou par évènement.

⇒ clause aggregate

➤ clause de franchise annuelle

Afin de limiter le coût de la réassurance, l'assureur peut conserver une partie du montant de sinistres au-delà de la priorité à hauteur de cette franchise.

➤ clause de plafond annuelle

On définit ici l'engagement maximum d'intervention du réassureur.

⇒ clause de reconstitution

En cas de fréquence importante ou de sinistre exceptionnel, le réassureur s'engage à rétablir la couverture, en contrepartie d'une prime complémentaire définie au préalable.

Le nombre de reconstitution est défini au moment de la souscription du traité.

⇒ clause de participation bénéficiaire

L'assureur peut recevoir une participation au bénéfice lorsque le solde réassurance est excédentaire.

⇒ clause de « no claim bonus »

Il est accordé une réduction sur la prime, si le traité n'a pas joué dans l'année.

○ Excédent de perte ou Stop-Loss (SL)

Le principe est identique à celui de l'excédent de sinistres. Les montants de la priorité et de la portée sont remplacés par un ratio technique (S/C).

Il est souscrit pour les risques cycliques (comme les événements climatiques ou catastrophes naturelles).

III.1.1.3 Exemple de traités sur les risques « catastrophes naturelles » et « événements climatiques »

Les sinistres sont enregistrés dans les garanties catastrophes naturelles si le bien endommagé se trouve dans une commune ayant fait l'objet d'un arrêté catastrophes naturelles. Les autres sinistres des événements climatiques sont enregistrés dans des différentes garanties événements climatiques : tempête, grêle, neige, inondation ...

a) Traités Catastrophes Naturelles

Pour le risque catastrophe naturelle, les assureurs français ont pour réassureur la CCR car elle offre une garantie illimitée couverte par une garantie de l'état.

Elle propose deux traités de réassurance : un quote-part et un stop loss.

Exemple de mise en application :

- Hypothèses retenues

Encaissement de l'assureur : 1.000.000 €

Pourcentage de cession en quote-part : 50 %

Rétention de l'assureur : 50 % soit 500.000 €

Franchise du stop-loss : 200 % de 500.000 € soit 1.000.000 €

Sinistralité annuelle = 10.000.000 € (en 1 ou n sinistres)

- Résultats par traité

Répartition au titre de la quote part :

- Assureur 50 % = 5.000.000 €

- CCR 50 % = 5.000.000 €

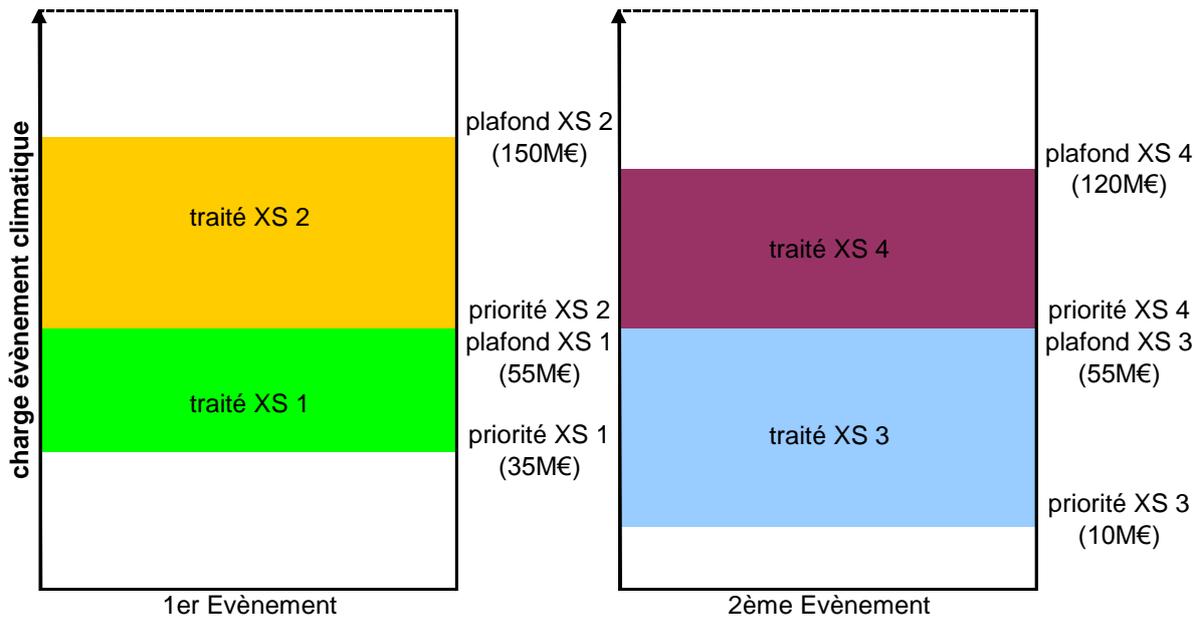
Répartition au titre du stop-loss :

