

Mémoire présenté devant l'ENSAE Paris
pour l'obtention du diplôme de la filière Actuariat
et l'admission à l'Institut des Actuaires
le 09/03/2023

Par : **Aya Slimen**

Titre : **Amélioration de la modélisation catastrophes
naturelles dans un portefeuille de réassurance**

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

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

Membres présents du jury de la filière

Nom :

Christian-Yann Robert

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

Entreprise :

SCOR
The Art & Science of Risk

Signature :

Directeur du mémoire en entreprise :

Nom : Feta Ramadani

Signature :

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

Secrétariat :

Bibliothèque :

Signature du responsable entreprise

Signature du candidat

Remerciements

Je saisis cette occasion pour exprimer ma reconnaissance à tous ceux qui ont contribué au bon déroulement de mon expérience chez SCOR dans un environnement professionnel et de qualité.

Je tiens à témoigner ma gratitude à Mr. Feta Ramadani et à Mme. Dulcy Gninghayé pour leurs disponibilités, leurs encouragements et pour m'avoir proposé un sujet aussi intéressant à explorer. J'exprime aussi toute ma reconnaissance à Mr. Henry Bovy pour son partage d'expérience et de connaissances pendant les différentes phases du projet.

Je souhaite également remercier tous les membres des équipes Cat chez SCOR de m'avoir bien accueillie et pour l'ambiance agréable qu'ils entretiennent au quotidien.

Je remercie mon tuteur académique M. Christian-Yann Robert ainsi que tout le corps professoral de l'ENSAE pour la qualité de la formation qu'ils m'ont offerte.

Finalement, je remercie toute ma famille, amis et tous mes proches qui m'ont soutenue durant toutes mes années d'études.

Note de synthèse

Contrairement à une majorité des risques en assurance, les catastrophes naturelles sont caractérisées par une faible fréquence et une forte sévérité. Les pertes engendrées peuvent fortement impacter un (ré)assureur. Une modélisation adéquate s'avère donc nécessaire afin d'y faire face.

Pour ce faire, les organismes d'assurance font appel aux modèles dits « vendeurs » pour évaluer leurs pertes potentielles. Ces modèles offrent une modélisation spécifique pour chaque *Region Peril*. Un *Region Peril* désigne l'association d'un péril et d'une région de survenance.

Ils fournissent en sortie un catalogue d'évènements dans une table appelée ELT (*Event Loss Table*) où on trouve une liste d'évènements simulés définis par leur identifiants et une probabilité annuelle d'occurrence.

Pour l'étude d'un contrat ou du portefeuille global, il suffit de simuler ces évènements sur $N = 100\,000$ années de simulation et on se sert des ordres de quantiles de pertes de chaque scénario (année x date x évènement) pour calculer la perte associée. L'agrégation des pertes de ces scénarios par année de simulation permet de créer un YLT (*Year Loss Table*) qui regroupe la liste des scénarios possibles sur les années de simulations et leurs pertes estimées.

Cette approche par exposition permet d'avoir une vue globale sur le risque puisqu'elle permet, grâce au YLT, de calculer la perte moyenne annuelle agrégée et d'autres mesures du risque comme l'OEP (*Occurrence Exceedance Probability*) et l'AEP (*Aggregate Exceedance Probability*).

Cependant, la couverture des modèles « vendeurs » est limitée géographiquement. Il existe des *Region Perils* couverts dans le portefeuille pour lesquels on ne trouve pas de modèles disponibles sur le marché.

Dans un contexte d'accumulation, on souhaite avoir une vue par scénario pour tous les *Region Perils* couverts dans le portefeuille afin de pouvoir effectuer des agrégations de pertes sur différents niveaux de granularité (par pays, par type de contrat...).

C'est ainsi qu'a été élaboré à SCOR en 2019 le *Synthetic Modelling* qui sert à offrir une vue par scénario des pertes associées aux contrats qui couvrent des *Region Perils* non modélisés en utilisant des tables de référence qui sont les YEQTs (*Year Event Quantile Table*). Il s'agit de créer des tables prédéfinies pour chaque *Region Peril* non modélisé qui sont utilisées pour discrétiser les distributions de fréquence-sévérité paramétrées par expérience au moment de la tarification des contrats. Pour chaque *Region Peril* on génère sur les $N = 100\,000$ années de simulation des scénarios avec des identifiants d'évènements spécifiques à ce *Region Peril* et tels que les scénarios sont ordonnés par ordre de perte décroissante.

Cette simulation se fait de manière que pour chaque année de simulation le nombre d'évènements suit une loi Poisson de paramètre $\lambda = 2$ (sauf quelques exceptions où $\lambda = 4$), la date et l'ordre de sévérité de scénarios sont aléatoires.

Grâce à cette méthode on dispose d'un catalogue d'évènements unique pour chaque *Region Peril* non modélisé qui permet ensuite de déduire les YLTs des contrats qui les couvrent.

L'objectif de ce mémoire est d'améliorer la représentation des catastrophes naturelles dans le portefeuille en intervenant sur les YEQTs issus du *Synthetic Modelling* et sur la liaison entre ces tables. Le projet peut donc se résumer sur 3 axes :

1. **Étude de la fréquence $\lambda = 2$ utilisée dans le *Synthetic Modelling*** : En effet, dans un contexte de changement climatique, il est intéressant de revoir cette hypothèse qu'est le nombre moyen d'évènements annuel de sinistres.
2. **Étude de la saisonnalité des catastrophes naturelles** pour que les dates des scénarios générés dans les YEQTs des *Region Perils* non-modélisés reflètent bien la nature du phénomène naturel en terme de nombre d'évènements par mois et de sévérité des sinistres qui arrivent dans certains mois par rapport aux autres.
3. **Étude de la corrélation entre différents *Region Perils*** : Étant donné que les identifiants d'évènements présents dans les YEQTs des différents *Region Perils* lui sont uniques, au moment de l'agrégation de pertes par année de simulation entre différents périls, on suppose qu'ils sont indépendants. Il s'agit d'une hypothèse à revoir pour introduire la dépendance entre *Region Perils* différents quand c'est nécessaire.

Étude de la fréquence

Le choix du paramètre $\lambda = 2$ pour caractériser le nombre d'évènements par année de simulation était fait en se basant sur des données de tarifications historiques et représentait un choix conservateur en accumulation.

Pour revoir cette hypothèse, après 3 ans, on se muni des fréquences utilisées pour la tarification des contrats couvrant les catastrophes naturelles pour extraire les fréquences moyennes utilisées pour chaque *Region Peril* non-modélisé. De plus, on utilise le YLT du portefeuille d'étude qui caractérise les pertes associées tous les scénarios des différents *Region Perils*.

Deux cas de figures se présentent :

- La fréquence moyenne de tarification < 2 : Dans ce cas on suggère une réduction de la taille du YEQT associé. Cependant, ceci ne nous fait pas gagner en précision en termes de pertes associées au *Region Peril* concerné. On garde donc l'hypothèse $\lambda = 2$.
- La fréquence moyenne de tarification > 2 : L'impact de générer plus de scénarios dans les YEQTs de ces *Region Perils* est aussi faible. En effet, cette modification ne concerne que quelques périls dont la part du portefeuille est négligeable (moins de 1% de la perte totale).

Même si l'historique de tarification suggère quelques modifications au niveau de la fréquence annuelle moyenne des sinistres pour certains *Region Perils*, on n'effectue pas ce changement vu l'absence d'impact sur nos études d'accumulation.

Étude de la saisonnalité

À partir de données historiques de sinistres, on se propose d'évaluer la saisonnalité d'évènements des *Region Perils* non-modélisés.

La faible fréquence des catastrophes naturelles limite le nombre de données historiques disponibles par *Region Peril*. Ainsi, on étudie la saisonnalité au niveau de clés plus globales : Péril x Sous-continent x Type de Climat. Chaque clé regroupe ainsi plusieurs *Region Peril* qui partagent le même péril et la même présence géographique de risque.

On agrège les pertes historiques pour chaque clé par mois et on calcule deux vecteurs :

- *Props* : La proportion du nombre d'évènements de chaque mois de l'année par rapport au nombre total d'évènements.

— *Sevs* : La proportion des pertes associées aux évènements de chaque mois de l'année par rapport à la perte totale.

On sélectionne ensuite les clés qui présentent une saisonnalité (vecteurs *props* et *sevs* non constants).

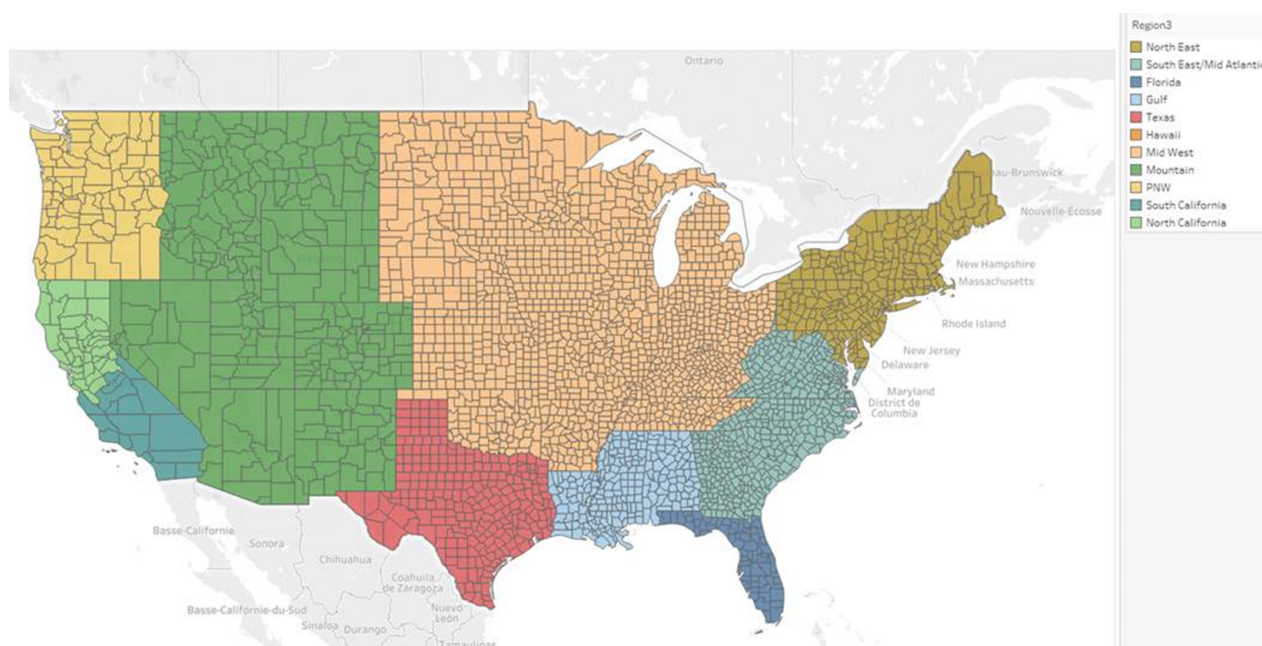
On introduit la saisonnalité dans notre modélisation en appliquant ce qui est observé pour les *Region Perils* modélisés.

La prise en compte de la saisonnalité des évènements dans le YEQT (d'un point de vue nombre et intensité) a un impact important sur la distribution des pertes sur l'année.

Étude de la corrélation

À partir de données historiques de sinistres, on commence par explorer la dépendance des sinistres extrêmes entre couples de clés. Cependant, ce niveau d'étude induit une surestimation de risque en introduisant la structure de corrélation observée sur une clé aux *Region Perils* concernés. On se propose donc de réduire notre étude géographiquement et de l'effectuer sur les périls aux États-Unis.

Pour ce faire, on dispose de pertes marché historiques de sinistres aux différents États des États-Unis. On regroupe les États sous 10 groupes différents :



Groupes des États aux États-Unis

On agrège les pertes par année pour chaque péril dans chaque groupe d'États et on se concentre sur la dépendance des couples de périls dans un même groupe.

En évaluant les mesures de dépendance entre couples, on détecte une dépendance entre les périls grêle et cyclone tropical dans les États du *South East/ Mid Atlantic* qu'on modélise en utilisant la copule Franck.

L'implémentation de cette structure de dépendance dans un contrat qui couvre ces deux périls montre qu'on a une augmentation de la perte maximale agrégée pour les différentes périodes de retour et que cette variation est plus visible pendant les basses périodes de retour.

Conclusions

Ce mémoire a permis d'explorer plusieurs axes d'amélioration de la modélisation des catastrophes naturelles dans un portefeuille de réassurance.

Pour les *Region Perils* non-modélisés, notre étude de la fréquence de modélisation ne suggère pas de modification pertinente pour l'entreprise contrairement à l'étude de saisonnalité qui change notre vue du risque en la considérant.

L'étude de corrélation sur l'exemple des États-Unis a montré l'importance de considérer la dépendance entre périls dans la modélisation pour éviter de sous-estimer le risque. Cette étude se poursuit pour d'autres couples de périls dans d'autres zones géographiques du monde.

Executive summary

Unlike most insurance risks, natural disasters are characterized by low frequency and high severity. The losses generated can strongly impact a (re)insurer. An adequate modeling is therefore necessary to face them.

To do this, insurance organizations use models known as vendor models to evaluate their expected losses. These models offer a specific modeling for each Region Peril. A Region Peril is defined by the combination of a peril and a region of occurrence.

These models provide as output a catalog of events in an table called ELT (*Event Loss Table*). For the study of a contract or of the global portfolio, it is sufficient to simulate these events over $N = 100\,000$ years of simulation and we use the orders of quantiles of losses for each scenario (year x date x event) to calculate the associated loss. Aggregating the losses of these scenarios by simulation year allows us to create a YLT (*Year Loss Table*) which gathers the list of possible scenarios over the simulation years and their estimated losses.

This exposure-based approach provides a global view of risk since it allows, thanks to the YLT, to calculate the aggregated annual average loss and other risk measures such as OEP (Occurrence Exceedance Probability) and AEP (Aggregate Exceedance Probability).

However, the coverage of the vendor models is limited geographically. There are some Region Perils covered in the portfolio for which there are no models available on the market.

In an accumulation context, we wish to have a view by scenario for all *Region Perils* covered in the portfolio in order to be able to perform loss aggregations on different levels of granularity (by country, by contract type...).

This is how the Synthetic Modelling was developed by SCOR in 2019 in order to provide a scenario-based view of losses associated with contracts that cover unmodeled *Region Perils* using reference tables that are YEQTs (*Year Event Quantile Tables*). This involves creating predefined tables for each unmodeled *Region Peril* that are used to discretize experimentally parameterized frequency-severity distributions at contract pricing. For each *Region Peril*, scenarios with event identifiers specific to that *Region Peril* are generated over the $N = 100\,000$ simulation years and such that the scenarios are ordered by decreasing loss.

This simulation is done in such a way that each year the number of events follows a Poisson distribution with a parameter of $\lambda = 2$ (with a few exceptions where $\lambda = 4$), the date and the order of severity of the scenarios are random.

This method allows us to have a unique catalog of events for each non-modelled *Region Peril* which then allows us to deduce the YLTs of the contracts that cover them

The objective of this thesis is to improve the representation of natural disasters in the portfolio by intervening on the YEQTs resulting from the *Synthetic Modelling* and on the link between these tables. The project can thus be summarized on 3 axes :

1. **Study of the frequency $\lambda = 2$ used in the *Synthetic Modelling*** : Indeed, in the

context of climate change, it is interesting to review this assumption which is the average number of annual loss events.

2. **Study of the seasonality of natural disasters** so that the dates of the scenarios generated in the YEQTs of the non-modelled Region Perils reflect well the nature of the natural phenomenon in terms of the number of events per month and the severity of the disasters that occur in certain months compared to others.
3. **Study of the correlation between different Region Perils** : Since the event identifiers present in the YEQTs of different Region Perils are unique to it, when aggregating losses by simulation year between different perils, they are assumed to be independent. This is an assumption to be revised to introduce dependence between different Region Perils when necessary.

Frequency Study

The choice of the parameter $\lambda = 2$ to characterize the number of events per simulation year was made based on historical pricing data and represented a conservative choice in accumulation. To revise this assumption, after 3 years, we use the frequencies used for the pricing of contracts covering natural catastrophes to extract the average frequencies used for each non-modeled region. In addition, we use the YLT of the study portfolio which characterizes the losses associated with all the scenarios of the different Region Perils.

There are two possible situations :

- The average pricing frequency < 2 : In this case we suggest a reduction in the size of the associated YEQT. However, this does not make us more accurate in terms of the losses associated with the concerned *Region Peril*. We therefore keep the assumption $\lambda = 2$.
- The average pricing frequency > 2 : The impact of generating more scenarios in the YEQTs of these perils is also small. This is because this change affects only a few region perils whose share of the portfolio is small (less than 1% of the total loss).

Although the pricing history suggests some changes in the average annual frequency of claims for some Region Perils, we do not make this change given the lack of impact on our accumulation studies.

Seasonality study

Using historical loss data, we propose to evaluate the seasonality of non-modelled Region Perils. The low frequency of natural disasters limits the amount of historical data available per Region Peril. Thus, seasonality is studied at the level of more global keys : Peril x Subcontinent x Climate Type. Each key thus groups together several Region Peril that share the same peril and the same geographical presence of risk.

We aggregate the historical losses for each key by month and compute two vectors :

- *Props* : The proportion of the number of events in each month of the year to the total number of events.
- *Sevs* : The proportion of losses associated with events in each month of the year to the total loss.

We then select the keys that show seasonality (non-constant *props* and *sevs* vectors).

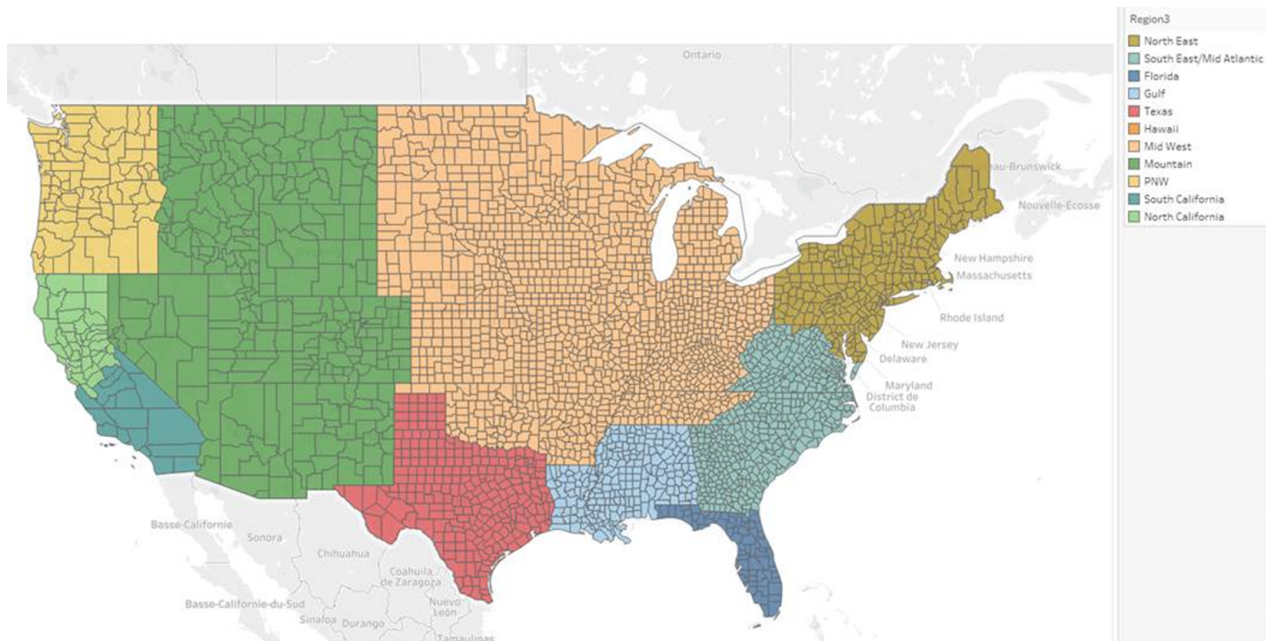
Seasonality is introduced into our modeling by applying what is observed for the modeled Region Perils.

Taking into account the seasonality of events in the YEQT (in terms of number and intensity) has a significant impact on the distribution of losses over the year.

Correlation study

Using historical loss data, we begin by exploring the dependence of extreme losses between pairs of keys. However, this level of study induces an overestimation of risk by introducing the correlation structure observed on a key to the concerned *Region Perils*. We therefore propose to reduce our study geographically and perform it on perils in the United States.

To do this, historical market losses of claims in different states of the United States are available. The states are grouped into 10 different groups :



Groups of States in the United States

Losses by year for each peril in each group of states are aggregated and the focus is on the dependence of pairs of perils within a group.

Evaluating the pairwise dependence measures, we detect a dependence between convective storm and tropical cyclone perils in the South East/ Mid Atlantic states that we model using the Franck copula.

Implementing this dependency structure in a contract that covers both perils shows that there is an increase in the aggregate maximum loss for the different return periods and that this variation is more visible at low return periods.

Conclusions

This dissertation explored several areas of improvement in modeling natural disasters in a reinsurance portfolio.

For the non-modelled Region Perils, our modelling frequency study does not suggest a relevant change for the company unlike the seasonality study which changes our view of the risk if we consider it.

The correlation study on the example of the United States has shown the importance of considering the dependence between perils in the modeling to avoid underestimating the risk. This study is continuing for other pairs of perils in other geographical areas of the world.

Resumé

Les portefeuilles de réassureurs sont de plus en plus exposés aux pertes causées par des événements de catastrophes naturelles dans le contexte actuel de changement climatique. En effet, le nombre de ces événements continue à augmenter ainsi que leurs intensités.

Afin d'évaluer leurs expositions en cas de sinistres cat-nat, les réassureurs se basent sur des modèles qu'on appelle « vendeurs ». Il s'agit de modèles probabilistes qui offrent un catalogue d'événements cat en répliquant des événements physiques possibles pour chaque péril dans chaque région.

Même si ces fournisseurs de modèles offrent une modélisation pour plusieurs périls dans le monde, certains périls ne sont toujours pas modélisés.

Dans un souci de cohérence et afin d'effectuer des études d'accumulation de risques cat, on a besoin de représenter les périls non-modélisés de la même manière que les périls modélisés par les modèles « vendeurs ».

On se base ainsi sur des données historiques de sinistres et de tarification afin d'améliorer la représentation des périls non-modélisés dans notre portefeuille de réassurance. En effet, on commence par étudier la fréquence annuelle d'occurrence de ces événements cat. Ensuite, une étude de la saisonnalité des événements catastrophes naturelles permet de voir l'impact de cette dimension temporelle sur la vue du risque.

Finalement, une étude de la corrélation historique entre les différents périls (modélisés et non-modélisés) permet d'évaluer leurs dépendances et de les introduire dans la modélisation du portefeuille pour éviter une sous-estimation des pertes.

Mots-clés : Réassurance, Catastrophes Naturelles, *Region Perils*, modélisation, YEQT, fréquence, saisonnalité, corrélation, copule.