- Assureur = 1.000.000 € (montant de la franchise)

- CCR = 5.000.000 € - 1.000.000 € = 4.000.000 €

Au total, la charge de CCR s'élève donc à 9.000.000 € tandis que celle de l'assureur reste limitée à 1.000.000 €.

b) Traités Evènements Naturels



Exemple de mise en application :

- Hypothèses retenues

Charge 1^{er} évènement : 140 M€
 Clause sur 1^{er} évènement : 20 XS 35 et 95 XS 55
 Charge 2^{ème} évènement : 130 M€
 Clause sur 2^{ème} évènement : 45 XS 10 et 65 XS 55

- Résultats par traité

1^{er} Evènement

Répartition au titre 20 XS 35 :

- Assureur = 35 M€
- Réassureur = 20 M€

Répartition au titre 95 XS 55 :

- Assureur = 0 M€
- Réassureur = 85 M€

2^{ème} Evènement

Répartition au titre 45 XS 10 :

- Assureur = 10 M€
- Réassureur = 45 M€

Répartition au titre 65 XS 55 :

- Assureur = 10 M€
- Réassureur = 65 M€

Au total, la charge du réassureur s'élève donc à 215 M€ tandis que celle de l'assureur reste limitée à 55 M€. Ce dernier a payé le montant de rétention (franchise) de 45 M€ mais également le montant dépassant le plafond 10 M€. En cas d'une reconstitution, l'assureur n'aurait pas payé au-delà le plafond mais la rétention serait passée à 70M€.

III.1.2 les provisions pour égalisation

Extrait de l'article R331-6 du code des assurances

« Ces provisions sont destinées à faire face aux charges exceptionnelles afférentes aux opérations garantissant les risques dus à des éléments naturels, le risque atomique, les risques de responsabilité civile dus à la pollution, les risques spatiaux, les risques liés au transport aérien, et les risques liés aux attentats et au terrorisme. »

Trois grandes familles de garanties sont impactées par ces provisions : catastrophes naturelles, événements climatiques et attentats.

Les provisions pour égalisation sont une réserve destinée à gérer les sinistres de fond (récurrents et à intensité limitée). Elles permettent de lisser la charge des événements sur plusieurs années. En effet, en cas de non événement majeur, le gain de l'année dote cette provision. Si l'année suivante l'évènement survient (charge de l'évènement > prime de l'année), la prime de l'année et la provision de l'année passée sont alors utilisées.