Abstract

Reinsurers' portfolios are increasingly exposed to losses caused by natural catastrophe events in the current context of climate change. Indeed, the number of such events continues to increase as well as their intensity.

In order to assess their exposure to cat-nat claims, reinsurers rely on so-called "vendor models". These are probabilistic models that offer a catalog of cat events by replicating possible physical events for each peril in each region.

Although these model vendors offer modeling for many perils around the world, some perils are still not modeled.

For consistency and in order to perform cat hazard accumulation studies, we need to represent the unmodeled perils in the same way as the perils modeled by the "vendor" models.

Thus, we use historical claims and pricing data to improve the representation of non-modeled perils in our reinsurance portfolio. Indeed, we start by studying the frequency of occurrence of these cat events. Then, a study of the seasonality of natural catastrophe events allows us to see the impact of this temporal dimension on the risk view.

Finally, a study of the historical correlation between the different perils (modeled and non-modeled) allows us to evaluate their dependencies and to introduce them into the modeling of the portfolio in order to avoid an underestimation of the losses.

Keywords : Reinsurance, Natural Disasters, Region Peril, Modelling, YEQT, frequency, seasonality, correlation, copula.

Table des matières

Note de synthèse	2
Executive summary	6
Resumé	9
Abstract	10
Introduction	16
1 Réassurance et catastrophes naturelles	17
1.1 Réassurance des catastrophes naturelles	17
1.1.1 Définition et particularités du risque catastrophe naturelles	17
1.1.2 Défis de la branche des catastrophes naturelles	17
1.2 Construction de la représentation des catastrophes naturelles en réassurance	18
1.2.1 Les <i>Region Perils</i>	18
1.2.2 Modélisation des <i>Region Perils</i>	19
1.3 Modèles cat-nat	20
1.3.1 Composantes	20
1.3.2 Sorties du modèle	21
1.4 <i>Synthetic Modelling</i>	25
1.4.1 Principe de la méthode	25
1.4.2 Déroulement de l'accumulation	26
1.4.3 YLT combinée	26
1.5 Objectifs du mémoire	27
2 Étude de la fréquence	28
2.1 Hypothèses et fonctionnement de la méthode actuelle	28
2.2 Présentation des données - <i>Loss Engines</i>	30
2.3 Extraction de la fréquence et observations	30
2.4 Implémentation et conclusions	32
2.4.1 Cas où la fréquence moyenne < 2	32
2.4.2 Cas où la fréquence moyenne > 2	33
2.4.3 Conclusions	35
3 Étude de la saisonnalité	36
3.1 Présentation des données utilisées	36
3.1.1 Données historiques de sinistres CAT de SCOR	36
3.1.2 Base de données du CRED	36
3.1.3 Défis de la donnée et choix de modélisation	37
3.2 Calculs des vecteurs de saisonnalité	38

3.2.1	Extraction de la saisonnalité	39
3.2.2	Moyenne Mobile Pondérée sur les vecteurs de saisonnalités	40
3.3	Implémentation de la saisonnalité	41
3.4	Impact de l'introduction de la saisonnalité sur un contrat	43
3.4.1	Avant termes et conditions	44
3.4.2	Après termes et conditions	45
3.4.3	Conclusions	47
4	Étude de la corrélation	48
4.1	Corrélation inter <i>Region Perils</i> dans le monde	48
4.1.1	Principe préliminaire	48
4.1.2	Résultats	49
4.2	Corrélation inter périls aux États-Unis	51
4.3	Théorie des copules	52
4.3.1	Copules bivariées	52
4.3.2	Mesures de dépendance	54
4.3.3	Dépendances de queues	55
4.3.4	Copules archimédiennes	56
4.3.5	Inférence statistique	56
4.4	Application sur les périls des États-Unis	58
4.4.1	Choix des couples pertinents	58
4.4.2	Estimation des copules	59
4.4.3	Implémentation et étude d'impact	62
	Conclusion	67
	Bibliographie	67
	Liste des abréviations	69
	A Preuve du lemme 1	70
	B Table des valeurs critiques du coefficient de corrélation de rangs de Spearman	71
	C Autres méthodes d'estimation de copules	73

Table des figures

1.1	Couverture géographique des périls par RMS	19
1.2	Composition du portefeuille étudié en fonction de la modélisation du péril	20
1.3	Composantes d'un modèle cat-nat	21
1.4	Courbes AEP et OEP du portefeuille d'étude	24
1.5	Déroulement de l'accumulation pour les <i>Region Perils</i> non-modélisés	26
2.1	Lien entre fréquence de tarification f et les YEQTs prédéfinis ($\lambda = 2$)	29
2.2	Distribution fréquence pour le loss engine du tableau 2.2	30
2.3	Diagrammes en moustaches des moyennes des fréquences observées des <i>Region Perils</i>	31
2.4	Passage de fréquence 2 à $\lambda < 2$	32
2.5	Passage de fréquence 2 à $\lambda > 2$	34
3.1	Évolution du nombre de données de pertes financières disponibles par rapport au nombre de catastrophes naturelles	37
3.2	Climat des pays du monde selon Köppen-Geiger	38
3.3	Exemples de vecteurs de saisonnalités observés pour le tremblement de terre et le cyclone tropical en climat tropical	39
3.4	Vecteurs de saisonnalité pour la grêle et les inondations aux pays de l'Amérique du nord à climat continental (Canada)	41
3.5	Saisonnalité observée pour <i>Region Perils</i> modélisés	42
3.6	Nombre d'évènements pas mois par <i>Region Peril</i> dans le YLT initial du contrat	43
3.7	Nombre d'évènements pas mois par <i>Region Peril</i> dans le YLT du contrat après introduction de la saisonnalité	44
3.8	Distribution initiale de pertes sur l'année par <i>Region Peril</i>	44
3.9	Distribution de pertes sur l'année par <i>Region Peril</i> après introduction de la saisonnalité	45
3.10	Distribution initiale des pertes sur l'année par <i>Region Peril</i> après T&C	46
3.11	Distribution de pertes sur l'année par <i>Region Peril</i> avec saisonnalité et après T&C	46
4.1	Nuage des points des rangs d'années d'évènements pour la grêle et les inondations aux pays de l'Amérique du nord au climat Tempéré	49
4.2	Dispersion géographique des <i>Region Perils</i> dépendants	50
4.3	Groupes des États aux États-Unis	52
4.4	Exemples de dépendogrammes de copules archimédiennes	57
4.5	Copules archimédiennes estimées à partir des observations	60
4.6	Distribution du paramètre estimé d'une copule Franck indépendante avec $N = 55$ observations	61
4.7	Distribution du paramètre estimé d'une copule Clayton indépendante avec $N = 55$ observations	62
4.8	Impact sur l'AEP et l'OEP après introduction de la dépendance entre les périls	65
4.9	Impact sur l'AEP après introduction de la dépendance entre les périls et après TC	66

Liste des tableaux

1.1	Liste des périls	18
1.2	Liste des sous-continentes	19
1.3	Exemple d'un ELT d'un <i>Region peril</i> modélisé	21
1.4	Partie de la YEQT pour les 3 évènements de l'ELT du tableau 1.3	22
1.5	Partie du YLT pour les 3 évènements de l'ELT du tableau 1.3	23
1.6	Exemple réduit du YEQT d'un <i>Region peril</i> synthétique	25
1.7	Partie de la YLT combinée du portefeuille d'étude	26
1.8	Partie du <i>header</i> lié au YLT du portefeuille d'étude	27
2.1	Fréquences historiques utilisées pour la tarification	28
2.2	Exemple d'un loss engine	30
2.3	Variation des pertes en réduisant les fréquences par péril	32
2.4	Variation des pertes en réduisant les fréquences par péril	34
2.5	Variations de fréquences >2 et impacts sur le portefeuille	34
3.1	Zones climatiques selon Köppen-Geiger	38
3.2	Nombre de <i>Region Perils</i> avec saisonnalité	40
3.3	Termes et conditions et généralité sur le contrat	45
3.4	Pertes trimestrielles (en euros)	47
4.1	Couples de clé retenus avec dépendance aux extrêmes	50
4.2	Nombre de sinistres historiques par péril	51
4.3	Copules archimédiennes	56
4.4	Couples d'étude de dépendance	58
4.5	Estimation des paramètres des copules et critères quantitatifs	59
4.6	Tableau qui illustre le réarrangement fait sur une colonne du vecteur y d'observations	63
4.7	Réarrangement des observations et années de simulation	63

Introduction

Les compagnies de réassurance souscrivent des risques à l'échelle mondiale. Elles sont donc exposées aux différentes catastrophes naturelles qui peuvent toucher le monde.

Les pertes engendrées en cas d'occurrence de ces événements peuvent atteindre des chiffres énormes et causer des dégâts irréversibles à une compagnie de réassurance. Ainsi, il est important pour ces organismes d'étudier ces risques grâce à une modélisation adéquate.

Afin d'évaluer ces risques catastrophes naturelles, les réassureurs utilisent des modèles catastrophes « vendeurs » tels que RMS et AIR. Ces modèles permettent de simuler des événements naturels pour estimer leurs impacts sur le portefeuille.

Même si ces modèles représentent un grand nombre de périls naturels tels que les tremblements de terre aux États Unis, certains périls ne le sont pas encore.

Pour ce mémoire, l'objectif est, d'une part, d'améliorer les méthodes de modélisation pour les périls qui ne sont pas modélisés par les logiciels vendeurs. Pour ce faire, l'idée est de s'inspirer des sorties des modèles cat-nat pour en produire les mêmes pour les périls non modélisés. D'autre part, on a pour objectif d'améliorer la modélisation de la dépendance entre les différentes catastrophes naturelles couvertes dans le portefeuille.

En première partie, on présentera des notions générales sur la réassurance des catastrophes naturelles et les outils de modélisation actuels.

Ensuite, on s'attardera sur l'étude de la fréquence et de la saisonnalité des catastrophes naturelles afin d'apporter une amélioration à la modélisation actuelle des périls non-modélisés par des logiciels « vendeurs ».

Finalement, on se concentrera sur la dépendance entre les périls différents dans l'objectif de l'introduire dans la modélisation actuelle.

Chapitre 1

Réassurance et catastrophes naturelles

Introduction

Une réassurance peut être définie comme « l'assurance de l'assureur ». Cette activité permet à une société d'assurance (la cédante) de céder à une société de réassurance une partie de son risque contre le règlement d'une prime de réassurance qui correspond au risque transféré. Par le biais de la réassurance, un assureur transfère une partie du risque qu'il porte et ainsi augmente sa capacité de souscription sans avoir besoin d'augmenter ses fonds propres.

Même si ce principe de transfert de risque est apparu pour la première fois en 1370 dans un contexte de commerce maritime vénitien, la réassurance moderne ne remonte qu'à la fin du *XIX^{me}* siècle en Allemagne suite à la révolution industrielle. Dès lors, plusieurs entreprises de réassurance ont été créées pour couvrir les marchés d'assurance mondiaux.

Ce marché connaît une forte évolution dans le monde. Le volume des primes de réassurance a augmenté de 60% entre 2010 et 2020 pour atteindre 320Mds\$. Parmi plusieurs sociétés de réassurance, SCOR occupe la 4^{eme} position après Munich Re, Swiss Re et Hannover Re en 2020 avec un chiffre d'affaires qui dépasse les 20 000Mds\$ [2].

Cependant, le marché de réassurance est confronté à des enjeux de plus en plus difficiles dans un contexte économique et social incertain étant donné la persistance des taux d'intérêts bas, la hausse de la sinistralité des catastrophes naturelles et l'impact de la pandémie covid-19.

1.1 Réassurance des catastrophes naturelles

1.1.1 Définition et particularités du risque catastrophe naturelles

En France, d'après l'article L125-1 du Code des assurances, une catastrophe naturelle est définie par « l'intensité anormale d'un agent naturel »[1].

En réassurance, on définit les catastrophes naturelles comme des événements de très faible fréquence et de forte sévérité. Ainsi, si le capital du réassureur exposé à ce risque n'est pas assez protégé, il peut être fortement impacté par des événements de catastrophes naturelles de façon irréversible. La bonne modélisation et l'étude de ces risques est primordiale aux organismes de réassurance.

1.1.2 Défis de la branche des catastrophes naturelles

Une modélisation adéquate des catastrophes naturelles favorise une meilleure tarification de la composante CAT des traités de réassurance et une meilleure gestion du portefeuille.

Cependant, l'estimation des pertes causées par des événements de catastrophes naturelles diffère de celle des autres risques couverts en assurance dont l'approche est plutôt fondée sur la

base des données de sinistres passés. Cette approche admet des limites concernant les risques cat-nat.

En effet, l'aspect basse fréquence d'occurrence des catastrophes naturelles ainsi que le faible horizon d'historique de données disponible rendent difficile la modélisation à partir des données historiques.

De plus, même si on avait des bases de données historiques suffisantes sur les sinistres catastrophes naturelles, le passé n'est pas forcément représentatif du futur vu que plusieurs facteurs changent. D'une part, les catastrophes naturelles peuvent être considérées comme risques liés à des manifestations climatiques qui sont en variation vu le changement climatique qui engendre une augmentation de la fréquence et de la sévérité de ces événements. D'autre part, les pertes sont fonction de la modernisation et du développement socio-économique. Plus précisément, le montant des pertes varie avec le nombre et la vulnérabilité des constructions touchées par le sinistre cat-nat d'où une variation des montants de pertes au fil des années.

1.2 Construction de la représentation des catastrophes naturelles en réassurance

1.2.1 Les *Region Perils*

La modélisation des catastrophes naturelles est spécifique à chaque péril et chaque région. C'est pour cette raison que la modélisation et l'estimation de pertes des sinistres catastrophes naturelles sont effectuées pour chacun de ces *Region perils*, par exemple les tremblements de terre aux États-Unis (USEQ).

Ci-dessous, la liste des périls chez SCOR :

Abréviation	Nom Du Péril	Signification (FR)
CS	Convective Storm	Grêle
DR	Drought	Sécheresse
EQ	Earthquake	Tremblement De Terre
ET	Extra Tropical Cyclone	Cyclone Extra-tropical
FL	Flood	Inondation
OP	Other Perils	Autres Périls
OW	Other Wind	Autres Périls Causés Par Le Vent
SP	Snow Pressure	Pression De Neige
TC	Tropical Cyclone	Cyclone Tropical
VO	Volcano	Volcan
WF	Wildfire	Feu De Forêt
WS	Windstorm	Tempête De Vent
WT	Winter Storm	Tempête De Neige

TABLE 1.1 – Liste des périls

Si on compte tous les périls pour chaque pays (226 pays), on se retrouve avec plus de 2400 *Region Perils*. Ce chiffre représente un grand nombre de *Region Perils* à étudier. D'autant plus que certains *Region Perils* n'ont pas essentiellement une signification physique ou ne représentent pas de risque sur le marché de réassurance.

Ainsi, les *Region Perils* sont réduits à 1042 qui permettent de partager les catastrophes naturelles dans le monde d'une manière significative et de bien représenter les périls associés aux différents pays d'où la présence des périls OP et OW.

Ces *Region Perils* peuvent être aussi regroupés en fonction d’une variable de localisation qui est le sous-continent. La liste des sous-continents est donnée dans le tableau 1.2 ci-dessous. Cette catégorisation par sous-continent permet de voir comment les périls sont distribués dans le monde.

Sous-continents		
Australia and New Zealand	Northern Africa	Southern Asia
Central Asia	Northern America	Southern Europe
Eastern Asia	Northern Europe	Sub-Saharan Africa
Eastern Europe	Pacific Islands	Western Asia
Latin America	South-eastern Asia	Western Europe

TABLE 1.2 – Liste des sous-continents

1.2.2 Modélisation des *Region Perils*

Parmi la liste des *Region Perils* couverts, on distingue deux groupes :

- ***Region Perils* modélisés** : Il s’agit des *Region Perils* pour lesquels SCOR utilise les modèles fournis par les vendeurs des modèles cat-nat tels que RMS et AIR.

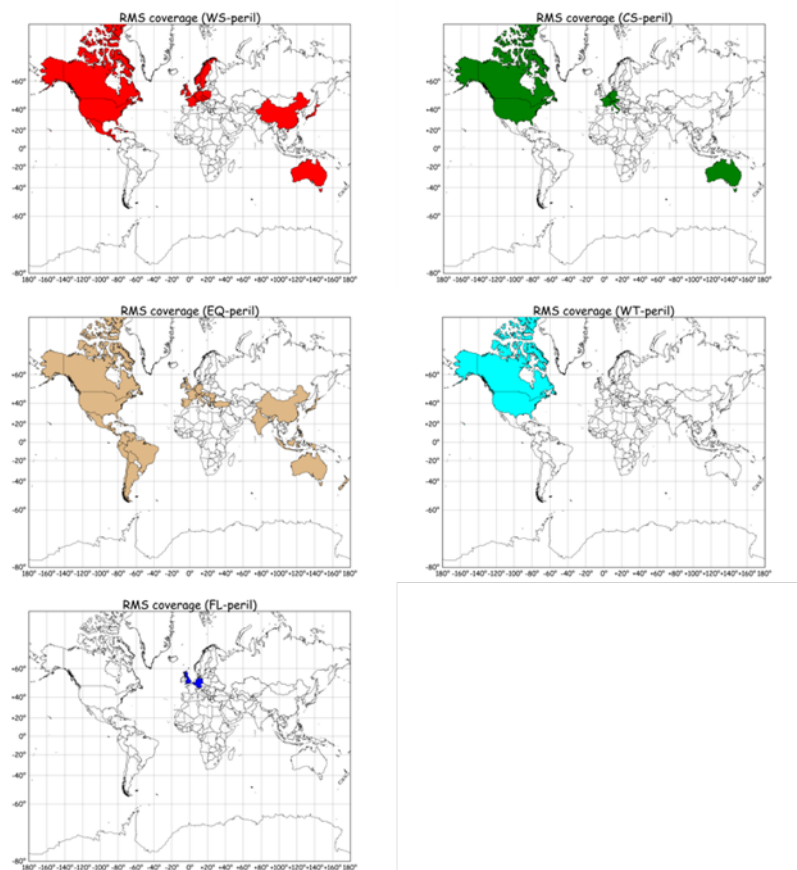


FIGURE 1.1 – Couverture géographique des périls par RMS

- ***Region Perils* synthétiques** : Ce sont les *Region Perils* qui ne sont pas modélisés par des modèles disponibles sur le marché.

Même si le nombre des *Region Perils* modélisés chez SCOR ne représente que 6% du nombre total des *Region Perils*, ceux-ci représentent environ 80% des pertes du portefeuille SCOR.

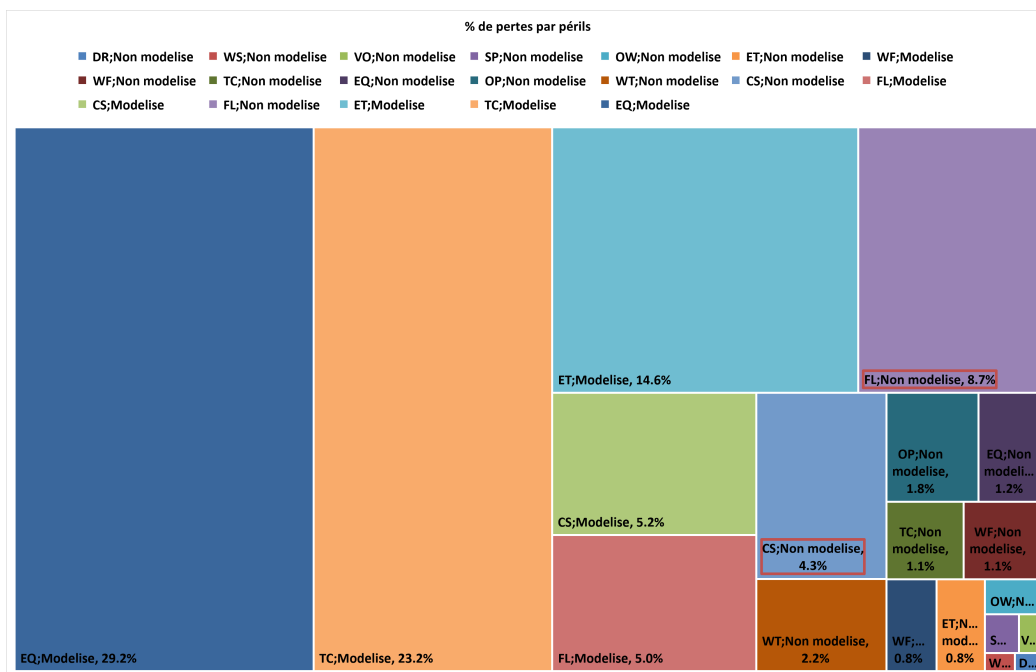


FIGURE 1.2 – Composition du portefeuille étudié en fonction de la modélisation du péril

On remarque qu’il y a des périls couverts dans le portefeuille qui ne sont pas modélisés par les modèles cat-nat tels que la sécheresse (DR), la pression de la neige (SP), les volcans (VO), les tempêtes de vent et de neige (WS et WT).

Même si ces périls ne sont pas modélisés dans le portefeuille, ils restent relativement moins importants par rapport aux *Region Perils* non-modélisés pour des périls tels que l’inondation (FL) et la grêle (CS).

1.3 Modèles cat-nat

1.3.1 Composantes

Un modèle cat-nat est une représentation simplifiée de la réalité. Il repose sur quatre modules clés pour permettre à un (ré)assureur d’évaluer le risque qu’il a dans un *Region peril* défini [11] :

- **Module Aléa** : Où, avec quelle fréquence et avec quelle intensité les événements catastrophes naturelles se produisent ?
En fonction de la localisation géographique, il permet de calculer la fréquence et l’intensité du risque en constituant un catalogue d’événements probables. Ces événements sont générés sur la base d’événements antérieurs et une connaissance approfondie des propriétés physiques des forces de la nature considérées.
- **Module Vulnérabilité** : Quel est le degré de dommage que peuvent subir les risques assurés en fonction de l’intensité de l’évènement ?
Ce module crée des courbes de vulnérabilité qui relient l’intensité physique de l’évènement (vitesse de vent, mouvement du sol...) à un taux de destruction en fonction des caractéristiques du bien (matériau de construction, nombre d’étages...).
- **Module Exposition** : Où se trouvent les biens qui sont couverts ?
Ce module fait le lien entre la sévérité des événements et la localisation des biens assurés pour aboutir au calcul de la perte brute à la suite d’un évènement.
- **Module Financier** : Quelle est la perte générée par les biens couverts en tenant compte

des conditions financières ?

Dans ce module, on applique les termes et conditions du contrat de couverture aux pertes brutes calculées afin d'aboutir à la charge finale pour l'assureur ou le réassureur.

Il faut ainsi quantifier ces modules en cas de sinistre catastrophe et les relier ensuite pour avoir un *Event Loss Table* (ELT) avec la perte nette moyenne de chaque évènement qui sera ensuite converti en *Year Loss Table* (YLT) qui contient les pertes par année de simulation et par évènement.