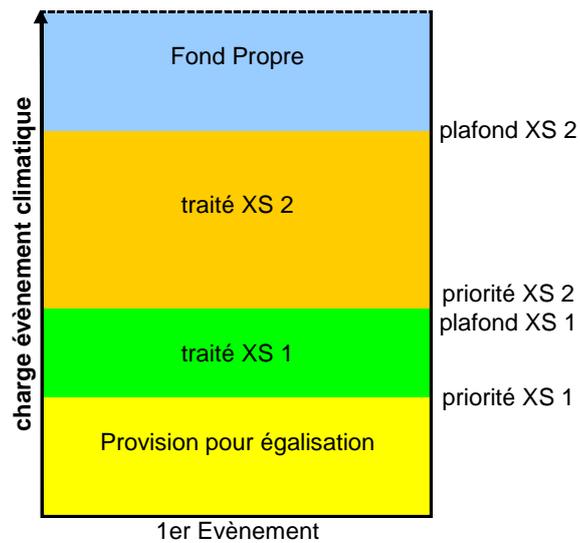
Le stock de cette provision ne peut pas excéder 300% de la cotisation nette de cession pour les risques événements climatiques et catastrophes naturelles (et 500% pour le risque attentat).

Méthode de calcul :

- ❖ Stock 31/12/N-1
- ❖ + Dotation = minimum entre
 - 75% du résultat (gain) comptable net de cession
 - 300% (prime de l'année – prime cédée - frais) - stock 31/12/N-1 (*dotation maximum*)
- ❖ - Reprise = maximum entre
 - 100% du résultat (perte) comptable net de cession
 - Stock 31/12/N-1

Ce calcul du montant de la dotation est réalisé au niveau d'une société et pour tous les produits intégrant des garanties événements climatiques. Cette dotation (ou reprise) est ensuite répartie par produit.

III.1.3 Illustration globale des mécanismes de protection



En somme, comme l'illustre le graphique ci-dessus, les provisions pour égalisation permettent ainsi de limiter (ou compenser) les rétentions du programme de réassurance.

Pour ne pas puiser sur tout le stock des provisions pour égalisation dès le premier évènement climatique important, un premier traité XS intervient avec une priorité basse (inférieure à 300% des cotisations nettes de cession).

Le second traité XS a pour objet de protéger les fonds propres de l'entreprise et il est choisi pour se couvrir contre un évènement très exceptionnel en définissant le plafond de ce traité (par exemple : évènement avec une période de retour de 200 ans).

Si la charge de l'évènement exceptionnel dépasse le plafond, le surcoût est alors puisé dans les fonds propres (dans le cas où le stock des provisions pour égalisation a été entièrement utilisé).

A travers cette illustration, nous montrons l'intérêt :

- des mécanismes de protection : Provision pour égalisation et traité de cession en excédent de sinistre.
- de connaître les risques évènements climatiques et plus particulièrement la tempête en définissant le niveau de plafond maximum et donc en mesurant le risque de toucher les fonds propres.

III.2. Détermination des ratios combinés des évènements climatiques

Pour tenir compte à plein des mécanismes de protection, il faut élargir le périmètre des évènements climatiques aux tempêtes du portefeuille professionnel et aux autres périls des évènements climatiques.

Des lois sont tout d'abord établies pour ces périmètres. Ensuite, des simulations stochastiques d'évènements sont proposées sur les lois de ces périmètres pour aboutir à une prime pure à chaque périmètre. A chacune de ces simulations, les traités de réassurance MAAF de 2012 ont été appliqués et les dotations des provisions pour égalisation sont alors déduites. Dès lors, les ratios combinés par produits seront déterminés.

III.2.1 Hypothèses sur les autres périmètres

Voici les hypothèses complémentaires à celles des tempêtes MRH.

III.2.1.1 Evaluation du risque tempête Professionnel

L'approche sur le portefeuille MRH aurait pu être appliquée sur celui professionnel : modélisation des nombres de sinistres et des coûts moyens selon les modèles linéaires généralisés pour l'obtention de la charge moyenne puis attribution de la même période de retour à chaque tempête.

L'approche retenue est plutôt de faire l'hypothèse que la courbe de charge de sinistres professionnelle est proportionnelle à celle de la MRH. Cette proportion est considérée comme le rapport de la somme assurée, informations fournies aux courtiers, entre la MRH et les professionnels.