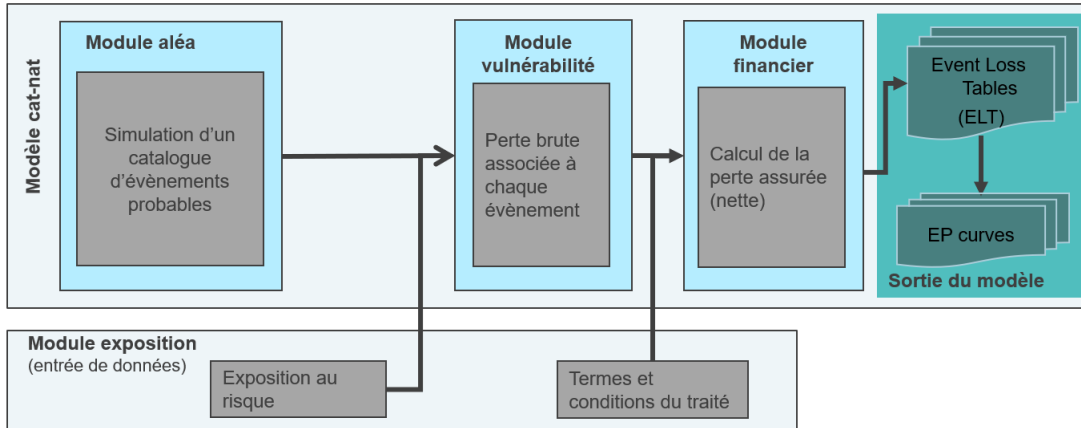


FIGURE 1.3 – Composantes d'un modèle cat-nat

1.3.2 Sorties du modèle

Les *Event Loss Tables* (ELT), *Yeart Event Quantile Tables* (YEQT) et *Year Loss Tables* (YLT)

En sortie de la modélisation d'un *Region Peril*, on obtient un tableau appelé l'ELT qui est un catalogue d'évènements simulés. Pour chaque évènement, on donne un identifiant, une probabilité annuelle d'occurrence, une perte nette moyenne et son écart type.

EventId	Loss \overline{m}_k	StdDev σ_k	ExpVal	Freq
2859413	2 490 152 649	3 383 274 750	228 300 024 092	0.008683591
2854000	10 713 770 140	6 317 251 490	1 144 033 982 897	0.007245175
2869738	988 623 735	1 680 694 584	1 222 058 672 887	0.004368581

TABLE 1.3 – Exemple d'un ELT d'un *Region peril* modélisé

Les évènements sont après simulés en utilisant leurs probabilités d'occurrence annuelle sur un nombre $N = 100\,000$ années de simulation. Le nombre d'années de simulation a été fixé pour le modèle interne SCOR et c'est le même nombre pour toutes les simulations de pertes dans les différentes branches pour garantir une agrégation consistante.

On associe à l'ELT une autre table qui accorde à chaque occurrence d'un évènement en une année de simulation un ordre de quantile. Cette table est spécifique pour chaque modélisation d'une catastrophe naturelle dans une localisation donnée. Elle est appelée le *Year Event Quantile Table* (YEQT).

Year	EventID	DateTime	Ordre p du quantile
30874	2859413	15/08/2022	0.5243
30163	2854000	24/02/2022	0.7589
38789	2859413	18/06/2022	0.9088
24472	2854000	19/02/2022	0.5803
50911	2854000	19/11/2022	0.8838
94772	2869738	05/09/2022	0.3908
88868	2869738	14/01/2022	0.5680
60868	2859413	02/01/2022	0.1581
73285	2859413	06/01/2022	0.7181
39430	2854000	20/10/2022	0.7956
37471	2869738	23/12/2022	0.4133
31314	2859413	17/02/2022	0.8965
71974	2859413	20/03/2022	0.0042
37902	2854000	02/04/2022	0.8326
43200	2859413	23/08/2022	0.7385
19510	2859413	13/05/2022	0.2276
85676	2854000	22/07/2022	0.5192
72371	2859413	18/12/2022	0.2276
861	2869738	23/12/2022	0.7759
81388	2854000	08/10/2022	0.1178

TABLE 1.4 – Partie de la YEQT pour les 3 évènements de l'ELT du tableau 1.3

Pour le calcul des montants des pertes, on commence par paramétrer une loi *Beta* qui représente les taux des dommages pour chaque évènement. Les paramètres α et β de cette loi, spécifique à chaque évènement, sont calculés en fonction de sa perte nette moyenne et de son écart type.

Définition 1 (Loi *Beta*). Soient $\alpha > 0$ et $\beta > 0$. La densité de la loi bêta de paramètres α et β est :

$$f_{\alpha,\beta}(x) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1} \mathbb{1}_{[0;1]}(x)$$

avec $\Gamma : t \mapsto \int_0^{+\infty} u^{t-1} e^{-u} du$ définie sur $]0; +\infty[$. L'espérance de cette loi est :

$$\frac{\alpha}{\alpha + \beta}$$

La variance de cette loi est :

$$\frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}.$$

Soit pour un évènement $k \in \{1; 2; \dots; n\}$ parmi les évènements présents dans l'ELT :

- \overline{m}_k la perte nette moyenne
- σ_k l'écart-type
- E_k l'exposition
- $\mu_k = \frac{\overline{m}_k}{E_k}$ l'espérance de la loi bêta de paramètre α_k et β_k ;
- $s_k = \frac{\sigma_k}{E_k}$ l'écart-type de la loi bêta de paramètre α_k et β_k .

En utilisant les formules de l'espérance et de la variance, nous pouvons calculer les paramètres α_k et β_k :

$$\begin{cases} \mu_k &= \frac{\alpha_k}{\alpha_k + \beta_k} \\ s_k^2 &= \frac{\alpha_k \beta_k}{(\alpha_k + \beta_k)^2 (\alpha_k + \beta_k + 1)} \end{cases} \iff \begin{cases} \alpha_k &= \mu_k \left(\frac{\mu_k (1 - \mu_k)}{\sigma_k^2} - 1 \right) \\ \beta_k &= \alpha_k \frac{1 - \mu_k}{\mu_k} \end{cases}$$

Ensuite, on utilise les ordres p de quantile de la YEQT, la loi *Beta* propre à chaque évènement pour déterminer le quantile de perte pour chaque occurrence d'un évènement dans une année de simulation dans le YEQT.

Year	EventID	DateTime	ordre p	α	β	quantile	ExpValue	Perte
30874	2859413	15/08/2022	0.5243	0.5249	47.5992	0.0058	3,29E+09	1,92E+07
30163	2854000	24/02/2022	0.7589	2.8400	300.4157	0.0125	5,50E+09	6,86E+07
38789	2859413	18/06/2022	0.9088	0.5249	47.5992	0.0308	3,29E+09	1,01E+08
24472	2854000	19/02/2022	0.5803	2.8400	300.4157	0.0094	5,50E+09	5,17E+07
50911	2854000	19/11/2022	0.8838	2.8400	300.4157	0.0161	5,50E+09	8,84E+07
94772	2869738	05/09/2022	0.3908	0.3449	426.0152	0.0001	1,66E+09	1,90E+05
88868	2869738	14/01/2022	0.5680	0.3449	426.0152	0.0004	1,66E+09	6,08E+05
60868	2859413	02/01/2022	0.1581	0.5249	47.5992	0.0005	3,29E+09	1,67E+06
73285	2859413	06/01/2022	0.7181	0.5249	47.5992	0.0129	3,29E+09	4,24E+07
39430	2854000	20/10/2022	0.7956	2.8400	300.4157	0.0133	5,50E+09	7,33E+07
37471	2869738	23/12/2022	0.4133	0.3449	426.0152	0.0001	1,66E+09	2,26E+05
31314	2859413	17/02/2022	0.8965	0.5249	47.5992	0.0287	3,29E+09	9,43E+07
71974	2859413	20/03/2022	0.0042	0.5249	47.5992	0.0000	3,29E+09	1,64E+03
37902	2854000	02/04/2022	0.8326	2.8400	300.4157	0.0143	5,50E+09	7,87E+07
43200	2859413	23/08/2022	0.7385	0.5249	47.5992	0.0140	3,29E+09	4,60E+07
19510	2859413	13/05/2022	0.2276	0.5249	47.5992	0.0010	3,29E+09	3,40E+06
85676	2854000	22/07/2022	0.5192	2.8400	300.4157	0.0086	5,50E+09	4,70E+07
72371	2859413	18/12/2022	0.2276	0.5249	47.5992	0.0010	3,29E+09	3,40E+06
861	2869738	23/12/2022	0.7759	0.3449	426.0152	0.0011	1,66E+09	1,87E+06
81388	2854000	08/10/2022	0.1178	2.8400	300.4157	0.0036	5,50E+09	1,98E+07

TABLE 1.5 – Partie du YLT pour les 3 évènements de l'ELT du tableau 1.3

Finalement, pour aboutir aux montants des dommages pour chaque évènement et chaque année de simulation, il suffit de multiplier le quantile trouvé par l'exposition.

On se retrouve ainsi avec un catalogue d'évènements qui regroupe les identifiants des évènements, des années de simulation et une perte pour chaque simulation d'évènement. On appelle cette table le *Year Loss Table* (YLT).

Les mesures de risque issues du YLT

À partir du YLT, on peut calculer la **moyenne du sinistre annuel attendu** (AAL : *Annual Aggregated Loss*) qui est donnée par la moyenne de toutes les pertes des sinistres du YLT.

Le YLT permet aussi d'avoir les **EP curves** (*Exceedance probability*) qui sont deux courbes donnant une meilleure visibilité sur le profil du risque, à savoir l'OEP (*Occurrence Exceedance Probability*) et l'AEP (*Aggregate Exceedance Probability*).

Avant de détailler ces deux courbes, on introduit la notion de **période de retour** d'une perte. Il s'agit d'une mesure donnant une idée sur la fréquence et la sévérité d'une perte. Elle est calculée telle que pour une perte i dans un YLT, $RP_i = \frac{N}{rang_i}$ où N est le nombre d'années de simulations et $rang_i$ c'est le rang de la perte i si on classe les pertes par ordre décroissant. Le résultat est lu tel que la perte i a une chance d'arriver tous les RP_i ans.

- **OEP** : c'est la probabilité que la perte maximale liée à un évènement au cours d'une année dépasse un certain montant.

Soient N la distribution de fréquence du modèle et X_1, \dots, X_N des variables aléatoires indépendantes et identiquement distribuées. Pour tout $i \in \{1; \dots; N\}$, X_i est la perte découlant de l'évènement i . L'OEP pour un seuil x est donnée par :

$$OEP(x) = \mathbb{P}(\max(X_1; \dots; X_N) > x)$$

Pour obtenir la courbe OEP, on calcule les pertes maximales pour chaque année, ensuite on calcule les périodes de retour des pertes maximales. Finalement, on trace les pertes maximales des années de simulation en fonction des périodes de retour.

- **AEP** : c'est la probabilité que la somme des pertes au cours d'une année dépasse un certain montant.

On considère les mêmes notations définies :

$$AEP(x) = \mathbb{P}\left(\sum_{i=1}^N X_i > x\right)$$

Pour tracer l'AEP, on commence par calculer les sommes des pertes du YLT par année de simulation, ensuite on calcule les périodes de retour pour les pertes agrégées. Finalement, on trace les pertes agrégées par années en fonction des périodes de retour.

On représente dans la figure 1.4 les courbes AEP et OEP du portefeuille d'étude avec une échelle changée.

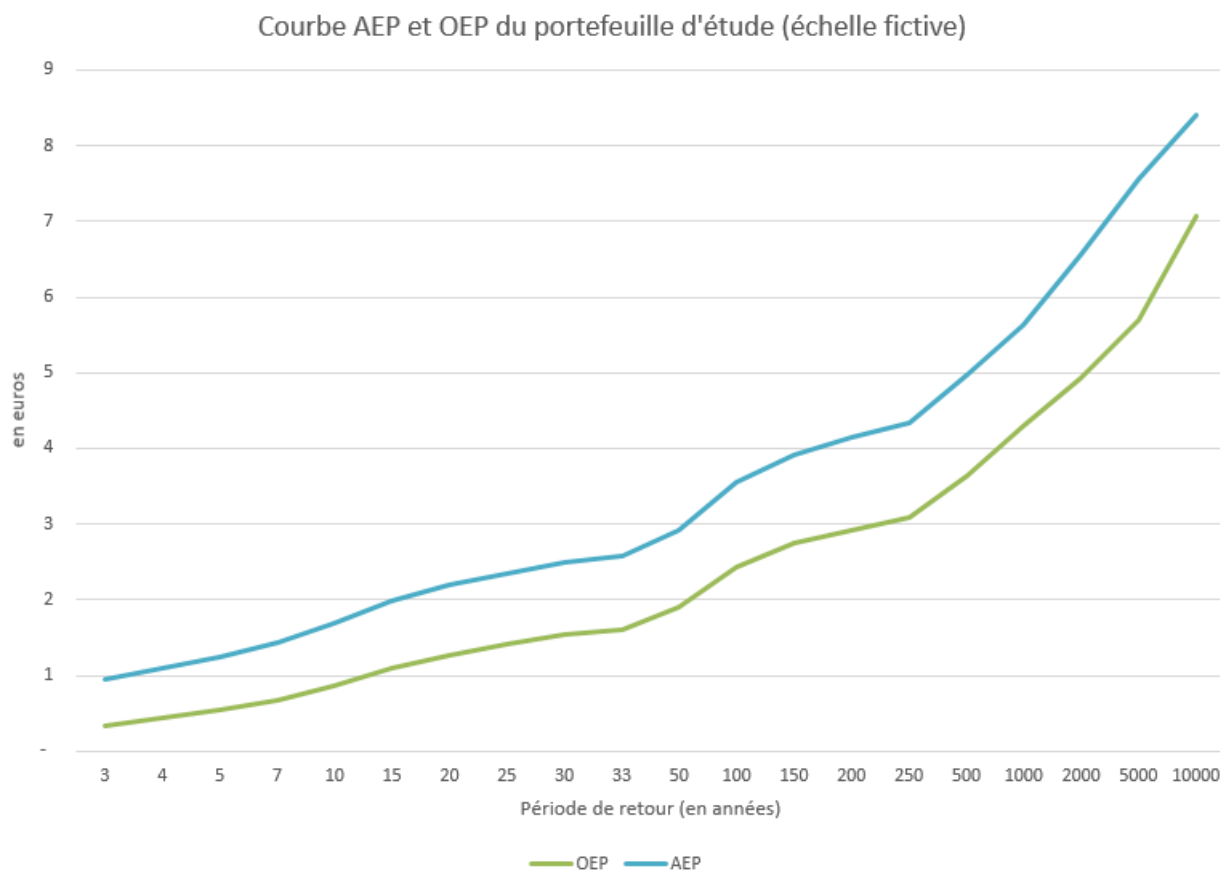


FIGURE 1.4 – Courbes AEP et OEP du portefeuille d'étude

1.4 *Synthetic Modelling*

En plus d'être utilisées pour la tarification des contrats, les sorties des modèles cat-nat permettent de générer des YLTs qui regroupent les pertes associées à plusieurs contrats et des *Region Perils* différents. En effet, les YLTs offrent une vision par scénario (on définira un scénario comme l'occurrence d'un événement défini par son EventId dans une année de simulation) qui peuvent ensuite être agrégés sur les années de simulation pour calculer les mesures de risques pertinentes pour les contrats étudiés (voire pour le portefeuille).

Cependant, cette approche n'est pas directement applicable pour les *Region Perils* non modélisés. En effet, les contrats qui couvrent des *Region Perils* non modélisés sont tarifés par expérience pour paramétrer des lois de fréquence-sévérité. Dans ce cas, on ne dispose pas de vision par scénario des pertes relatives à ces contrats vu qu'on n'a pas les catalogues d'événements associés nécessaires pour déduire leurs YLTs. D'où, on ne peut pas générer un YLT global du portefeuille qui regroupe toutes les pertes des contrats souscrits.

C'est ainsi que SCOR a mis en place la méthode de *Synthetic Modelling* depuis 2019 pour générer les mêmes sorties des modèles cat-nat pour des *Region Perils* non modélisés. Cette méthode sert à créer un YEQT propre à chacun de ces *Region Perils* non modélisés afin de déduire les YLTs des contrats qui les couvrent et qui sont homogènes avec les YLTs issus de la modélisation cat-nat et ainsi agrégeables pour déduire la YLT combinée du portefeuille (qu'on définira après).

1.4.1 Principe de la méthode

Le principe du *Synthetic Modelling* consiste en la génération de YEQT propre à chaque *Region Peril* non modélisé. Pour chacun, il s'agit de générer un catalogue d'événements de référence qui contient une liste de scénarios simulés sur les années de simulation (avec $N = 100\,000$ années de simulations) et ordonnés par ordre de sévérité de perte associée.

Ces YEQTs ont été générés sur $N = 100\,000$ ans de simulations suivant ce principe :

- Le nombre d'événements par année de simulation suit une loi Poisson de paramètre $\lambda = 2$ sauf pour les *Region Perils* USFL et USCT (les inondations et la grêle aux États Unis) où le paramètre de la loi Poisson est $\lambda = 4$. On tire ainsi N observations $(x_i)_{1 \leq i \leq N}$ de loi Poisson de paramètre λ .
- On réplique chaque année $i = 1 \dots N$, x_i fois et on lui attribue un scénario. On se retrouve en moyenne avec $L = N \cdot \lambda$ lignes¹ scénarios (environ 200 000).
- Les identifiants des événements sont spécifiques à chaque *Region peril* ce qui assure une décorrélation totale entre les événements des *Region Perils* différents mais une corrélation parfaite entre les pertes des contrats couvrant le même *Region Peril*.
- Les dates introduites dans les YEQTs sont totalement arbitraires.
- On ordonne ensuite la liste des scénarios de façon arbitraire et on leur attribue des rangs (1 est le rang associé au scénario générant la perte la plus élevée).

Year	EventId	DateTime	Rank
59804	1500460955	26/09/2020 06 :27	1
86848	1500353110	21/06/2020 14 :39	2
16237	1500385280	20/05/2020 15 :09	3
86888	1500358800	09/11/2020 19 :46	4

TABLE 1.6 – Exemple réduit du YEQT d'un *Region peril* synthétique

1. Si $X \sim Poi(\lambda)$, $\mathbf{E}(X) = \lambda$

On obtient donc une table avec en moyenne $L = N \cdot \lambda$ lignes. Cette table sert ensuite comme référence pour chaque *Region Peril* pour discrétiser les lois de fréquence-sévérité utilisées dans la tarification des différents contrats et obtenir les YLTs associés pour enfin les utiliser dans nos études d'accumulation en créant la YLT combinée du portefeuille. On détaillera mieux le processus de discrétisation dans le chapitre suivant.

1.4.2 Déroulement de l'accumulation

Les YEQTs prédéfinis pour chaque *Region Peril* synthétique permettent de discrétiser les distributions fréquence-sévérité établis dans la tarification des contrats pour générer des YLTs. Pour chacun de ces contrats, on discrétise la distribution de sévérité sur $N = 100\,000$ années de simulation et on fait appel au YEQT prédéfini de ce *Region Peril* pour faire correspondre les pertes obtenues aux scénarios de référence en respectant leurs rangs.

Cette opération permet d'avoir le YLT de chacun des contrats couvrant des *Region Perils* synthétiques. Ils offrent une vision par scénario homogène aux sorties des modèles cat-nat.

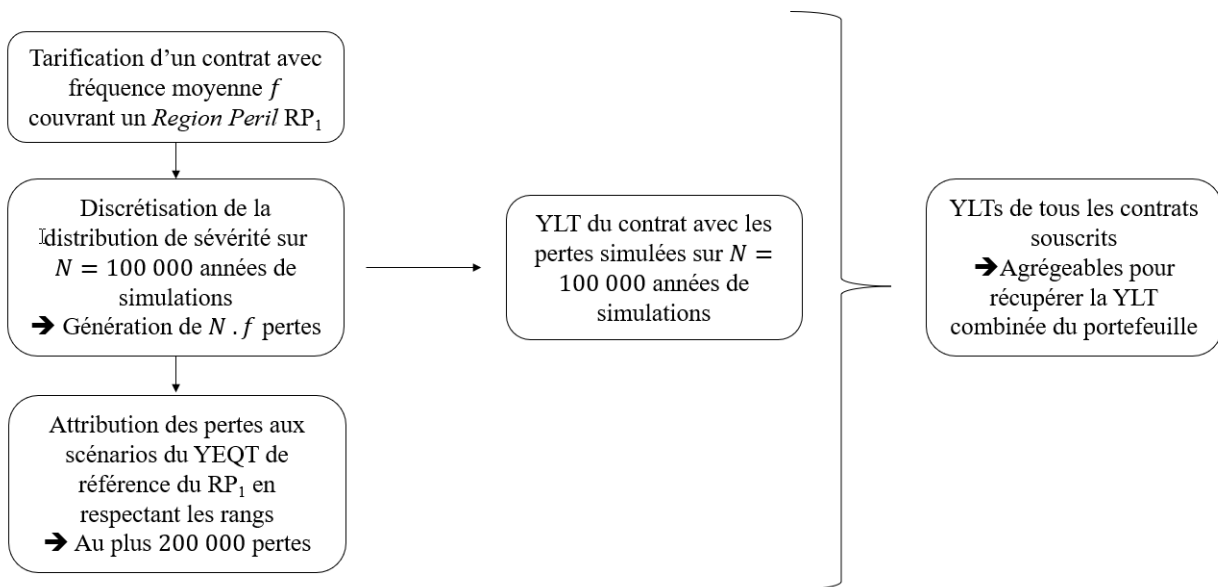


FIGURE 1.5 – Déroulement de l'accumulation pour les *Region Perils* non-modélisés

1.4.3 YLT combinée

La YLT combinée du portefeuille est la YLT qui regroupe toutes les pertes relatives aux différents contrats par année de simulation. Elle est créée en se basant sur les YLTs en sortie des modèles pour les contrats qui couvrent des *Region Perils* modélisés et des YLTs générés grâce à la méthode du *Synthetic Modelling* pour les *Region Perils* non modélisés.

YLT Segment ID	Event ID	Simulated Period ID	Event Date-Time	Loss
2807	15100000	1105	17/08/2013 00 :48	1 558 860 030
2904	2862386	420	29/07/2017 23 :32	1 221 711 620
2807	15006168	994	16/06/2013 00 :56	1 175 443 970
1130	3171597	2309	23/01/2013 13 :57	1 150 528 640

TABLE 1.7 – Partie de la YLT combinée du portefeuille d'étude

En utilisant la YLT combinée et une table appelée *header* qui contient les détails des contrats souscrits ainsi que les *Region Perils* couverts, on peut récupérer tous les détails des contrats présents dans le portefeuille.

YLTSegmentID	RegionPeril	CoverType	LossUnitCur	DomCur	Premium
1130	EUET	Prop...	EUR	EUR	55 032 234
2807	USEQ	Prop...	EUR	USD	52 201 388
2904	USTC	Prop...	EUR	USD	12 101 819

TABLE 1.8 – Partie du *header* lié au YLT du portefeuille d'étude

1.5 Objectifs du mémoire

L'idée de ce mémoire est d'améliorer la représentation des catastrophes naturelles dans notre portefeuille pour des études d'accumulation.

On commence par revoir le processus de génération des YEQTs prédéfinis des *Region Perils* synthétiques.

D'une part, on réévalue la fréquence utilisée pour la génération de ces YEQTs de référence. En effet, dans un contexte de changement climatique et une augmentation de la fréquence d'occurrence des catastrophes naturelles, on se demande si les 200 000 scénarios de référence suffisent pour discrétiser les lois de sévérité utilisées dans la tarification des contrats. Pour ce faire, les données de tarification permettent de voir si l'hypothèse $\lambda = 2$ permet d'absorber toutes les pertes possibles sur les $N = 100\,000$ années de simulation en comparant λ à f la fréquence moyenne de tarification.

D'autre part, on se propose d'introduire des dates représentatives du phénomène physique sous-jacent aux scénarios de référence des YEQTs des *Region Perils* synthétiques. Cette dimension temporelle peut contribuer à une meilleure vue du risque du portefeuille en comparant les résultats trimestriels attendus avec ce qui était estimé.

Finalement, on explore les dépendances entre les *Region Perils* différents afin d'éviter une sous-estimation du risque du portefeuille vu que la modélisation actuelle ne les prend pas en compte. En effet, pour les *Region Perils* modélisés, les modèles sont faits indépendamment de façon que, même chez un même modélisateur, il n'y a pas une prise en considération des dépendances qui peuvent avoir lieu entre deux régions périls différents. De plus, dans notre modélisation actuelle des *Region Perils* synthétiques, on suppose qu'ils sont indépendants du fait qu'on attribue des événements spécifiques à chacun. On envisage donc d'examiner des données historiques pour détecter les éventuelles dépendances possibles et les intégrer dans notre modélisation.

Chapitre 2

Étude de la fréquence

Introduction

Les YEQTs sont des tirages prédéfinis pour chaque *Region Peril*. Pour les modèles CAT, comme vu précédemment, ils sont dérivés du catalogue d'évènements physiques du modèle en question. Pour les *Region Perils* synthétiques, le processus est inversé : les modélisations arrivent à posteriori et sont sujettes à variations. Les variations peuvent être dues à un changement de modèle, une sinistralité ayant évolué, un changement dans la méthodologie. Par conséquent, le choix de fréquence λ dans la création du YEQT doit pouvoir absorber ces changements.

Les premiers YEQTs pour SCOR ont été définis en 2019 avec les statistiques de l'année précédente avec $\lambda = 2$ pour tous les périls, sauf pour les inondations et tempêtes de neige aux États Unis où $\lambda = 4$.

Dans le contexte du changement climatique, une révision de la fréquence d'évènements dans les YEQTs peut être nécessaire et fera l'objet de ce chapitre.

2.1 Hypothèses et fonctionnement de la méthode actuelle

Comme défini au premier chapitre, le *Synthetic Modelling* permet de déduire les YLTs des contrats qui couvrent des *Region Perils* non modélisés. Les hypothèses faites pour définir cette méthode, remontent à 3 ans.

En effet, en se référant sur des données des contrats souscrits, la moyenne de fréquence f de tarification utilisée était de 1.24 pour les différents contrats.

Dist. fréquence	Nombre de contrats	Fréq Maximale	Moyenne de la Fréq
Bernouilli	452		
Discrète	54		
Binomiale Négative	2	9	9
Poisson	5728	103.65	1.02
Autres dist. de Panjer	964	60	2.51
Total	7200	103.65	1.24

TABLE 2.1 – Fréquences historiques utilisées pour la tarification

Ainsi, le choix $\lambda = 2$ semblait conservateur pour générer les YEQTs référence des *Region Perils* non modélisés pour discrétiser la majorité des lois fréquence-sévérité utilisées dans la tarification en fonction de la fréquence moyenne f du contrat en absorbant tous les scénarios de pertes possibles.

En pratique, dans le cas des *Region Perils* non modélisés, on se base sur ces YEQTs prédéfinis et sur les distributions fréquence-sévérité élaborées dans la tarification pour déduire le YLT du contrat. Pour chaque contrat tarifé, on récupère le YEQT du *Region Peril* concerné et on discrétise la distribution de fréquence sur $N.f$ scénarios possibles avec f la fréquence moyenne de tarification auxquels on attribue finalement les pertes en se basant sur la distribution de sévérité.

Vu que dans la tarification la fréquence f utilisée peut être différente de $\lambda = 2$, on distingue deux cas de figures :

- Fréquence de tarification $f < 2$: nous conservons la fréquence de tarification et on récupère les $N.f$ premiers scénarios du YEQTs auxquels on attribue des pertes. Le YEQT aura donc une longueur $L < 200\ 000$.
- Fréquence de tarification $f > 2$: on effectue une troncature des pertes de rang $> 200\ 000$ pour 100 000 années de simulation.

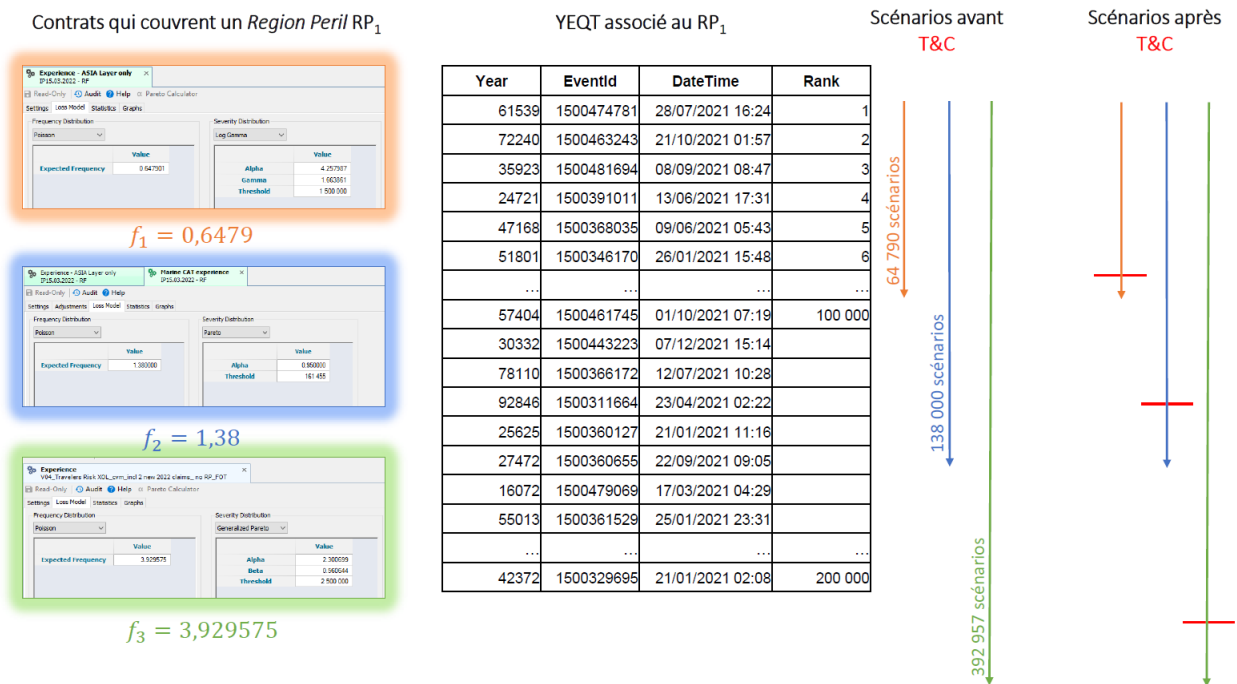


FIGURE 2.1 – Lien entre fréquence de tarification f et les YEQTs prédéfinis ($\lambda = 2$)

Ainsi, chaque *Region Peril*, couvert dans un contrat, a un catalogue d'évènements de longueur $L \leq 200\ 000$ lignes simulés sur $N = 100\ 000$ années de simulation. L'agrégation des pertes introduites dans les YEQTs par année de simulation permet d'avoir le YLT correspondant à un contrat.

Même si le choix de $\lambda = 2$ semble convenir à nos études d'accumulation, l'historique sur lequel on s'est basé a évolué ainsi que les modèles de tarification. Ainsi on se demande si cette hypothèse est toujours valable.

2.2 Présentation des données - *Loss Engines*

Les loss engines sont des modules extraits de l'outil de tarification interne à SCOR (xAct). Ils contiennent pour chaque traité souscrit, les distributions de pertes après application des termes et conditions du contrat. Pour notre étude de réévaluation de la fréquence, on va considérer les données de tarification les plus récentes à notre disposition et qui sont celles de l'année 2021.

UW_Year	Pricing_ID	Loss_Engine_ID	ARC_Data
2021	11751188	11751327	{"Version" :3,"Severity" :[900000, 900369.25,...], "Percentile" :[1E-05, 8E-05, 0.00048,...], "Frequency" : 2.127844}

TABLE 2.2 – Exemple d'un loss engine

Les *loss engines* contiennent 3 informations :

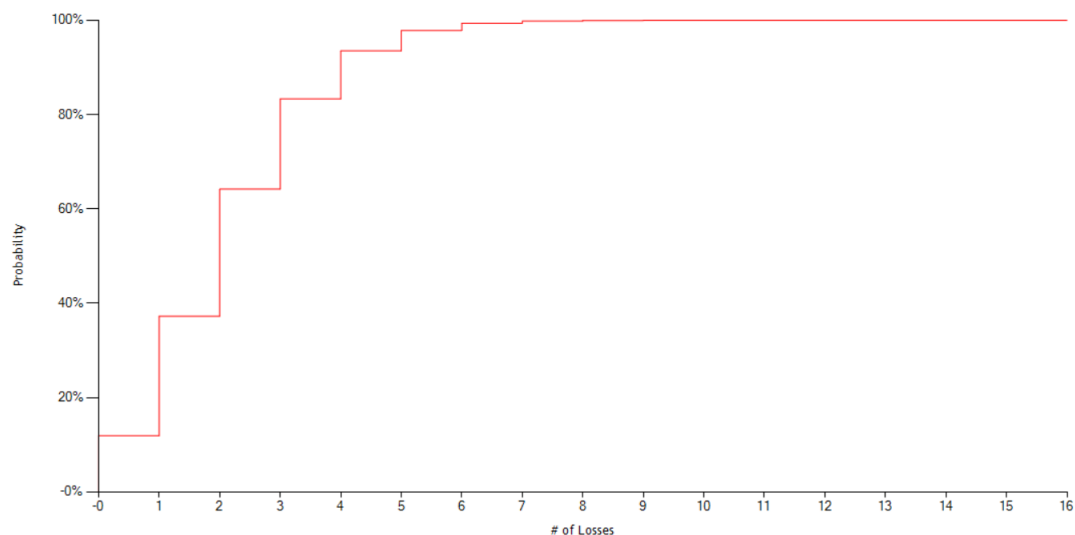


FIGURE 2.2 – Distribution fréquence pour le loss engine du tableau 2.2

- Severité : un vecteur de pertes nettes cumulées
- Percentile : probabilité de pertes
- Fréquence : la fréquence moyenne de la loi utilisée pour ajuster la fréquence

Chaque *loss engine* est lié à un contrat d'où à un *Region Peril*. On utilise la base de données des contrats pour lier ces *Loss Engines* avec les détails des contrats pour récupérer le *Region Peril* concerné.

Pour notre étude, nous nous intéressons aux fréquences avant termes et conditions. Les données issues des contrats non-proportionnels, étant biaisées par les franchises contractuelles, sont exclues. Nous ne garderons finalement que les distributions qui sont liées aux contrats proportionnels car elles représentent le risque pur.

2.3 Extraction de la fréquence et observations

On utilise les clés définissant chaque *loss engine* à savoir le Pricing_ID et le Loss_Engine_ID pour récupérer les détails du contrat et le *Region Peril*.

On se retrouve donc avec une ou plusieurs observations de fréquences moyennes de tarification

pour chaque *Region Peril* couvert.

On calcule, pour chaque *Region Peril* synthétique la moyenne des observations de fréquence de tarification.

Les fréquences moyennes calculées pour chaque *Region Peril* sont plutôt toujours inférieures à 2 sauf pour les périls CS et FL où on retrouve des *Region Perils* dont la fréquence dépasse 2 (figure 2.3). Ces observations de fréquence moyenne semblent s'aligner avec l'hypothèse actuelle puisque dans la majorité des cas la fréquence de tarification $f < 2$.

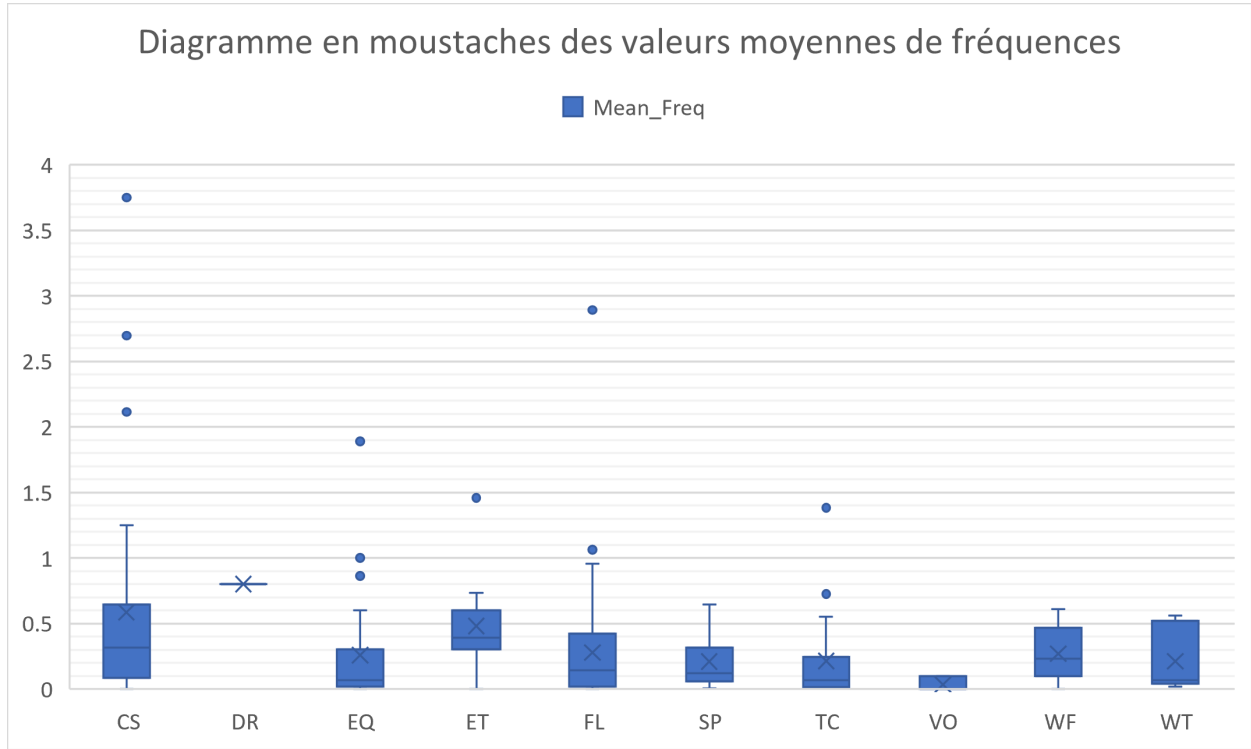


FIGURE 2.3 – Diagrammes en moustaches des moyennes des fréquences observées des *Region Perils*

Ainsi, en utilisant des YEQTs générés avec un paramètre $\lambda = 2$ on dispose d'une base de scénarios possibles plus large à partir de laquelle on peut tirer les scénarios relatifs à chaque contrat mais on se demande si on a intérêt à réduire la taille de ces YEQTs prédéfinis puisqu'ils contiennent plus de scénarios que nécessaire dans une grande partie des cas.

D'autre part, vu l'existence de cas où la fréquence moyenne annuelle d'évènements dépasse $\lambda = 2$ on peut suggérer une augmentation de la taille des YEQTs de référence pour couvrir tous les scénarios possibles.

2.4 Implémentation et conclusions

Après calcul de la fréquence moyenne pour chaque *Region Peril*, l'objectif est de voir l'impact de cette réduction ou augmentation de fréquence d'évènements annuels des YEQTs sur le portefeuille SCOR.

On dispose du YLT (combiné) du portefeuille SCOR et du fichier *header* qui permet de retrouver les *Region Perils* correspondant à ce catalogue d'évènements. On peut ainsi extraire pour chaque *Region Peril* son catalogue d'évènements classés par ordre de perte décroissante.

2.4.1 Cas où la fréquence moyenne < 2

Dans le cas où la fréquence moyenne des observations est < 2 , on génère les nouveaux YEQTs de ces *Region Perils* avec un nombre moyen d'évènements qui suit une loi poisson de paramètre $\lambda < 2$. Dans ce cas, le YEQT aura moins de scénarios que le YEQT pré-établi.

Pour mesurer l'impact de la réduction de fréquence, on choisit de faire le test avec des *Region Perils* du FL, CS et EQ dont les nouvelles fréquences suggérées sont < 2 puisqu'ils sont les périls les plus importants en termes de pertes parmi les *Region Perils* non-modélisés (représentent plus de 50% des pertes non-modélisées).

Pour ce faire, on génère les nouveaux YEQTs avec les nouvelles fréquences et dont la longueur serait évidemment inférieure à $L = 2.N / N = 100\ 000$ années de simulation.

Étant donné que les identifiants d'évènements sont spécifiques à chaque *Region Peril* et que les nouveaux YEQTs générés contiennent moins d'évènements que les YEQTs utilisés dans le YLT initial, le changement de fréquence va résulter en la réduction du nombre d'évènements de ces *Region Perils* modifiés dans le YLT.

Ces changements sont effectués en supprimant les lignes du YLT en excès.

Year	Event_Id	Date	Rank		Year	Event_Id	Date	Rank
71974	2859413	13/05/2022	1	Passage à une fréquence $\lambda < 2$, la taille du YEQT diminue (on génère $x \sim 100\ 000 \times \lambda$ évènements) au lieu de $L \sim 100\ 000 \times 2$	84611	2868922	04/07/2022	1
37902	2854000	22/07/2022	2		39003	2866119	22/04/2022	2
43200	2859413	18/12/2022	3		18431	2859321	27/01/2022	3
19510	2869738	23/12/2022	4		14822	2863788	21/03/2022	4
...
39 862	2861648	18/02/2022	x-2		8809	2857066	06/08/2022	x-2
999	2864977	03/10/2022	x-1		25111	2869513	19/06/2022	x-1
51 829	2861860	23/10/2022	x		94620	2867626	28/11/2022	x
...
54 406	2863642	04/03/2022	L-2					
898	2867303	22/02/2022	L-1					
73285	2854757	02/06/2022	L					

FIGURE 2.4 – Passage de fréquence 2 à $\lambda < 2$

On effectue à chaque fois pour un péril le changement de fréquence et on mesure l'impact sur le portefeuille d'étude (cf. tableau 2.4) :

Portefeuille	Variation de la perte moyenne annuelle en %
Après changement de fréquence des périls CS	-0.21%
Après changement de fréquence des périls EQ	-0.05%
Après changement de fréquence des périls FL	-0.44%

TABLE 2.3 – Variation des pertes en réduisant les fréquences par péril

On remarque que les variations en pertes sont assez réduites suite à la réduction de fréquence. Ceci revient au principe de génération des YEQTs en définissant le nombre d'évènements par année de simulation avec une loi Poisson pour laquelle on peut démontrer ce lemme (cf. annexe A).

Lemme 1. Soit N une variable aléatoire de loi $\mathcal{P}(\lambda)$ avec $\lambda > 0$ et D une variable aléatoire telle que $D | N \sim \mathcal{B}(N, p)$ avec $p > 0$. Alors

$$N - D \sim \mathcal{P}(\lambda(1 - p)).$$

En effet, on se donne le vecteur d'années de simulation $A = (1, \dots, N)$ tel que $N = 100\,000$. On attribue à chaque année i de simulation, un nombre d'évènements X_i tiré avec une loi Poisson de paramètre $\lambda = 2$.

Pour générer les YEQTs prédéfinis, on réplique les années de simulation suivant le nombre d'évènements X_i et on les réarrange. On aboutit enfin à un catalogue d'année de simulations de longueur moyenne $L = 200\,000 \approx \lambda \cdot N$ lignes et où l'année i se réalise X_i fois.

Le passage à $\lambda_1 < 2$ induit que le catalogue d'évènements serait de longueur $L_1 = \lambda_1 \cdot N < L$. Dans ce cas, il s'agirait, à partir du YEQT prédéfini de tirer un échantillon aléatoire de longueur L_1 . Ainsi, la proportion de conserver une réalisation d'une année i serait de $p = \frac{L_1}{L}$.

D'où, le nombre d'évènements par année qui restent dans le catalogue après tirage suit une loi Poisson de paramètre $\lambda_1 = \lambda \cdot (1 - \frac{L_1}{L}) = \lambda - \lambda_1$. Finalement, par la propriété de la somme de lois de poisson, l'échantillon tiré aurait le paramètre λ_1 .

Étant donné que dans le cas où la tarification est faite avec une fréquence $f < 2$, le fait de tirer son YEQT d'un YEQT généré avec une loi Poisson de paramètre $\lambda > f$ n'a pas d'impact puisque le YEQT final serait de fréquence égale à celle du pricing.

Conclusions

Aucune modification à la baisse n'est suggérée pour la fréquence utilisée pour générer les YEQTs de référence vu que ça ne permet pas de gagner de façon remarquable en précision ni en temps de calculs. De plus, en partant du principe de la méthode du *Synthetic Modelling*, pour la discrétisation de la loi de sévérité, si la fréquence de tarification $f < 2$ on ne se sert que du nombre de scénarios dont on a besoin en tirant que les $f \cdot N < 200\,000 / N = 100\,000$ premiers scénarios du YEQT de référence.

2.4.2 Cas où la fréquence moyenne > 2

Dans cette partie, on se concentre sur les changements des fréquences pour les *Region Perils* où la nouvelle fréquence moyenne est > 2 . Ce n'est le cas que pour 4 *Region Perils* à savoir ARCS, FRCS, GBCS et CNFL qui correspondent respectivement à la grêle en Argentine, France, Grèce et aux inondations en Chine.

Dans ce cas, on est amené à générer plus d'évènements pour compléter les YEQTs. On réintègre les nouveaux YEQTs ensuite dans le YLT du portefeuille en faisant correspondre les rangs pour le nombre d'évènements initial et en générant de nouvelles pertes pour les rangs inexistantes par interpolation linéaire à partir de la plus petite perte du *Region Peril* dans le portefeuille initial jusqu'à 0.