III.2.1.2 Evaluation des autres évènements climatiques

Deux approches ont été retenues pour compléter l'exhaustivité des périmètres évènements climatiques :

- Produit Automobile : tempêtes et grêle des évènements majeurs

Pour ce périmètre, les courbes fournies par nos courtiers ont été exploitées afin de déterminer la prime pure des évènements exceptionnels.

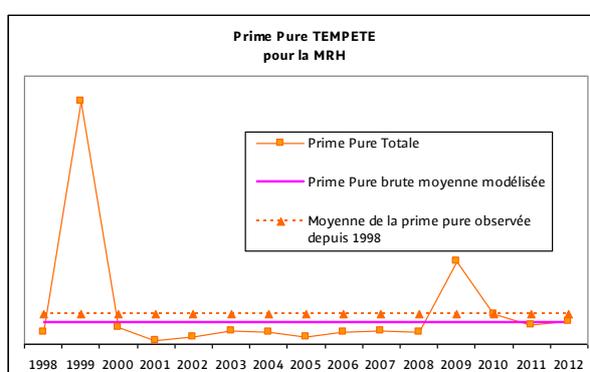
La même démarche que la tempête MRH a été utilisée. Afin de retenir que les évènements exceptionnels, seuls les résultats de la courbe au-delà 5 ans ont été pris en compte.

- Autres périmètres

Pour les périmètres restants (tempêtes et grêles automobile des évènements non exceptionnels, inondation, grêle MRH et Professionnelle ...), une analyse des primes pures historiques de l'ensemble des garanties climatiques hors les évènements exceptionnels est alors réalisée sur les produits, professionnels et automobile.

Il a été retenu une loi normale On ne note pas de tendance particulière, c'est pourquoi les paramètres de la loi sont les moyennes et les écarts types empiriques par branche.

III.2.1.3 Benchmark sur les primes pures historiques.



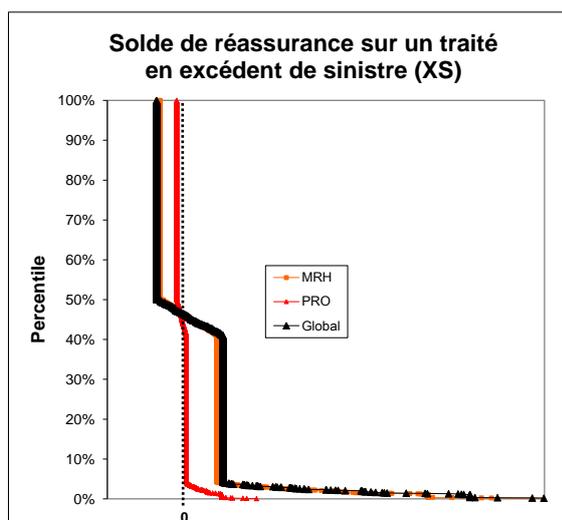
Ce graphique illustre les primes pures observées par survenance avec la moyenne empirique.

Il permet de montrer la pertinence de la prime pure moyenne modélisée.

III.2.2 Analyse des résultats des traités et des provisions pour égalisation

A chaque simulation, de tempêtes MRH et d'autres évènements, correspondant à une année, l'ensemble du programme de réassurance du groupe MAAF ont été mis en jeu (excédent de sinistre des différents réassureurs en tenant compte des « aggregate » et des reconstitutions ainsi que des Stop Loss) puis les dotations et les reprises des provisions pour égalisation sont alors calculées.