On sait déjà que la dernière perte dans les YLTs serait égale à 0 et le choix d'effectuer une interpolation linéaire pour générer les nouvelles pertes était le plus simple étant donné qu'on n'a pas d'informations disponibles par rapport au comportement des pertes.

Year	Event_Id	Date	Rank
15734	2858418	11/09/2022	1
1321	2862361	19/06/2022	2
14008	2859584	12/08/2022	3
93209	2861983	28/04/2022	4
...	...	04/01/2022	...
52709	2860643	27/06/2022	L-2
24105	2865692	05/05/2022	L-2
22980	2859371	14/02/2022	L ★

Passage à une fréquence $\lambda > 2$, la taille du YEQT augmente (on génère $x \sim 100\ 000 \times \lambda$) événements

au lieu de $L \sim 100\ 000 \times 2$

Les pertes générées par ces événements ajoutés, décroît linéairement de la perte engendrée par l'évènement ★ jusqu'à 0.

Year	Event_Id	Date	Rank
71974	2859413	13/05/2022	1
37902	2854000	22/07/2022	2
43200	2859413	18/12/2022	3
19510	2869738	23/12/2022	4
...
39 862	2861648	18/02/2022	L-2
999	2864977	03/10/2022	L-2
51 829	2861860	23/10/2022	L
...
54 406	2863642	04/03/2022	x-2
898	2867303	22/02/2022	x-1
73285	2854757	02/06/2022	x

FIGURE 2.5 – Passage de fréquence 2 à $\lambda > 2$

L'impact de cette modification se résume dans ce tableau :

<i>Region Peril</i> modifié	Fréq	Variation de la perte associée	Part du portefeuille
CNFL	2.89	1%	0.41%
ARCS	2.7	14%	0.04%
FRCS	2.11	1%	0.35%
GBCS	3.75	36%	0.03%

TABLE 2.4 – Variation des pertes en réduisant les fréquences par péril

Même si l'impact semble plus important sur les pertes moyennes d'un *Region Peril* comparés avec d'autres, l'impact sur le portefeuille globalement est presque nul :

Portefeuille	Variation de la perte moyenne annuelle en %
Après changement de fréquence des périls CS	0.02%
Après changement de fréquence des périls FL	-0.001%

TABLE 2.5 – Variations de fréquences >2 et impacts sur le portefeuille

Conclusions

En vue du faible impact de l'augmentation de la fréquence utilisée dans la génération des YEQTs de référence des *Region Perils* concernés, aucune modification n'est effectuée.

Ces faibles variations peuvent être le résultat de différents facteurs tels que :

- **La part de ces *Region Perils* dans les pertes SCOR** : En effet, les *Region Perils* qui présentent des fréquences > 2 contribuent avec moins de 1% aux pertes totales. Ceci justifie l'impact presque nul de l'ajout de nouveaux événements dans leurs YEQTs initiaux.
- **Notre choix/limite de modélisation** : Notre point de départ est le YLT du portefeuille SCOR sur lequel on effectue des changements. En augmentant la fréquence d'évènements d'un *Region Peril*, nous agissons plutôt sur les pertes faibles (les rangs les plus élevés). En effet, on ajoute des événements dont la perte maximale est plus faible que toutes les pertes initiales incluses dont le YLT global. Ainsi, la modification, étant faite sur les pertes les plus basses, pourrait aussi justifier les résultats.

2.4.3 Conclusions

Les impacts de variation de la fréquence moyenne d'occurrence annuelle des sinistres des *Region Perils*, sont faibles sur le portefeuille général :

- Dans le cas d'une réduction de λ : Le fait de nous rapprocher le plus de la moyenne utilisée pour la tarification ne nous fait pas gagner en précision puisqu'en réalité lorsque la fréquence de tarification est λ on tire le nombre nécessaire des scénarios du YEQT prédéfini.
- Dans le cas d'une augmentation de λ : Les *Region Perils* concernés par une augmentation de fréquence de tirage représentent une faible partie du portefeuille d'études d'où l'absence d'impact de ce changement ce qui fait qu'on maintient notre hypothèse initiale du $\lambda = 2$.

Finalement, dans les deux cas de figure étudiés, les modifications suggérées pour la fréquence ne fournissent visiblement pas de résultats très importants. Ainsi, on garde l'hypothèse $\lambda = 2$ pour la loi de Poisson qui caractérise la fréquence annuelle des sinistres.

Chapitre 3

Étude de la saisonnalité

Introduction

Les YEQTs générés pour les *Region Perils* synthétiques attribuent des dates arbitraires aux occurrences des événements. Cependant, on peut observer que pour certains périls, les occurrences sont naturellement concentrées pendant une saison de l'année et leurs occurrences les plus sévères en termes de pertes ont une tendance d'arriver aux mêmes mois.

L'objectif de cette partie est de mettre à jour les dates de ces YEQTs prédéfinis avec des dates bien représentatives du phénomène, à savoir attribuer des mois d'occurrence qui respectent le comportement naturel de ces événements.

3.1 Présentation des données utilisées

3.1.1 Données historiques de sinistres CAT de SCOR

C'est une base de données qui contient les détails d'environ 3000 sinistres de catastrophes naturelles qui ont engendré des pertes chez SCOR entre les années 2001 et 2021. Les pertes présentes dans ce document sont indexées. L'indexation des pertes permet de mieux les ajuster en considérant l'évolution actuelle du marché et le changement de la valeur du risque sous-jacent.

On y retrouve des sinistres enregistrés pour plus de 280 *Region Perils* différents dans le monde. Cependant, le nombre de sinistres par *Region Peril* n'est pas réparti de façon équilibrée.

En effet, on retrouve 36% des *Region Perils* avec uniquement un seul sinistre enregistré pour l'historique de 20 ans et des cas extrêmes comme la grêle aux États Unis (USCS) où on a 560 données historiques.

3.1.2 Base de données du CRED

Il s'agit d'une base de données élaborée par Le Centre de recherche sur l'épidémiologie des catastrophes (CRED) qui est une unité de recherche de l'université de Louvain en Belgique. Le champs de recherche du CRED comprend deux axes principaux : les catastrophes naturelles et les situations de crise causées par des troubles civils, des conflits ou autres [3].

Il y était élaboré depuis 1988 une base de données internationale de catastrophes (EM-DAT). Cette base de données contient des informations sur l'occurrence et les effets de plus de 18 000 catastrophes de masse dans le monde entier, de 1900 à nos jours y compris les catastrophes naturelles, les épidémies et plusieurs autres types de catastrophes.

On s'intéresse aux informations sur les catastrophes naturelles dont les pertes sont mentionnées. On a remarqué que les données sont plus consistantes et exploitables depuis 1990 (cf. figure

3.1). On se réduit ainsi à l'utilisation des données entre 1990 et 2021.

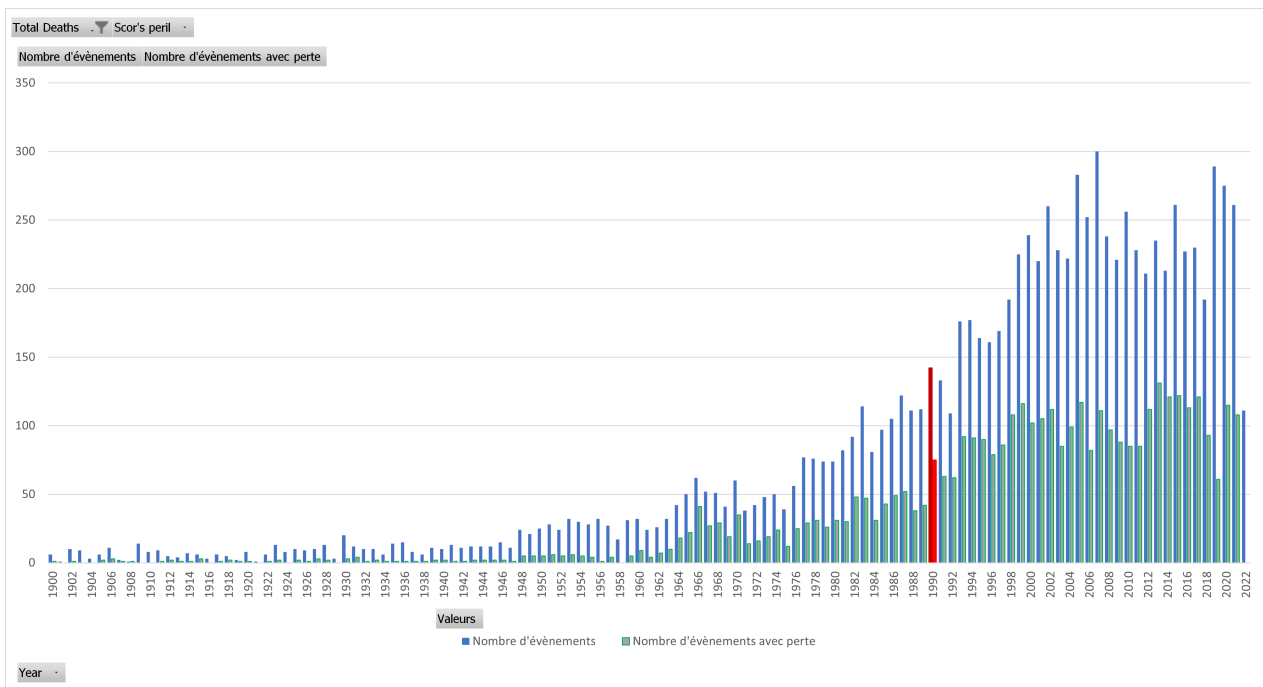


FIGURE 3.1 – Évolution du nombre de données de pertes financières disponibles par rapport au nombre de catastrophes naturelles

Pour les catastrophes naturelles, on retrouve la date de l'évènement, son type qu'on convertit en périls comme définis chez SCOR, le pays utilisé pour ensuite déduire le *Region Peril* concerné et les pertes totales indexées.

On obtient finalement, grâce à cette base de données, les informations concernant presque 4000 sinistres de catastrophes naturelles.

3.1.3 Défis de la donnée et choix de modélisation

L'objectif de cette étude est initialement d'introduire des changements sur les YEQTs des *Region Perils* synthétiques. Pour ce faire, on aurait besoin de données historiques avec ce niveau de granularité, ce qui n'est pas évident vu le caractère de faible fréquence des catastrophes naturelles.

C'est pour cette raison, qu'on a choisi d'étudier les périls non pas au niveau des pays (ou régions) mais plutôt sur un niveau plus global qui est celui des sous-continentes.

Ensuite, on s'est aperçu qu'étudier les périls sur ce niveau large introduisait du bruit dans les données. D'où on a introduit une variable qui est le type de climat pour étudier les caractéristiques des *Region Perils* historiquement.

Selon la classification des climats de Köppen-Geiger, on se retrouve avec 5 grandes zones climatiques [8] :

Zone climatique	Définition
A	Tropical
B	Sec
C	Tempéré
D	Continental
E	Polaire

TABLE 3.1 – Zones climatiques selon Köppen-Geiger



FIGURE 3.2 – Climat des pays du monde selon Köppen-Geiger

Ces choix ont été fondés sur le fait que :

- On étudie des catastrophes naturelles qui sont en lien étroit avec le climat.
- L'agrégation des données sur un niveau de sous-continent x type de climat était assez granulaire pour conserver des tendances (saisonnalité, mêmes zones géographiques, mêmes types de constructions...) et moins granulaire qu'au niveau des *Region Perils* pour réussir à avoir des données exploitables.

D'où, dans ce qui suit, on appellera clé (le niveau sur lequel on travaille), la combinaison : Péril x Sous-continent x Type de climat.

Si on regroupe les *Region Perils* qui ont le même péril, le même sous-continent et dont les pays ont le même type de climat sous une même clé, on se retrouve avec 165 clés qui peuvent être définis.

En faisant cette catégorisation des *Region Perils*, on passe de 299 *Region Perils* synthétiques couverts à étudier à 142 clés dont certaines regroupent jusqu'à 11 *Region Perils* synthétiques telle que la clé *FL x Southern Europe x C* qui regroupe les *Region Perils* d'inondations dans les pays de l'Europe du Sud dont le climat est tempéré et qui présentent en moyenne 4 millions d'euros de perte moyenne annuelle.

3.2 Calculs des vecteurs de saisonnalité

Les calculs et analyses de la saisonnalité sont faits au niveau des clés comme définis dans la partie 3.1.3.

3.2.1 Extraction de la saisonnalité

Nous disposons de deux bases de données avec des sinistres historiques et leurs pertes engendrées. Pour chaque base de données, on procède de la manière suivante pour toute clé :

1. Pour chaque mois, on calcule le pourcentage des évènements de ce mois par rapport au nombre total d'évènements dans l'historique. On obtient ainsi un vecteur de proportions *props* qui caractérise le nombre des évènements qui arrivent dans chaque mois par rapport au nombre total d'évènements associés à la clé d'étude.
2. On calcule la somme des pertes des évènements de chaque mois de l'historique et on calcule la proportion de ses pertes par rapport aux pertes totales historiques engendrées par les sinistres de cette clé. On obtient ainsi un vecteur qui caractérise la sévérité des évènements de chaque mois *sevs*.

On commence par tracer les vecteurs *props* pour les différentes clés pour visualiser les saisonnalités.

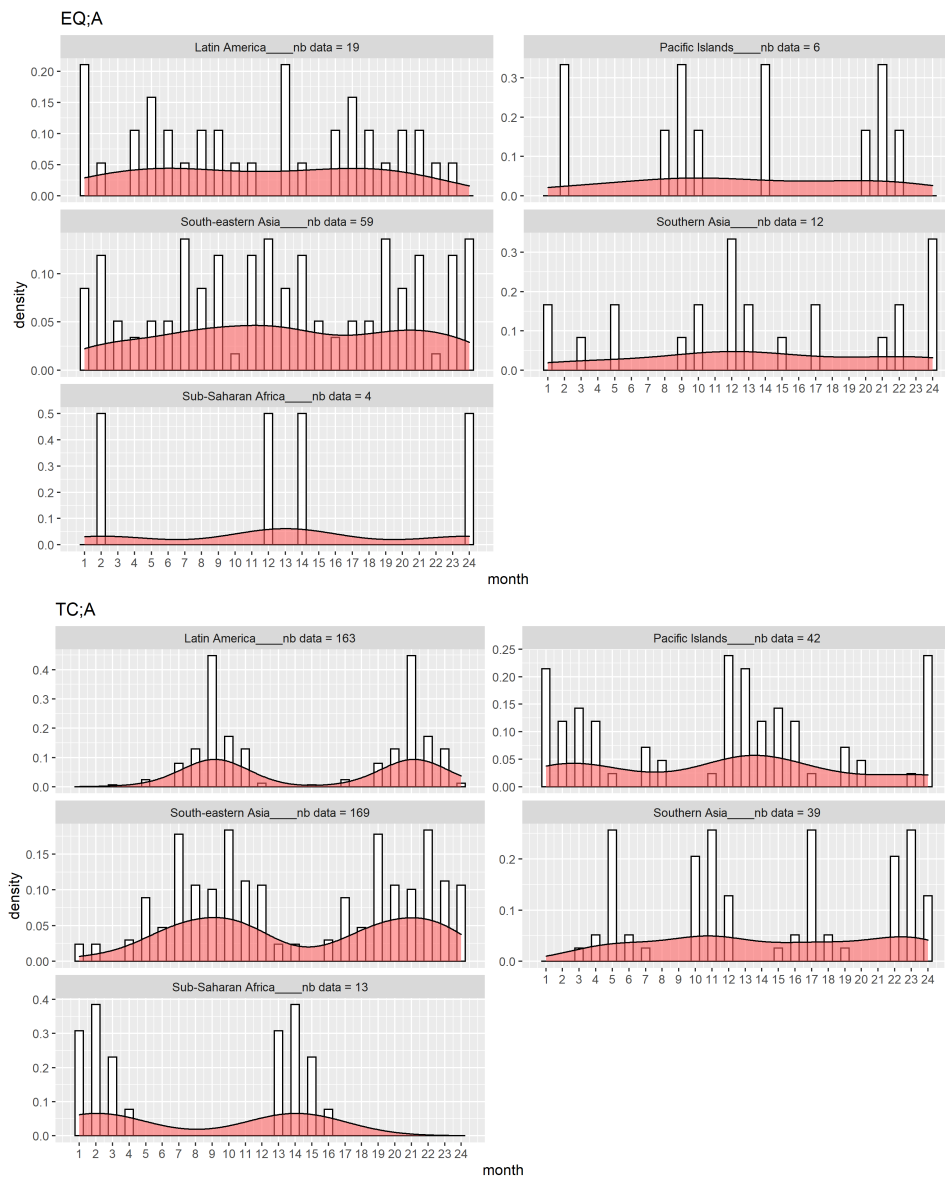


FIGURE 3.3 – Exemples de vecteurs de saisonnalités observés pour le tremblement de terre et le cyclone tropical en climat tropical

On peut remarquer qu'il y a des clés où on peut détecter des saisonnalités visibles (qui concernent

des périls tel que FL, TC, CS...) alors que pour d'autres périls on n'observe pas de saisonnalité à savoir les tremblements de terre et les périls qui regroupent plusieurs *Region Perils* (OP et OW).

Ensuite, on évalue le nombre de données disponibles pour chaque clé depuis les deux bases de données disponibles (CRED et les sinistres SCOR) pour choisir les clés pertinentes à étudier et quelle base de données à utiliser. On retrouve dans le tableau suivant la liste des périls et clés à étudier ainsi que le nombre total des *Region Perils* touchés :

Péril	Nombre de clés	Nombre de <i>Region Perils</i>
FL	24	86
TC	12	17
CS	10	25
WT	7	8
WF	3	4
SP	3	4
VO	2	2
ET	2	4
DR	1	1
Total général	64	151

TABLE 3.2 – Nombre de *Region Perils* avec saisonnalité

3.2.2 Moyenne Mobile Pondérée sur les vecteurs de saisonnalités

L'enjeu du faible nombre de données historiques persiste. En effet, on remarque que les vecteurs de saisonnalité (*props* et *sevs*) ne sont pas lisses et on observe parfois des pics de saisonnalité énormes.

Ainsi, pour pouvoir mieux représenter la réalité et lisser les données, on effectue une moyenne mobile pondérée sur ces vecteurs (centrée sur 3 valeurs) telle que

$$\tilde{x}_m = 0.25(x_{m-1} + x_{m+1}) + 0.5x_m$$

où

$$\begin{cases} x_m \text{ est la valeur observée au mois } m \\ \text{Si } m = 1, m - 1 = 12 \\ \text{Si } m = 12, m + 1 = 1 \end{cases}$$

Grâce à cette transformation, on se retrouve avec des vecteurs de saisonnalité qui présentent moins de pics irréguliers et plus lisses sur l'année (cf figure 3.4).

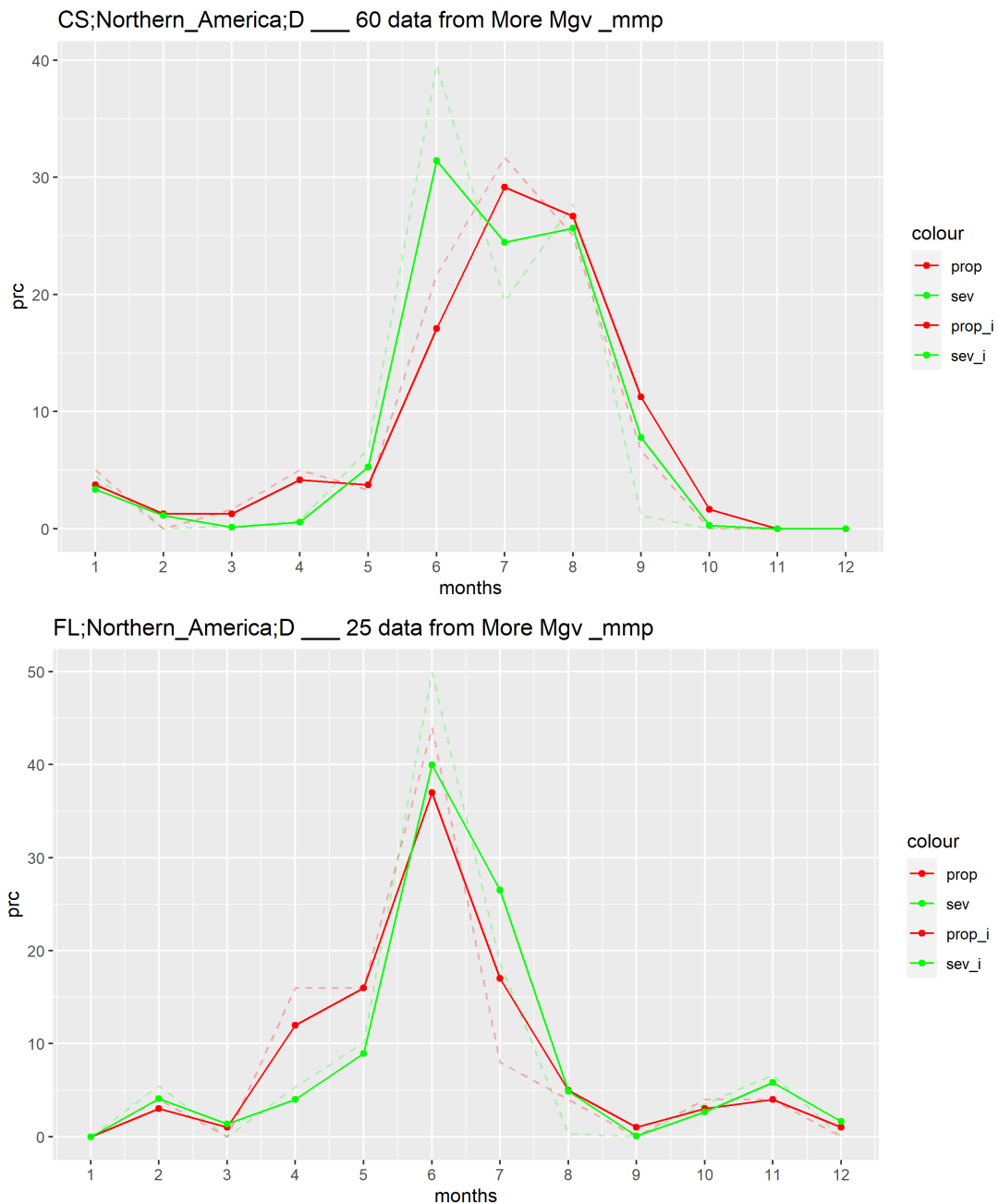


FIGURE 3.4 – Vecteurs de saisonnalité pour la grêle et les inondations aux pays de l’Amérique du nord à climat continental (Canada)

3.3 Implémentation de la saisonnalité

Nous disposons, pour chaque clé, de deux vecteurs qui caractérisent la saisonnalité en terme de nombre d’évènements par mois et de sévérité des évènements par mois. Afin de pouvoir utiliser ces deux vecteurs pour générer les dates dans les YEQTs, on a fait une analyse de la saisonnalité pour des *Region Perils* modélisés :

- Le cyclone extra-tropical en Europe (EUET)
- Le cyclone tropical au Japan (JPTC)
- Le cyclone tropical aux États Unis (USTC)

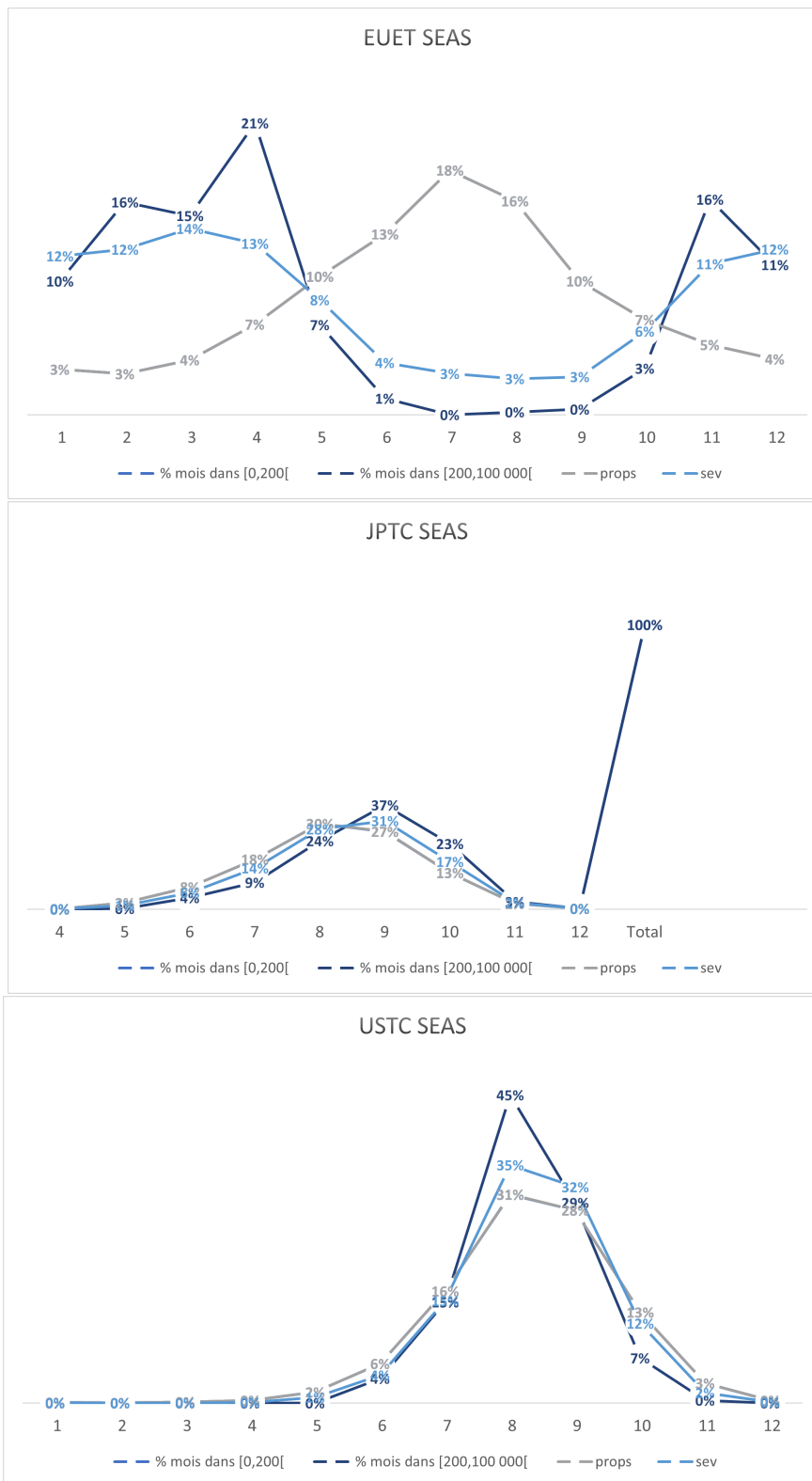


FIGURE 3.5 – Saisonnalité observée pour *Region Perils* modélisés

On partage les évènements des YEQTs de ces 3 *Region Perils* en deux classes :

- Les évènements dont la période de retour est < 200 : les évènements dont les pertes sont faibles mais qui ont la plus grande fréquence d'occurrence. Ils regroupent les évènements fréquents et de sévérité relativement faible.
- Les évènements dont la période de retour est > 200 : ce sont les évènements extrêmes, qui ont une chance d'arriver tous les 200ans ou plus mais qui engendrent des pertes

importantes.

On remarque que la courbe des proportions des événements de chaque mois dans la tranche de période de retour < 200 se superpose à la courbe de saisonnalité globale du *Region Peril* (*props*). D'autre part, les proportions de mois dans la tranche des événements les plus sévères suit la même variation que le vecteur *sevs*.

Ainsi, les événements dont la période de retour < 200 sont les événements qui alimentent la saisonnalité du *Region Peril* en terme de nombre d'événements. Cependant, les événements de période de retour > 200 sont ceux qui imposent la distribution des pertes par mois.

Finalement, on adopte le même principe observé pour les *Region Perils* modélisés pour introduire la saisonnalité aux YEQTs des *Region Perils* synthétiques.

3.4 Impact de l'introduction de la saisonnalité sur un contrat

Afin de mesurer l'impact de l'introduction des vecteurs de saisonnalité générés suite à cette étude dans les YEQTs des *Region Perils* synthétiques, on se donne un contrat couvrant les périls inondations, grêle et tempête de neige au Canada (CAFL, CACS et CAWT). Ces *Region Perils* ne sont pas modélisés dans le portefeuille et appartiennent au sous-continent : Amérique du Nord et de type de climat Continental (D).

On récupère les YLTs de ce contrat pour ces trois *Region Perils* (contiennent les pertes brutes qu'engendrent les catalogues d'événements de CAFL, CACS et CAWT pour ce contrat). On regroupe les trois YLTs des différents *Region Perils* et on remplace les dates d'événements avec les nouvelles dates qui contiennent la saisonnalité.

On voit bien que la modélisation initiale ne prend pas compte de la saisonnalité :

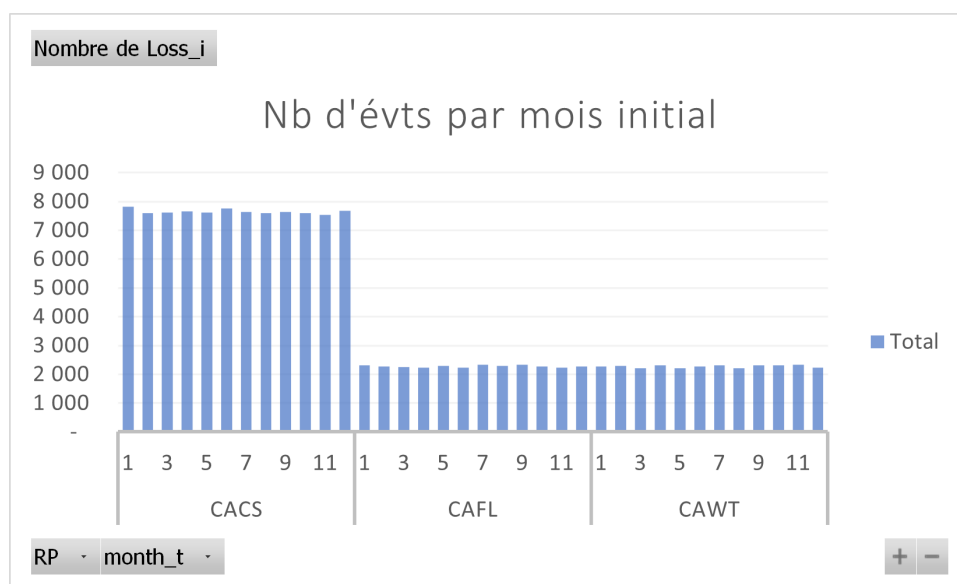


FIGURE 3.6 – Nombre d'événements pas mois par *Region Peril* dans le YLT initial du contrat

Pour les 3 périls, le nombre d'événements est constant sur tous les mois.

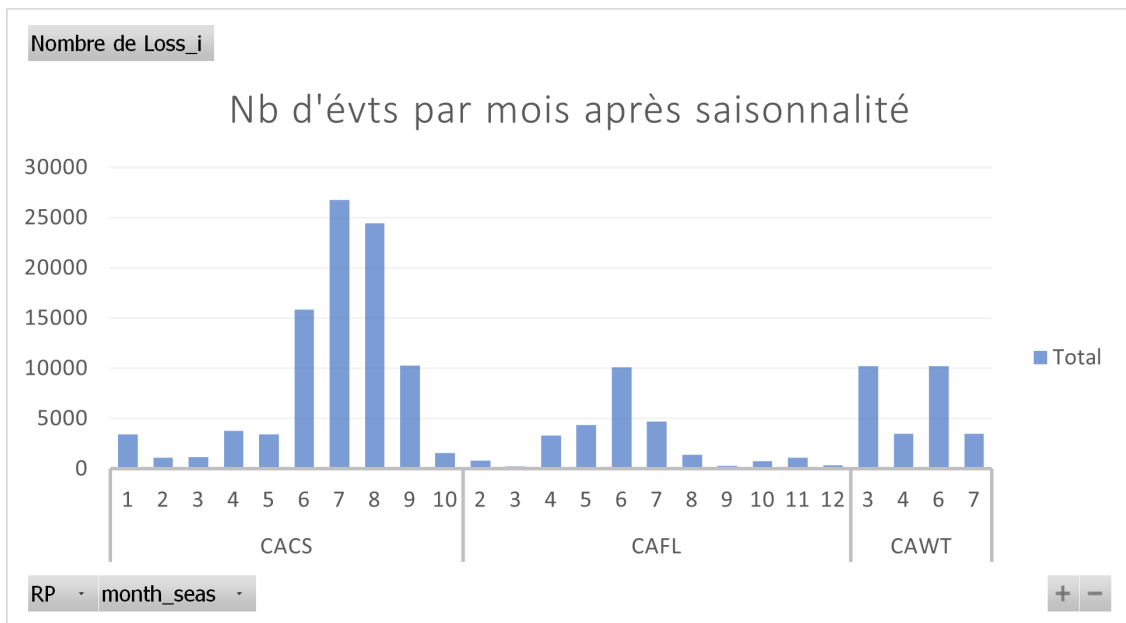


FIGURE 3.7 – Nombre d’évènements pas mois par *Region Peril* dans le YLT du contrat après introduction de la saisonnalité

Après prise en considération de la saisonnalité, on vérifie que le nombre d’évènements par mois dans le YLT comme présenté dans la figure 3.7 respecte cette dimension temporelle.

3.4.1 Avant termes et conditions

On calcule la distribution de la perte moyenne pour ce contrat par trimestre et la contribution de chaque *Region Peril* avant et après introduction de la saisonnalité (cf. figures 3.8 et 3.9).

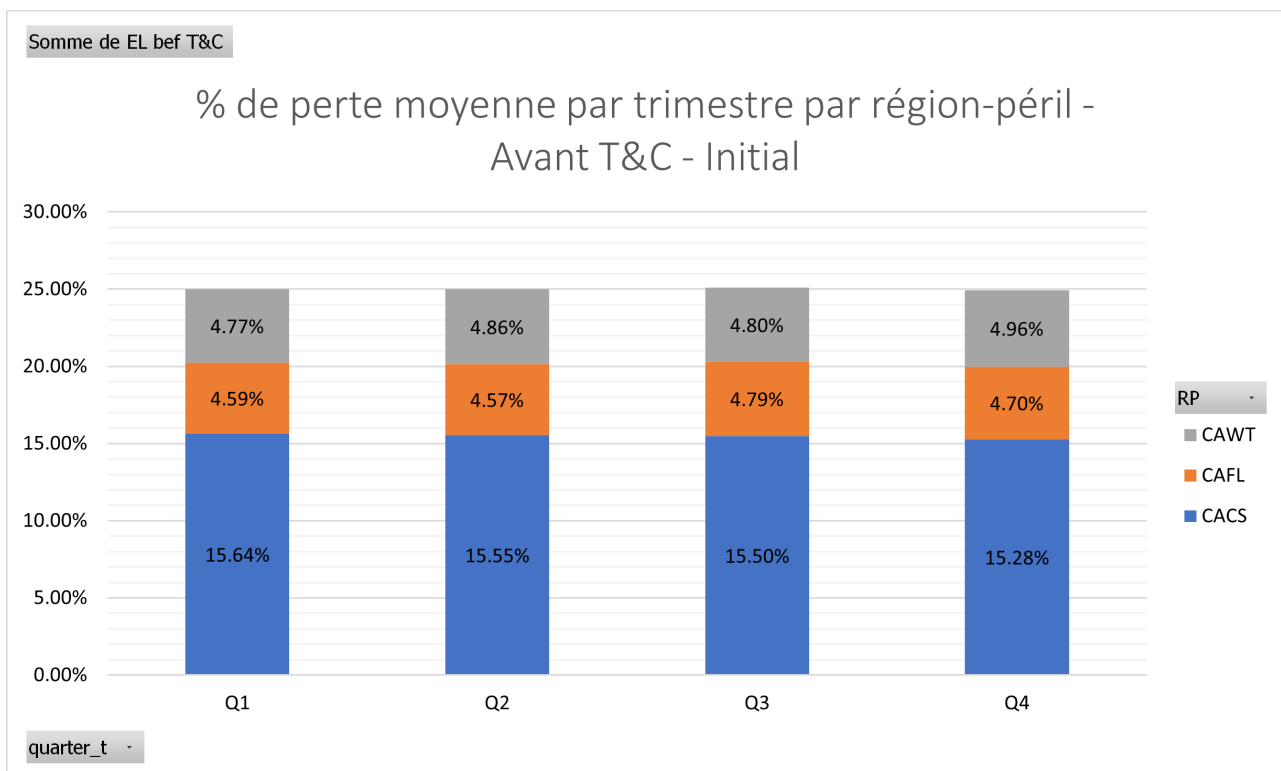


FIGURE 3.8 – Distribution initiale de pertes sur l’année par *Region Peril*

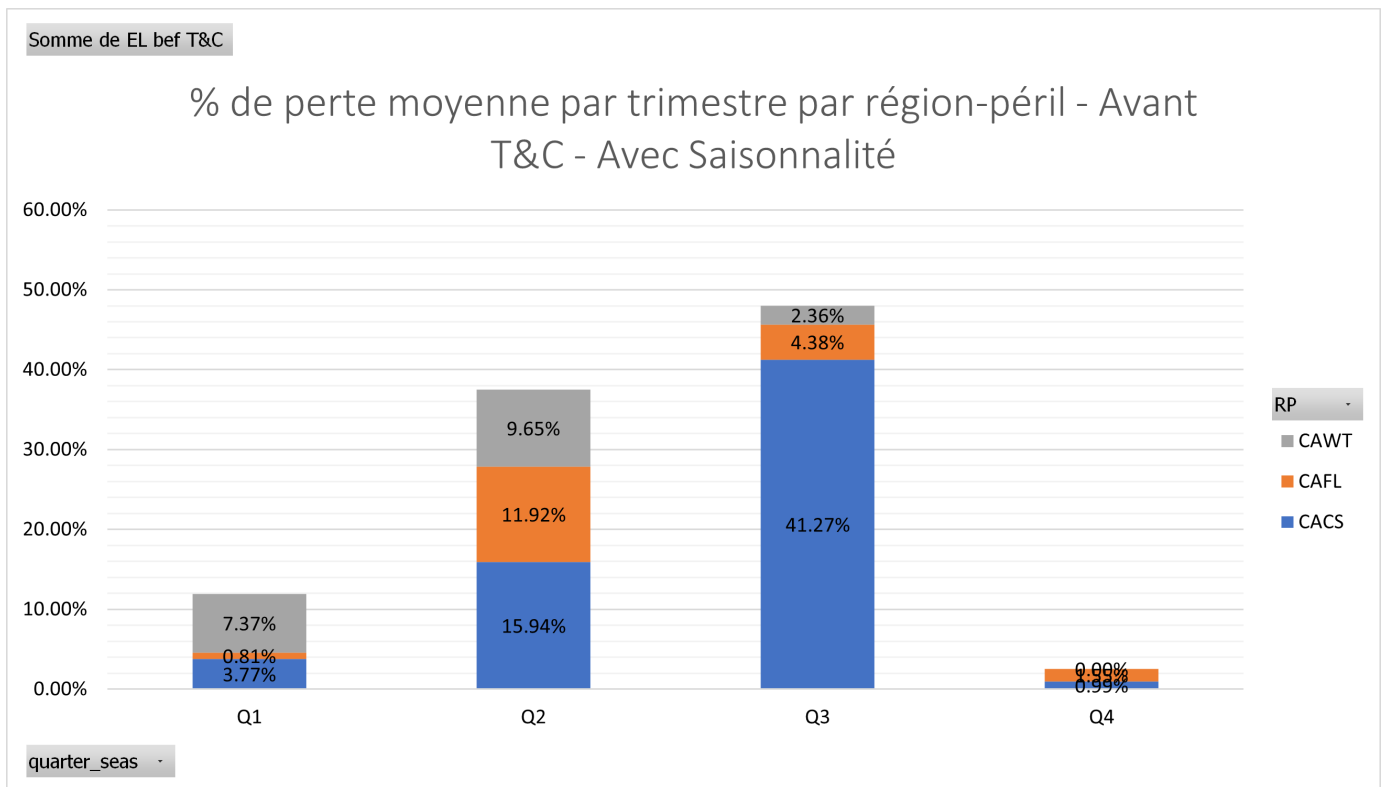


FIGURE 3.9 – Distribution de pertes sur l’année par *Region Peril* après introduction de la saisonnalité

Avant introduction de la saisonnalité, la perte moyenne est la même pour les quatre trimestres avec la même contribution de chaque *Region Peril*. Cependant, en introduisant la saisonnalité, la perte est plus concentrée au troisième trimestre pour atteindre son minimum au dernier trimestre de l’année. De plus, la contribution des périls aux pertes trimestrielles varie vu qu’elle dépend de leurs saisonnalités.

3.4.2 Après termes et conditions

On se propose maintenant de voir l’impact de la saisonnalité sur un contrat non-proportionnel de type Cat XL : 20 000 000 XS 10 000 000 euros.

Attachement Point	10 000 000
Limit	20 000 000
Expected Loss (EL)	796 840
Loss On line (LoL)	3.98%
Return Period of LoL	25 years
Probability of attachement	7 years
Probability of exhaustion	104 years

TABLE 3.3 – Termes et conditions et généralité sur le contrat

On introduit ainsi une tranche de couverture moyenne. En effet, on a une chance d’atteindre la franchise une fois tous les 7 ans et de perdre toute la limite une fois tous les 25 ans avec une perte annuelle moyenne qui est presque égale à 80 000 euros.

On étudie la variation des pertes trimestrielles et la contribution de chaque *Region Peril* dans la perte totale après ajout des T&C (cf. figures 3.10 et 3.11).

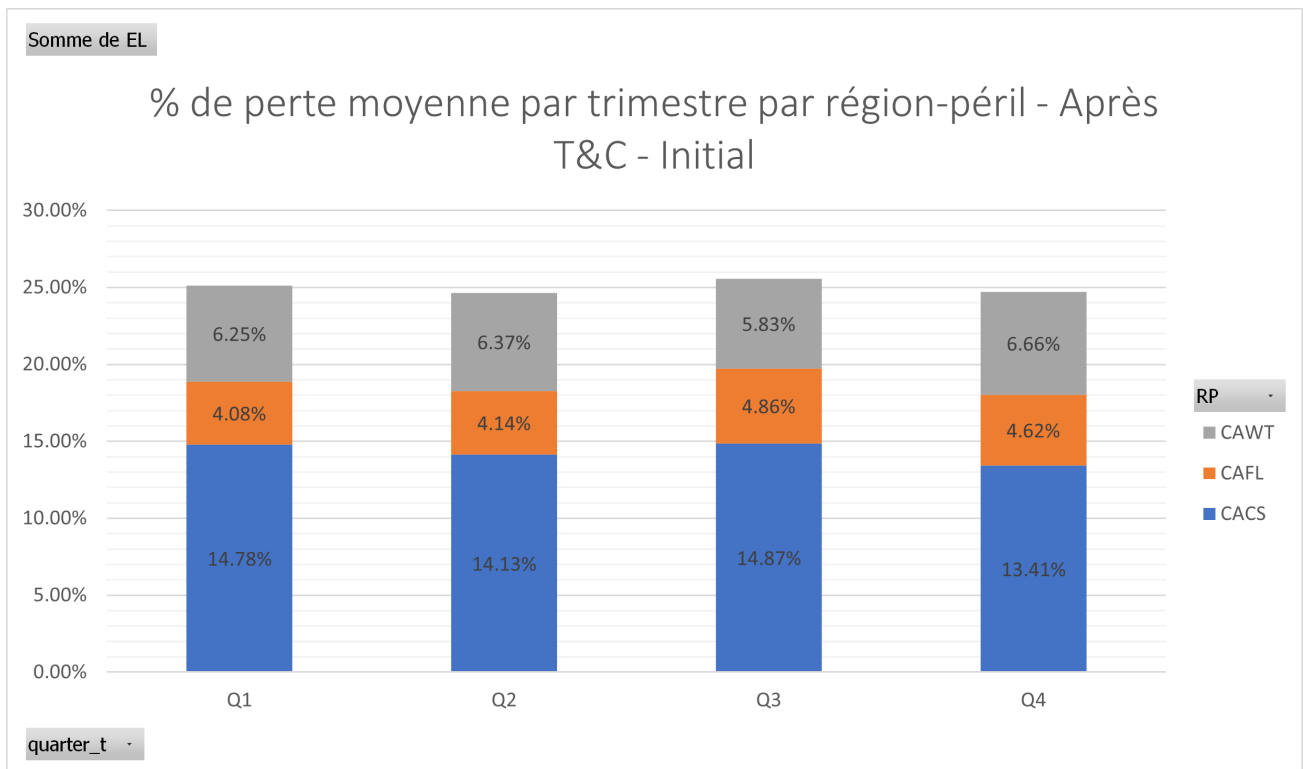


FIGURE 3.10 – Distribution initiale des pertes sur l’année par *Region Peril* après T&C

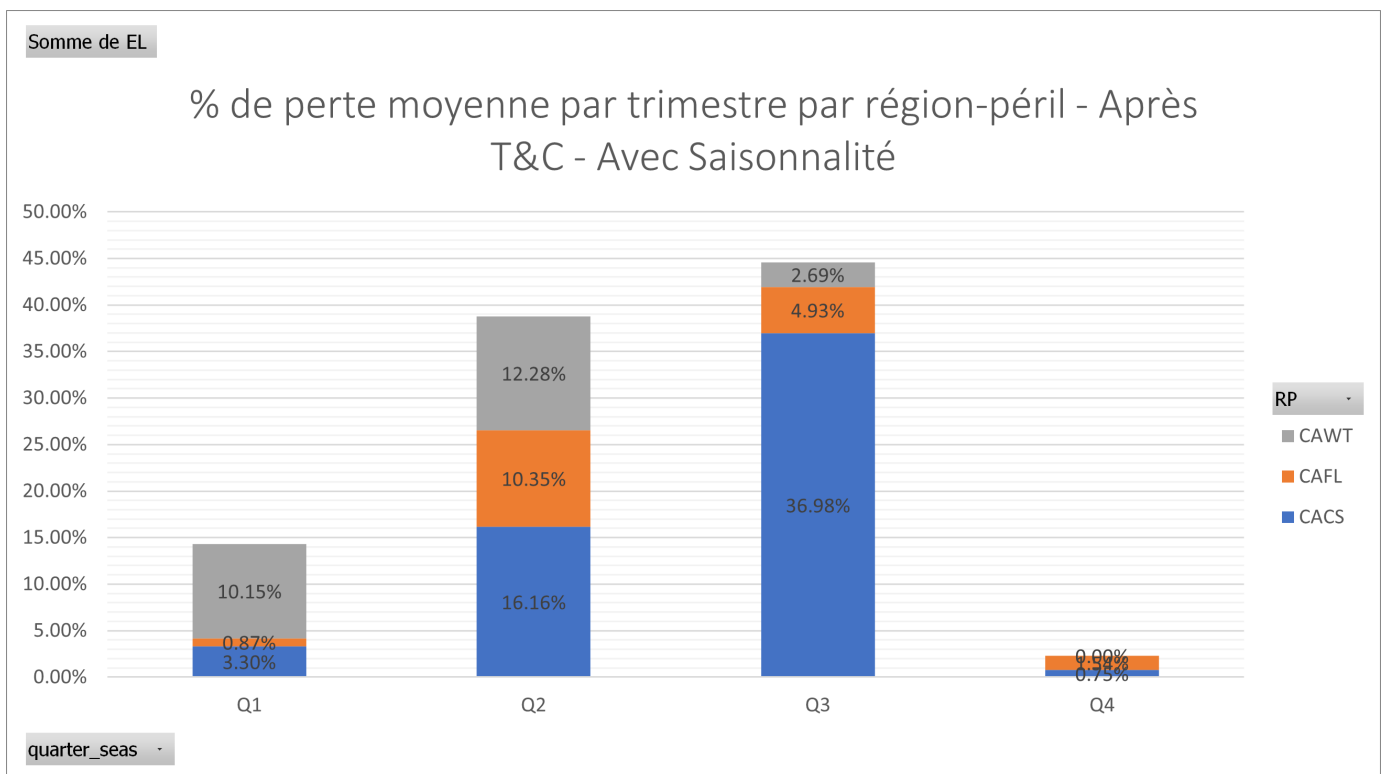


FIGURE 3.11 – Distribution de pertes sur l’année par *Region Peril* avec saisonnalité et après T&C

Après termes et conditions, on conserve presque la même distribution des pertes sur l’année que ce qu’on a observé dans la partie 3.4.1 pour les deux cas avec et sans saisonnalité. Cependant, le changement se voit au niveau de la contribution des périls aux pertes du trimestre. En effet, la contribution des pertes causées par la grêle (CACS) diminue et celle des tempêtes de

neige augmente (CAWT). Ceci nous permet de dire qu'une partie des pertes du CS n'atteint pas la franchise et n'est pas couverte par ce contrat alors que le péril WT est celui qui engendre des pertes plus sévères.