III.2.2.1 Exemple de résultat d'un traité XS

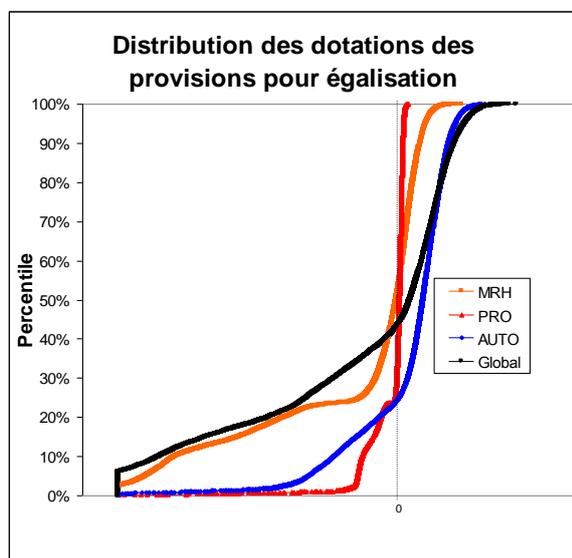


Le graphique est excentré sur la droite. Les valeurs de gauche (négatives) correspondent aux primes cédées qui sont supérieures aux charges cédées. Les valeurs extrêmes de gauche sont bornées aux primées cédées et de droite à la portée du traité moins les primes cédées

Globalement, on note que la prime cédée (et prime de reconstitution) est supérieure à celle des sinistres cédées (déduction des « aggregate ») sur 50% des cas. Ce traité avec sa priorité intervient donc en moyenne tous les 2 ans.

Les résultats sur les produits professionnels sont plus faibles que ceux sur les produits MRH, de part le nombre de contrats en portefeuille.

III.2.2.2 Dotations des provisions pour égalisation



Le graphique est également excentré sur la gauche. En effet, les courbes à gauche correspondent aux reprises qui sont limitées aux stocks. A droite, ce sont les dotations qui sont limitées à 300% des cotisations.

Les provisions pour égalisation sont dotées dans 60% des cas : 40% pour les produits MRH, 55% pour les professionnels et 80% pour l'automobile. En somme, les provisions pour égalisation progressent 3 années sur 5. Or, dans le cas contraire, la reprise peut être plus conséquente et ainsi, en moyenne les provisions pour égalisation ont plutôt tendance à se réduire.

III.2.3 Compte de résultat par branche

Afin de connaître la suffisance du tarif actuel de l'ensemble des garanties des événements climatiques, les cotisations acquises moyennes au 31.12.2012 par branche sont comparées à leur prime pure moyenne brute de réassurance issue des simulations et aux différents frais moyen, dans un premier temps, pour l'obtention du ratio combiné brut de réassurance.

Dans un second temps, les cessions (cotisation moyenne et prime pure) sont ajoutées au ratio combiné pour une vision nette de réassurance.

Lorsque le ratio net de réassurance est supérieur à 100% alors le tarif est jugé insuffisant. Il est également jugé insuffisant au-delà de 95% au regard des dotations des provisions pour égalisation. En effet, lorsque le résultat net de réassurance de l'exercice est positif alors seul 75% du bénéfice dote les provisions pour égalisation (contre 100% des pertes pour la reprise).

Ainsi, lorsque le ratio combiné net de réassurance est supérieur à 95%, le tarif doit être majoré soit au niveau de chaque branche, soit toutes branches confondues.

IV. Synthèse

Afin d'estimer une charge des tempêtes exceptionnelles, un tirage d'évènement est réalisé sur une courbe de charge de sinistres en fonction de la probabilité de survenance (la période de retour). Pour définir la courbe, il a été évalué une charge pour 25 tempêtes historiques en fonction de la structure actuelle du portefeuille à l'aide des modèles linéaires généralisés. Puis, on définit une période de retour à chacune de ces tempêtes. Enfin, une courbe est calibrée selon une loi béta. Ainsi, pour simuler des tempêtes, un nombre d'évènement annuel a été modélisé selon une loi binomiale négative avec un tirage sur la courbe définie précédemment.

Les charges des tempêtes récurrentes ont été approximées selon une loi normale.

Ces simulations permettent d'estimer la charge annuelle moyenne brute de cession et de mesurer les impacts de mécanismes de protections (réassurance et provision pour égalisation) avec des hypothèses complémentaires sur les autres périls climatiques pour l'ensemble des produits intégrant des évènements climatiques.