3.4.3 Conclusions

On remarque d'après les deux cas, que l'introduction de la saisonnalité dans la modélisation des *Region Perils* impacte la distribution des pertes sur l'année. Ceci permet d'avoir une meilleure visibilité sur l'exposition du portefeuille sur l'année et devient très intéressant surtout quand il s'agit de tarifier un contrat d'une durée qui n'est pas annuelle ou pluriannuelle (couverture sur un nombre de mois non divisible par 12).

Ci-dessous, les pertes trimestrielles moyennes du contrat après l'application des termes et conditions :

EL in Q1	114 099
EL in Q2	309 118
EL in Q3	355 340
EL in Q4	18 284

TABLE 3.4 – Pertes trimestrielles (en euros)

En prenant en compte la saisonnalité, on voit que la prime technique varie considérablement en fonction de la période de couverture. En effet, celle-ci vaut moins de 150 000 euros pour une couverture sur Q4-Q1 et s'élève à environ 750 000 euros si la couverture est pour Q2-Q3.

L'ajout de la saisonnalité dans la modélisation des *Region Perils* synthétiques montre que la prime pure ne varie pas linéairement sur l'année ainsi, en la considérant, on pourrait optimiser la vue du risque pour la cédante et pour le (ré)assureur en adaptant la couverture et la prime en fonction de la période à couvrir.

De plus, SCOR étant une société cotée en bourse, elle se doit de présenter des résultats de manière trimestrielle. La société peut donc comparer son résultat modélisé moyen au résultat obtenu pour chaque trimestre ce qui lui donne un nouvel indicateur sur ses propres résultats.

Chapitre 4

Étude de la corrélation

Introduction

La modélisation des catastrophes naturelles se fait moyennant des YEQTs dont les évènements sont spécifiques au *Region Peril* étudié. Ainsi, au moment de l'accumulation, en agrégeant les pertes générées par différents *Region Perils*, on effectue une hypothèse forte qui est l'indépendance entre les pertes générées par deux *Region Perils* différents.

Cette modélisation engendre une sous-estimation du risque et des pertes issues des catastrophes naturelles dans le cas où certains sont réellement dépendants.

Dans ce chapitre, on vérifie s'il existe des *Region Perils* corrélés dans l'optique d'introduire cette structure de dépendance dans le portefeuille SCOR.

4.1 Corrélation inter *Region Perils* dans le monde

4.1.1 Principe préliminaire

L'étude est effectuée au niveau des clés comme définies dans 3.1.3. On se sert des données historiques des sinistres de la base de données CRED (présentée dans la partie 3.1.2) pour explorer la corrélation des pertes annuelles causées par les différentes clés. L'historique utilisé s'étale sur 32 ans (de 1990 à 2021).

On commence notre étude de corrélation en se concentrant plutôt sur les couples de clés qui présentent des dépendances aux extrêmes.

Pour ce faire, on calcule les pertes historiques annuelles causées par les catastrophes naturelles pour chaque clé qu'on classe ensuite par ordre décroissant de façon que l'année avec la perte la plus importante pour une clé ait le rang 1.

Étant donné le faible nombre d'années d'historiques, on considère comme observation extrême les 25% évènements les plus sévères pour chaque clé, ce qui donne, au meilleur cas où on a 32 observations, 8 évènements extrêmes. Ensuite, on trace des dépendogrammes de rangs des pertes pour chaque couple de clés. On obtient des figures comme dans l'exemple 4.1.

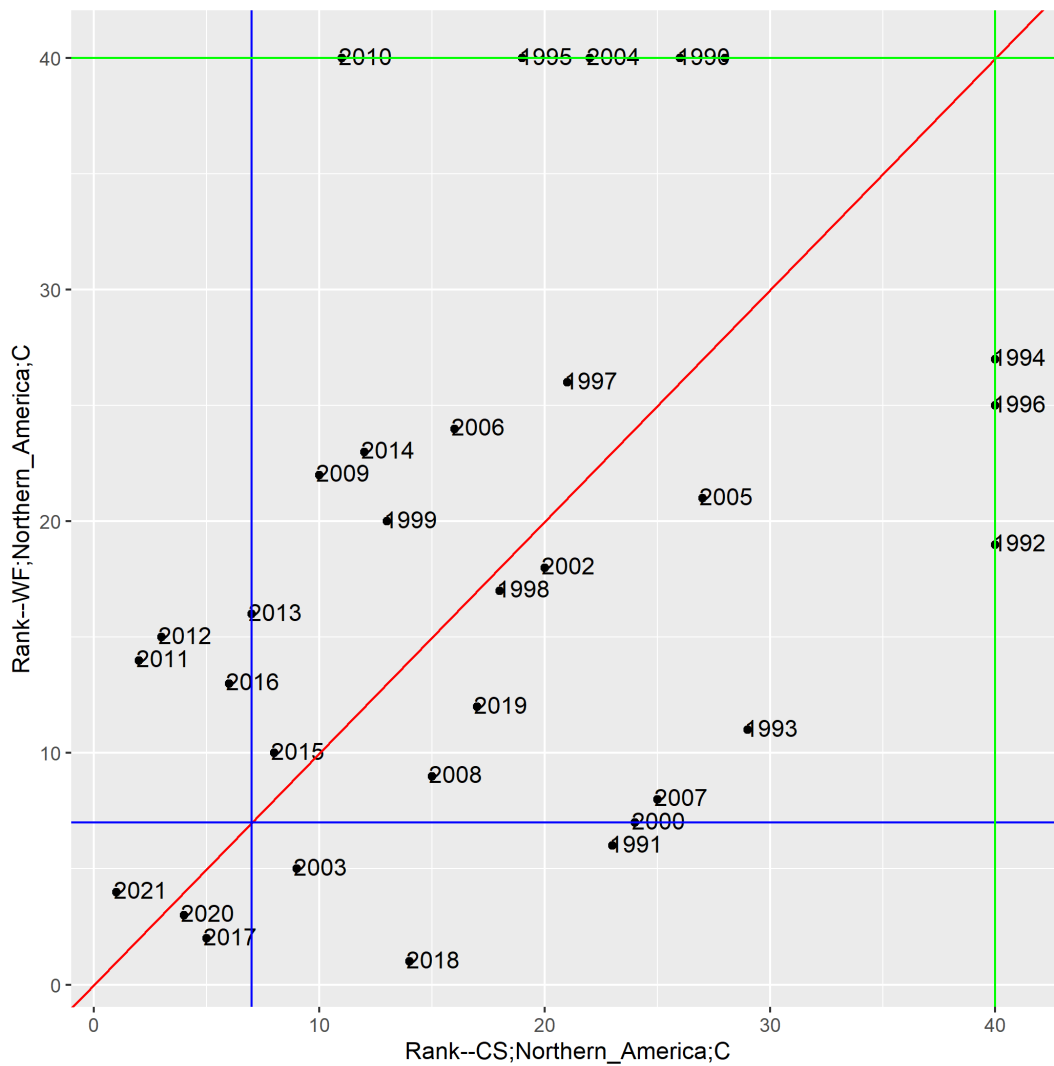


FIGURE 4.1 – Nuage des points des rangs d’années d’évènements pour la grêle et les inondations aux pays de l’Amérique du nord au climat Tempéré

Afin d’évaluer la corrélation aux extrêmes, on définit comme métrique le pourcentage des points qui figurent dans le cadran extrême (le coin bas à gauche) par rapport aux couples d’observations annuelles.

Puisqu’on a considéré comme extrême les 25% observations de rangs les plus faibles, en partageant le plan uniformément, on se retrouve avec un total de 16 cadrans. Dans une hypothèse d’indépendance parfaite, on retrouverait que 6.25% des couples d’observations dans chaque cadran.

Cette métrique nous permet alors de filtrer les couples de clés qui peuvent présenter une dépendance aux extrêmes en choisissant ceux dont cette métrique définie dépasse ce seuil de 6.25%.

4.1.2 Résultats

On calcule la métrique de dépendance aux extrêmes pour tous les couples de clés parmi lesquels on sélectionne les couples de clés pertinents :

1. On exclut les couples qui ne dépassent pas le seuil de dépendance ou ceux qui ont des dépendances difficilement interprétables. En effet, on commence par regarder les couples de clés qui ont au moins un paramètre de clé commun (même péril, même sous-continent

ou même type de climat ou même couple de ces paramètres).

2. On sélectionne les couples auxquels SCOR est exposée.

Parmi plus de 10 000 couples de clés possibles, on se retrouve avec 9 couples de clés sélectionnés :

Clé1	Clé2	nb_occurrence commune	Métrique_corrélation
TC;Eastern_Asia;B	TC;South-eastern_Asia;C	27	7%
CS;Northern_America;C	WF;Northern_America;C	24	13%
TC;Eastern_Asia;C	TC;South-eastern_Asia;C	22	9%
ET;Western_Europe;C	ET;Western_Europe;D	10	20%
ET;Northern_Europe;C	ET;Western_Europe;C	8	25%
ET;Northern_Europe;C	ET;Western_Europe;D	7	29%
CS;Northern_America;C	CS;Southern_Europe;C	7	14%
EQ;Southern_Europe;C	EQ;Western_Asia;C	7	14%

TABLE 4.1 – Couples de clé retenus avec dépendance aux extrêmes

Pour une meilleure visualisation, on représente ces corrélations sur une carte du monde pour observer leur dispersion géographique :



FIGURE 4.2 – Dispersion géographique des *Region Perils* dépendants

On peut observer l'apparition de clusters de dépendance tels que les périls causés par le vent en Europe, les cyclones tropicaux en Asie.

L'implémentation de ces structures de dépendances entre ces groupes observés peut avoir un impact énorme vu que ces dépendances seront introduites au niveau des *Region Perils* et ainsi on se retrouve à imposer des dépendances entre plusieurs *Region Perils* avec une étude effectuée avec très peu d'observations. On se propose ainsi, pour évaluer la dépendance et son impact, de limiter l'étude géographiquement.

4.2 Corrélation inter périls aux États-Unis

Dans la suite de notre étude des corrélations, on a choisi de se concentrer sur les corrélations entre les périls différents au sein des États des États-Unis. En effet, le marché de réassurance aux États-Unis est segmenté, de façon que les cédantes de chaque État couvrent essentiellement des risques géographiquement localisés dans le même État. Ainsi, une étude de la corrélation par État permet de prendre des décisions par rapport à la couverture offerte pour les différentes cédantes.

Données utilisées

Afin d'effectuer notre étude sur la corrélation entre les périls aux États-Unis, on dispose d'une base de données dans laquelle on retrouve l'historique de pertes indexées causées par des sinistres catastrophes naturelles avec répartition de la perte totale sur chaque État touché.

Péril	Nombre de sinistres historiques
CS	1402
WT	238
TC	119
WF	55
EQ	15
FL	10
VC	2
Total	1841

TABLE 4.2 – Nombre de sinistres historiques par péril

L'historique s'étale de 1968 à 2022.

On compte 52 États dans les États-Unis pour lesquels on agrège les pertes par localisation géographique sous 10 groupes comme présentés dans la carte suivante :

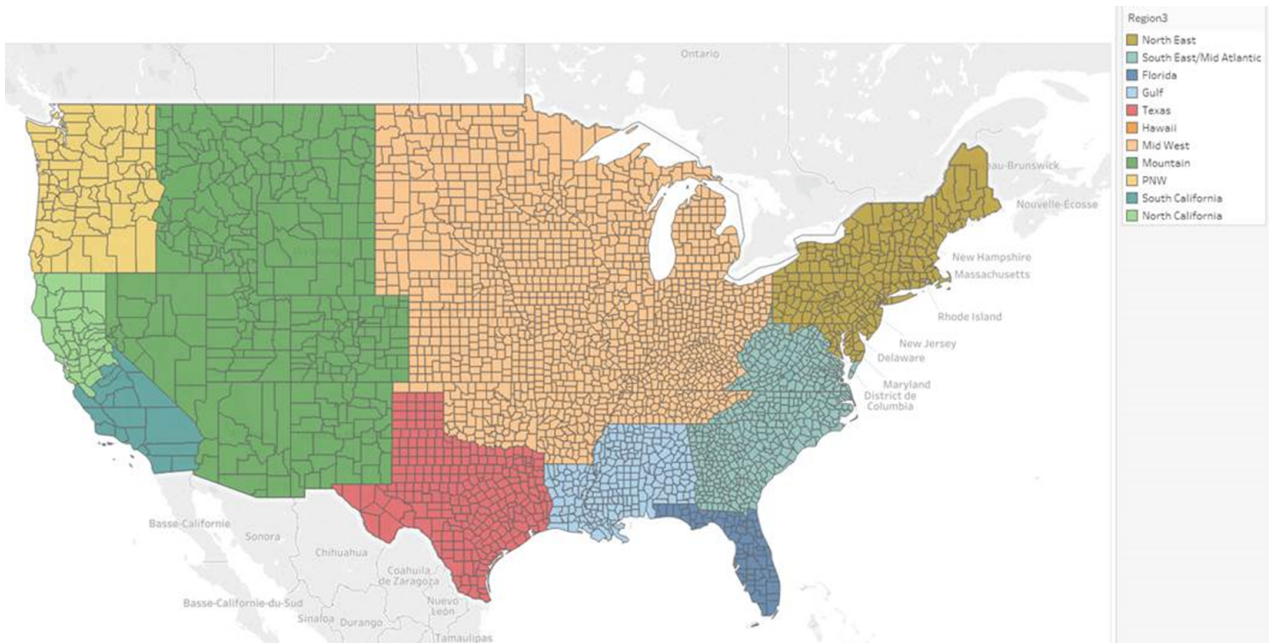


FIGURE 4.3 – Groupes des États aux États-Unis

À ce niveau, on utilise ces données de pertes annuelles indexées par péril pour chaque groupe d'États pour évaluer la corrélation entre périls différents au sein d'un même État.

La corrélation entre les événements causés par une même classe de sinistres cat-nat entre différents groupes d'États n'est pas le sujet de notre étude puisque celle-ci est systématiquement intégrée du fait que la modélisation des périls aux États-Unis se fait au niveau du pays.

4.3 Théorie des copules

L'objectif de ce chapitre est de détecter les couples qui sont corrélés et d'introduire cette corrélation dans la modélisation, à savoir, corréler les années de survenance des pertes. Pour ce faire, on opte pour la modélisation de cette corrélation avec les copules. C'est un outil fréquemment utilisé en gestion des risques afin de décrire complètement la dépendance entre deux risques. On commence donc par introduire des notions fondamentales de la théorie de copules bivariées et les méthodes d'estimation.

4.3.1 Copules bivariées

Définition 2 (Copule bivariée). *Une copule C bivariée est la fonction de répartition d'un vecteur de variables aléatoires $U = (U_1, U_2)$ dont les lois marginales sont des lois uniformes sur $[0, 1]$.*

Avec les copules, on ne travaille plus en nombre ou en montant mais en rang. Autrement dit, dans la pratique on transforme linéairement les réalisations x_1, \dots, x_n en pseudo-observations u_1, \dots, u_n où $u_i = \frac{\text{Rang}(x_i)}{n+1}$ pour tout i de 1 à n .

Théorème 1 (Théorème de Sklar). *Soit $H : \overline{\mathbb{R}}^2 \rightarrow \mathbb{R}$ une fonction de répartition de marges F_1, F_2 . Alors il existe une copule C telle que :*

$$\forall x \in \mathbb{R}^2, H(x_1, x_2) = C(F_1(x_1), F_2(x_2))$$

De plus, si F_1 et F_2 sont continues, alors C est unique. Sinon, C est uniquement déterminée sur $S \text{ upp}(F_1) \times \text{Supp}(F_2)$ avec $S \text{ upp}(F_1)$ le support de F_1 .

Inversement, si C est une copule et si F_1 et F_2 sont des fonctions de répartition, alors la fonction définie par :

$$F(x_1, x_2) = C(F_1(x_1), F_2(x_2))$$

est une fonction de répartition bivariée avec les fonctions de répartition marginales F_1 et F_2 .

Le théorème de Sklar est un théorème fondamental de la théorie des copules. En effet, il permet de dire que toutes les dépendances entre deux variables aléatoires sont contenues dans une copule tandis que la loi marginale modélise la distribution de chaque variable aléatoire.

Densité de copule

La fonction de densité associée à une copule C est définie par :

$$c(u_1, u_2) = \frac{\partial^2}{\partial u_1 \partial u_2} C(u_1, u_2)$$

Dans le cas continu, pour la fonction de répartition bivariée d'un vecteur $X = (X_1, X_2)$ avec

$$F_{X_1, X_2}(x_1, x_2) = C(F_{X_1}(x_1), F_{X_2}(x_2))$$

on a la fonction de densité associée :

$$f_{X_1, X_2}(x_1, x_2) = c(F_{X_1}(x_1), F_{X_2}(x_2)) \cdot f_{X_1}(x_1) \cdot f_{X_2}(x_2)$$

où $f_{X_i}(x_i)$ est la fonction de densité de X_i : $f_{X_i}(x_i) = \frac{dF_{X_i}}{dx_i}(x_i)$

Bornes de Fréchet

Soit C une copule définie sur $[0, 1]^2$. Alors la copule C est encadrée par les bornes de Fréchet :

$$C^-(u_1, u_2) \leq C(u_1, u_2) \leq C^+(u_1, u_2)$$

avec :

$$\begin{cases} C^-(u_1, u_2) = \max(u_1 + u_2 - 1; 0) & : \text{borne inférieure de Fréchet} \\ C^+(u_1, u_2) = \min(u_1; u_2) & : \text{borne supérieure de Fréchet} \end{cases}$$

Copule de survie

Soit $C(u_1, u_2)$ une copule définie sur $[0, 1]^2$. On peut alors définir une copule de survie $\bar{C}(u_1, u_2)$, associée à la copule C et reliant les fonctions de survie marginales \bar{F}_{X_1} et \bar{F}_{X_2} selon l'expression :

$$\bar{F}_{X_1, X_2}(x_1, x_2) = \bar{C}(\bar{F}_{X_1}, \bar{F}_{X_2})$$

On peut expliciter cette expression comme suit :

$$\mathbb{P}(U_1 > u_1, U_2 > u_2) = \bar{C}(1 - u_1, 1 - u_2)$$

Enfin, nous avons la formule suivante liant la copule de survie \bar{C} à la copule initiale C :

$$\bar{C}(u_1, u_2) = C(1 - u_1, 1 - u_2) + u_1 + u_2 - 1$$

4.3.2 Mesures de dépendance

Dans cette partie, on définit les métriques qui permettent de quantifier la dépendance entre deux variables aléatoires.

On commence par définir les propriétés d'une mesure de dépendance :

Propriétés 1. Une mesure de dépendance $\pi(X_1, X_2)$ entre les variables X_1 et X_2 doit posséder les propriétés suivantes :

1. Symétrie : $\pi(X_1, X_2) = \pi(X_2, X_1)$
2. Normalisation : $-1 \leq \pi(X_1, X_2) \leq 1$
3. Comonotonie : $\pi(X_1, X_2) = 1$ si et seulement si X_1 et X_2 sont comonotones
4. Antimonotonie : $\pi(X_1, X_2) = -1$ si et seulement si X_1 et $-X_2$ sont comonotones
5. Invariance : pour toute fonction strictement croissante $\Psi : \mathbb{R} \rightarrow \mathbb{R}$, on a :

$$\pi(\Psi(X_1), X_2) = \begin{cases} \Psi(X_1, X_2) & \text{si } \Psi \text{ est croissante} \\ -\Psi(X_1, X_2) & \text{si } \Psi \text{ est décroissante} \end{cases}$$

Coefficient de corrélation linéaire

Définition 3 (Coefficient linéaire de Pearson). Pour deux variables aléatoires X et Y à valeurs dans \mathbb{R} , le coefficient de corrélation linéaire de Pearson est défini par :

$$\rho(X, Y) = \frac{\mathbb{E}[X \cdot Y] - \mathbb{E}[X] \cdot \mathbb{E}[Y]}{\sigma(X) \cdot \sigma(Y)}$$

Propriétés 2. Soient X et Y deux variables aléatoires à valeurs dans \mathbb{R} . On a les propriétés suivantes :

1. $\rho(X, Y) \in [-1, 1]$
2. $\rho(X, Y) = 1 \Leftrightarrow \exists a \geq 0, Y = a \cdot X + b$ ps
3. $\rho(X, Y) = -1 \Leftrightarrow \exists a \leq 0, Y = a \cdot X + b$ ps
4. Si X et Y sont indépendants, alors ils sont décorrélés : $\rho(X, Y) = 0$.