Pour juger du niveau de tarif, ces charges brutes et nettes de cession simulées ont été ajoutées aux cotisations cédées et aux frais puis rapportées aux cotisations brutes de cession pour élaborer un ratio combiné. L'analyse des dotations des provisions pour égalisation a montré que le ratio combiné doit être en deçà de 95% pour être en capacité de doter.

Conclusion Générale

Le risque tempête est un enjeu majeur pour les assureurs puisqu'en cas de dépassement du plafond de réassurance, les fonds propres sont utilisés pour indemniser les sinistrés et la solvabilité de l'entreprise peut être remise en cause.

Les modèles linéaires généralisés ont permis d'établir des courbes de sinistralité dépendant de l'intensité du vent. Ces courbes sont opérationnelles pour les produits Multirisques Habitation en fonction de critères de portefeuilles et de zones géographiques. *Elles permettront à COVEA de gérer les flux de sinistres et estimer leur coût lors de la survenance d'une prochaine tempête.*

A l'aide de cette modélisation et des hypothèses sur les périodes de retour de tempêtes exceptionnelles, les simulations ont permis d'estimer le coût annuel d'une tempête et plus largement les événements climatiques. Afin de juger le tarif actuel, des éléments comptables complémentaires ont été intégrés dans la modélisation à travers des mécanismes de protection comme la réassurance et les provisions pour égalisation. Suite à cette étude, *MAAF Assurances a été amené à réviser son niveau de tarif.*

Dans le prolongement de ce mémoire, il peut être relevé quelques axes d'approfondissement.

La modélisation des courbes de la sinistralité pourrait encore être approfondie par une étude de zonage géographique plus fine pour affiner les estimations et limiter leur volatilité.

L'étude de sensibilité sur les périodes de retour estimés par les coutiers a montré l'impact important sur la prime pure et donc sur le tarif. Les pistes pour estimer les périodes de retour pourraient être d'analyser les données sur la distribution des vitesses de vent par station météorologique et les corrélations entre ces stations ou d'influer sur les paramètres météorologiques des tempêtes historiques (trajectoires, déplacement et vitesse du courant-jet ...). Pour cela, il serait nécessaire alors de se rapprocher des experts météorologiques.

Les réflexions sur ce mémoire pourraient contribuer à des études complémentaires sur les programmes de réassurances en ajustant des priorités et des plafonds différents et d'en connaître le tarif économique et le capital requis (SCR) du risque catastrophes naturelles dans le cadre de solvabilité 2.

Bibliographie

Daniel DESCHEEMAEKERE et Bernard PERRON (1996) : « Les événements naturels en France : étude du risque tempête pour une compagnie d'assurances »

Nataliya DRAGIEVA (2008) : « Construction d'un intervalle de confiance par la méthode de bootstrap et test de permutation », 54-61

Bauwens V. et de Walhin JF Larcier (2008) : « La titrisation du risque d'assurance », *Larcier*

Jacques BLONDEAU et Christian PARTRAT (2003) : « La réassurance Approche technique », *Economica*

Michel DENUIT et Arthur CHARPENTIER (2005) : « Mathématiques de l'assurance non-vie TomeII : tarification et provisionnement », *Economica*, 70-177

Ministère de l'Écologie, du Développement durable et de l'Énergie (Mai 2013) : « Collection Prévention des risques naturels : Les tempêtes », http://www.developpement-durable.gouv.fr/IMG/pdf/Tempetes_16-05-2013.pdf.

Présentation (18/19 mars 2013 à Tours) sur « Les innovations face au développement des catastrophes non-vie »

« Les tempêtes en Europe : un risque en expansion » n°91 Septembre 2012 des cahiers de l'assurance de RISQUES

Sites de météorologie :

www.meteofrance.com

www.alertes-meteo.com

www.passion-meteo.net

www.meteonet.org

Site scientifiques :

www.cnrs.fr