La réciproque de la propriété 4 est fautive sauf dans le cas où les variables sont gaussiennes.

Le coefficient de corrélation linéaire n'est pas indépendant par transformations monotones des lois marginales : $\forall X \in \mathbb{R}^+, \rho(X, Y) \neq \rho(X^2, Y)$. Ainsi, cette mesure ne respecte pas la propriété d'invariance.

On introduit dans ce qui suit des mesures de dépendance alternatives à savoir le Rho de Spearman et le Tau de Kendall. Elles permettent de mesurer la corrélation entre les rangs des observations, à la différence du coefficient de corrélation linéaire qui lui apprécie la corrélation entre les valeurs des observations. Elles offrent par ailleurs l'avantage de s'exprimer simplement en fonction de la copule associée au couple de variables aléatoires.

Rho de Spearman

Définition 4 (Rho de Spearman). Pour un couple (X, Y) de variables aléatoires, le Rho de Spearman est défini par :

$$\rho_S(X, Y) = 3 \times \mathbb{P}[(X_1 - X_2) \cdot (Y_1 - Y_2) > 0] - 3 \times \mathbb{P}[(X_1 - X_2) \cdot (Y_1 - Y_2) < 0]$$

où (X_2, Y_2) est indépendant de (X_1, Y_1) et possède les mêmes lois marginales.

Propriétés 3. *Propriétés.* Soit (X, Y) un couple de variables aléatoires. On a les propriétés suivantes :

1. $\rho_S(X, Y) \in [-1, 1]$
2. Si X et Y sont indépendantes, alors $\rho_S(X, Y) = 0$.
3. Si X et Y sont comonotones, alors $\rho_S(X, Y) = 1$.
4. Si X et Y sont antimonotones, alors $\rho_S(X, Y) = -1$.
5. Soient Φ_1 et Φ_2 deux fonctions monotones (croissantes ou décroissantes). Alors,

$$\rho_S(\Phi_1(X), \Phi_2(Y)) = \rho_S(X, Y)$$

Ainsi le Rho de Spearman satisfait les propriétés de mesure de dépendance souhaitées.

Tau de Kendall

Définition 5 (Tau de Kendall). *Pour un couple (X, Y) de variables aléatoires, le Tau de Kendall est défini par :*

$$\tau(X, Y) = \mathbb{P}[(X_1 - X_2) \cdot (Y_1 - Y_2) \geq 0] - \mathbb{P}[(X_1 - X_2) \cdot (Y_1 - Y_2) \leq 0]$$

où (X_1, Y_1) et (X_2, Y_2) sont deux versions indépendantes du couple (X, Y) .

Cette métrique peut être exprimée en fonction de la concordance des paires : $\tau(X, Y) = \mathbb{P}[\text{Paires concordantes}] - \mathbb{P}[\text{Paires discordantes}]$

Propriétés 4. *Soit (X, Y) un couple de variables aléatoires. On a alors :*

1. $\tau(X, Y) \in [-1, 1]$.
2. $X \perp Y \rightarrow \tau(X, Y) = 0$.
3. Si X et Y sont comonotones, alors $\tau(X, Y) = 1$.
4. Si X et Y sont antimonotones, alors $\tau(X, Y) = -1$.
5. Soient Φ_1 et Φ_2 deux fonctions monotones (croissantes ou décroissantes). Alors,

$$\tau(\Phi_1(X), \Phi_2(Y)) = \tau(X, Y)$$

4.3.3 Dépendances de queues

L'idée de dépendance de queue fournit une description de la dépendance au niveau des queues de distribution, ce qui est très intéressant dans l'étude de l'occurrence simultanée de valeurs extrêmes. Il s'agit d'une mesure locale, contrairement au tau de Kendall et au rho de Spearman, qui mesurent la dépendance tout au long de la distribution.

Définition 6 (Coefficient de dépendance de queue supérieure). *Pour un couple de variables aléatoires (X, Y) , le coefficient de dépendance de queue supérieure est défini, sous conditions d'existence, par :*

$$\lambda_U(X, Y) = \lim_{\kappa \rightarrow 1^-} \mathbb{P}(X > F_X^{-1}(\kappa) \mid Y > F_Y^{-1}(\kappa))$$

Définition 7 (Coefficient de dépendance de queue inférieure). *Pour un couple de variables aléatoires (X, Y) , le coefficient de dépendance de queue inférieure est défini, sous conditions d'existence, par :*

$$\lambda_L(X, Y) = \lim_{\kappa \rightarrow 0^+} \mathbb{P}(X < F_X^{-1}(\kappa) \mid Y < F_Y^{-1}(\kappa))$$

4.3.4 Copules archimédiennes

On se propose de présenter la famille de copules archimédiennes dans ce mémoire vu qu'elles sont mieux adaptées à notre étude. En effet, contrairement aux copules elliptiques et à la copule gaussienne, les copules archimédiennes permettent de modéliser des structures de dépendance très variées à savoir les dépendances asymétriques (coefficients de dépendance de queue supérieure et inférieure différents).

Définition 8 (Générateur strict). *On appelle générateur strict une fonction $h : [0, 1] \rightarrow \overline{\mathbb{R}}$ telle que :*

- h est strictement décroissante
- h est continue
- $h(1) = 0$
- $\lim_{x \rightarrow 0} h(x) = +\infty$

Définition 9 (Copule archimédienne). *On appelle copule archimédienne de générateur h la copule $C : [0, 1]^n \rightarrow [0, 1]$ définie par $C(u_1, \dots, u_n) = h^{-1}(h(u_1) + \dots + h(u_n))$*

Copule	$h(u)$	$C(u, v)$	λ_L	λ_U	τ_a
Gumbel	$(-\ln u)^a, a \geq 1$	$\exp\left(-\left[(-\ln u)^a + (-\ln v)^a\right]^{1/a}\right)$	0	$2 - 2^{1/a}$	$1 - 1/a$
Franck	$-\ln\left(\frac{e^{-au}-1}{e^{-a}-1}\right), a \neq 0$	$-\frac{1}{a} \ln\left[1 + \frac{(e^{-au}-1)(e^{-av}-1)}{(e^{-a}-1)}\right]$	0	0	$1 - \frac{4(1-D_1(a))}{a}$
Clayton	$\frac{(u^{-a}-1)}{a}, a > 0$	$(u^{-a} + v^{-a} - 1)^{\frac{1}{a}}$	$2^{-\frac{1}{a}}$	0	$\frac{a}{a+2}$

TABLE 4.3 – Copules archimédiennes

La copule de Gumbel ne capture que les dépendances positives et a la propriété de représenter les risques avec une structure de dépendance accentuée dans la queue supérieure.

La copule de Franck permet de modéliser à la fois les dépendances positives et négatives. On remarque aussi que cette copule n'a pas de dépendance de queue.

Comme la copule de Gumbel, la copule de Clayton ne peut modéliser que des dépendances positives. Contrairement à la copule de Gumbel, elle prévoit de tenir compte de la dépendance aux événements de faible intensité.

4.3.5 Inférence statistique

On compte plusieurs méthodes fréquemment utilisées pour l'estimation du paramètre de la copule : méthode des moments, méthode du maximum de vraisemblance exact, inférence sur les marginales et méthode du maximum de vraisemblance canonique...

On va présenter dans cette partie les méthodes basées sur le maximum de vraisemblance qu'on utilisera dans la suite.

Méthode du maximum de vraisemblance

En se basant sur le théorème de Sklar, la densité jointe de la distribution F est donnée par l'expression :

$$f(x_1, \dots, x_d) = c(F_1(x_1), \dots, F_d(x_d)) \prod_{i=1}^d f_i(x_i)$$

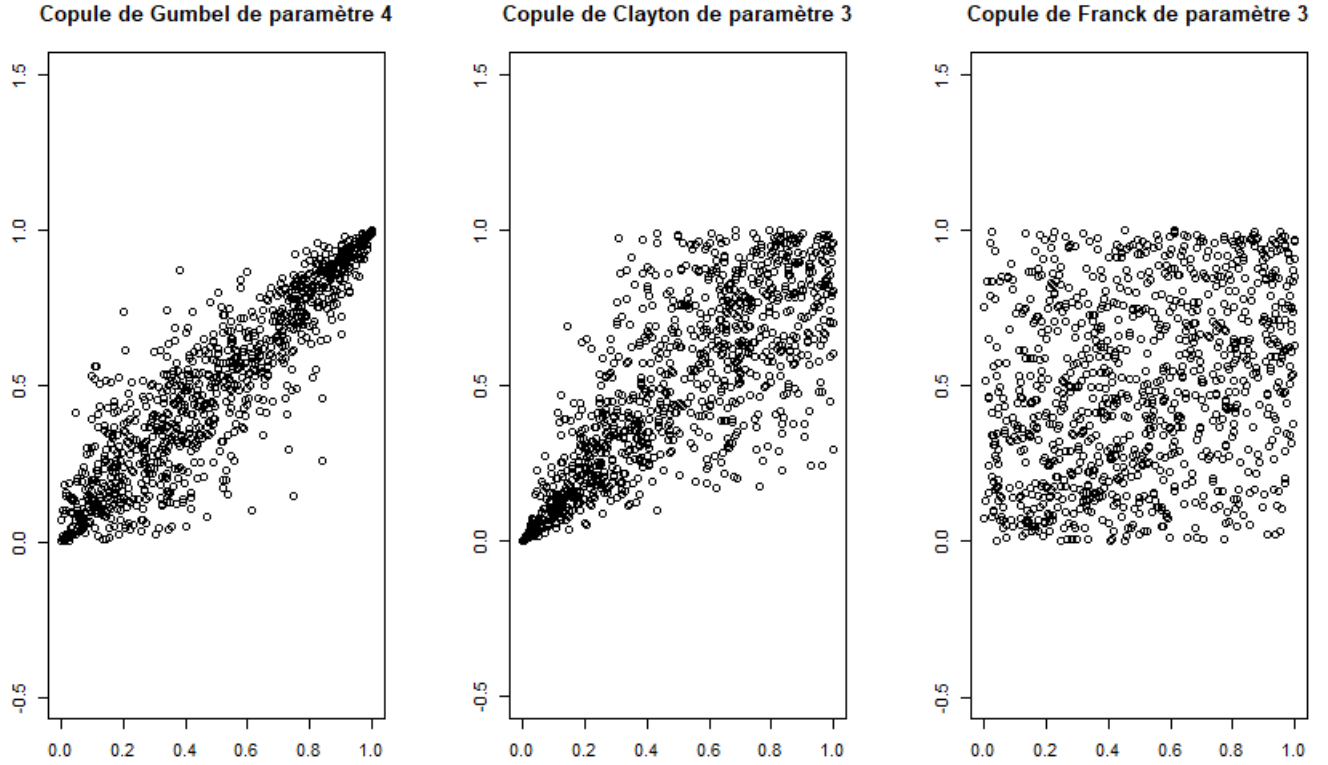


FIGURE 4.4 – Exemples de dépendogrammes de copules archimédiennes

On dispose des réalisations $(X_1^{(1)}, \dots, X_d^{(1)}), \dots, (X_1^{(n)}, \dots, X_d^{(n)})$ et on cherche à estimer $\theta \in \Theta$, la log-vraisemblance est donnée par :

$$l(\theta) = \sum_{t=1}^n \ln c(F_1(x_1^t), \dots, F_d(x_d^t)) + \sum_{t=1}^n \sum_{k=1}^d \ln f_k(x_k^t)$$

On note $\hat{\theta}_{ML}$ = l'estimateur de θ tel que $l(\hat{\theta}_{ML}) \geq l(\theta), \forall \theta \in \Theta$ et qui vérifie la propriété de normalité asymptotique : $\sqrt{T}(\hat{\theta}_{ML} - \theta_0) \xrightarrow{L} N(0, I^{-1}(\theta_0))$, avec $I(\theta_0)$ la matrice d'information de Fisher.

Cependant, cette méthode peut engendrer des temps de calcul longs si on est en grande dimension puisqu'elle estime parallèlement les paramètres des lois marginales et les paramètres de la structure de dépendance. De plus, l'estimation de la copule est sensible à une éventuelle erreur d'estimation des marginales car elles sont directement liées au calcul de la log-vraisemblance.

Méthode du maximum de vraisemblance canonique (CML)

Cette méthode ne nécessite pas d'avoir recours à l'estimation des lois marginales. En effet, on commence par transformer les observations $(x_d^i)_{1 \leq i \leq n}$ disponibles en pseudo-observations $(u_d^i)_{1 \leq i \leq n} = \frac{\text{Rang}(x_d^i)}{n+1}$ et dont la fonction de répartition \hat{F}_d sont tels que :

$$\hat{F}_d(x) = \frac{1}{n+1} \sum_{i=1}^n \mathbf{1}(X_d^{(i)} \leq x).$$

Ensuite, on applique la méthode de maximum de vraisemblance pour déterminer $\hat{\theta}$ en remplaçant les fonctions de répartition par les fonctions de répartition empiriques \hat{F}_d .

Critères quantitatifs de validation de copules

La sélection de copules doit être relative à l'élément modélisé. Ainsi, lorsqu'on essaie de modéliser un événement extrême, il faut se concentrer sur la famille des copules extrêmes. Cela supprime la copule gaussienne.

Dans le cas d'une famille spécifique, différents critères statistiques peuvent être comparés, soit pour confirmer la copule, soit non.

Les critères associés au log de vraisemblance sont : la log-vraisemblance $\ln \mathcal{L}$, critère d'Aikake (AIC) $2k - 2 \ln \mathcal{L}$ ou critère de Schwarz (BIC) $-2 \ln \mathcal{L} + k \ln n$ où k est le nombre de paramètres à estimer et n est la taille de l'échantillon.

4.4 Application sur les périls des États-Unis

On dispose des pertes annuelles historiques indexées pour chaque péril dans les groupes d'États comme présentés dans la partie des données (cf. 4.3). On souhaite modéliser la dépendance entre deux périls différents dans un même groupe d'États moyennant les copules archimédiennes. On commencera par choisir les couples de périls appartenant à un même groupe qui sont pertinents pour notre étude. Ensuite, on détaillera l'implémentation et on mesurera l'impact de prendre en compte cette dépendance sur la couverture des risques.

4.4.1 Choix des couples pertinents

On calcule le rho de Spearman pour chaque couple (49 couples disponibles) ainsi que le nombre d'années de survénance de sinistres pour chaque péril dans un groupe d'États. On calcule aussi la métrique de corrélation aux extrêmes qu'on a définie au début du chapitre (cf. 4.1.1).

En prenant en compte ces critères de choix, on se propose de concentrer notre étude sur deux couples de périls dans les États du *South East/Mid Atlantic* à savoir :

- CS et TC : la grêle et le cyclone tropical.
- CS et WT : la grêle et la tempête de neige.

Ce groupe *South East/Mid Atlantic* réunit 5 différents États : Georgia, North Carolina, South Carolina, Virginia et West Virginia.

<i>South East/Mid Atlantic</i>	Couple 1	Couple 2
peril_x	CS	CS
peril_y	TC	WT
rho de spearman	38%	5%
nombre d'occurrences de x	53	53
nombre d'occurrences de y	30	31

TABLE 4.4 – Couples d'étude de dépendance

L'étude des ces périls dans le groupe *South East/Mid Atlantic* est intéressante vu que cette région représente la 3^{ème} plus grande perte historique par rapport aux autres groupes d'États en considérant les pertes causées par des événements des 3 périls TC, CS et WT.

D'après la table du coefficient de corrélation des rangs de Spearman de deux variables aléatoires et pour un échantillon de 55 années d'observations avec un seuil 5% bilatéral, les valeurs critiques du rho de spearman est $r_s = \pm 0.266$ [13].

Ainsi, la valeur du rho du Spearman pour le couple 1 laisse apparaître une dépendance significative entre les deux périls, ce qui n'est pas le cas pour le couple 2 dont la dépendance peut être due au hasard.

4.4.2 Estimation des copules

On utilise la bibliothèque *copula* de R afin de choisir la copule archimédienne la mieux représentative des dépendances entre les couples retenus et pour calculer le paramètre de copule associée.

On utilise la méthode CML pour estimer les paramètres de copules vu que ça réduit le nombre de paramètres à estimer en se séparant des marginales.

On calcule, pour chaque copule estimée les critères quantitatifs AIC, BIC et log-vraisemblance.

Couples de perils (x/y)	CS/TC	CS/WT
paramètre_Gumbel	1,3193	1,0440
AIC_Gumbel	-4,2050	1,8151
BIC_Gumbel	-2,1976	3,8225
Log-MV_Gumbel	3,1025	0,0924
Paramètre_Clayton	0,7567	0,2357
AIC_Clayton	-5,6866	1,1938
BIC_Clayton	-3,6792	3,2011
Log-MV_Clayton	3,8433	0,4031
Paramètre_Franck	2,5622	0,3330
AIC_Franck	-6,2569	1,8630
BIC_Franck	-4,2496	3,8703
Log-MV_Franck	4,1285	0,0685

TABLE 4.5 – Estimation des paramètres des copules et critères quantitatifs

On représente ci-dessous les dépendogrammes des pseudo-observations ainsi que les copules estimées pour chaque couple.

(**NB.** Les pseudo-observations qui s'alignent horizontalement représentent les années où on a une perte non nulle générée pour la clé de l'axe des abscisses et 0 perte pour la clé de l'axe des ordonnées.)

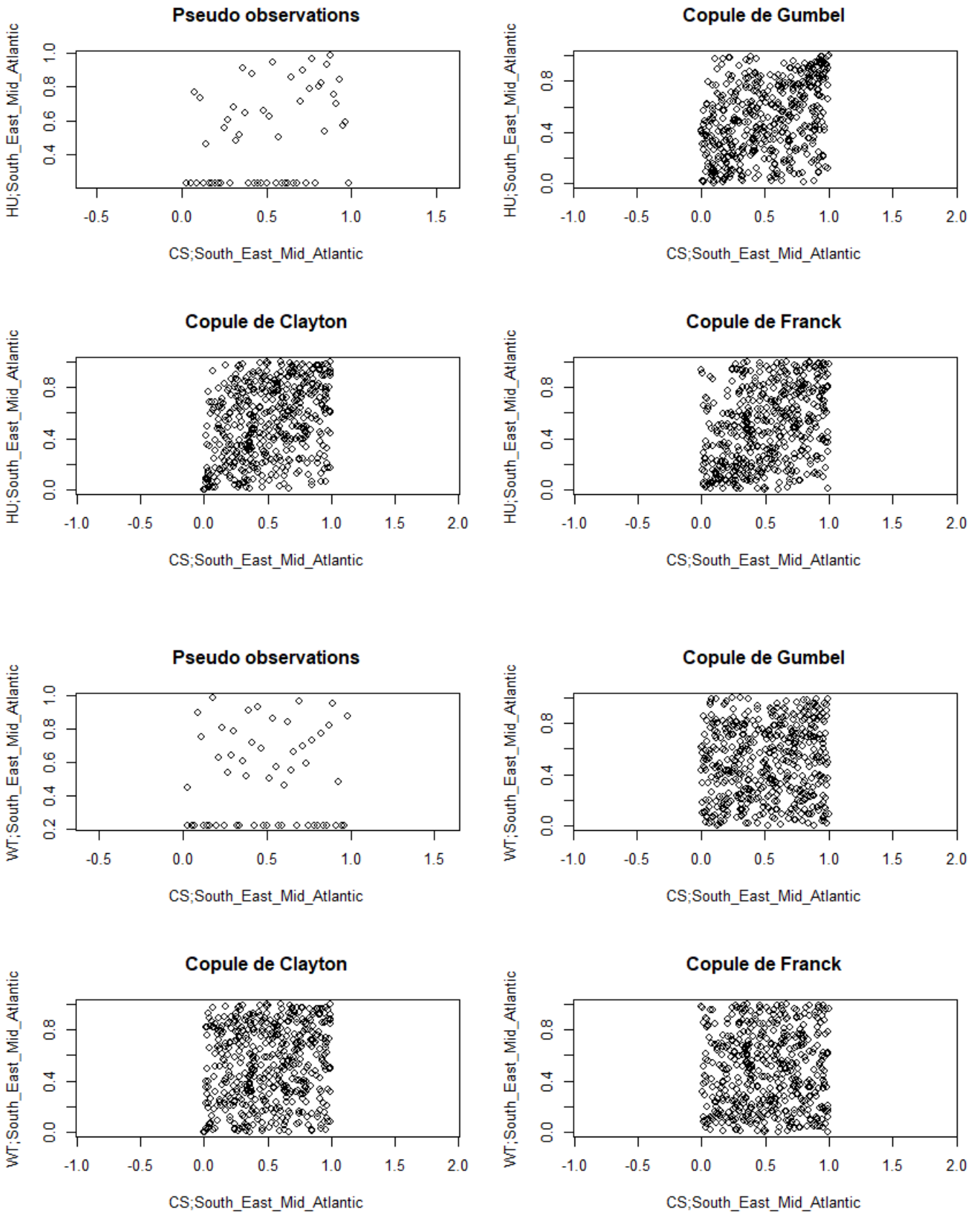


FIGURE 4.5 – Copules archimédiennes estimées à partir des observations

En se basant sur les critères quantitatifs, on peut retenir la copule Franck pour la modélisation de dépendance du premier couple (CS et TC) et celle de Clayton pour le deuxième couple (CS et WT).

On remarque que pour le deuxième couple, le paramètre de Clayton estimé est proche de 0 qui représente le cas de la copule indépendante.

Afin de valider les paramètres calculés à partir de 55 observations et de s'assurer qu'on n'a pas déduit la dépendance suite à un hasard des données, on effectue un test d'hypothèse.

Test d'hypothèse de l'indépendance

Afin de vérifier si la copule retenue ainsi que son paramètre estimé représentent un cas de dépendance entre les périls, on procède ainsi, pour chaque copule (Franck et Clayton) :

1. On simule $m = 100\,000$ copules indépendantes et ce en générant deux variables aléatoires uniformes indépendantes avec $N = 55$ observations (comme dans notre cas de données).
2. On estime le paramètre de copule (Franck et Clayton) à chaque scénario m .
3. On visualise la distribution des paramètres estimés pour chaque copule dans le cas indépendant.
4. On se fixe un seuil de confiance de 5% et on évalue la disposition du paramètre de copule retenu dans notre étude par rapport à la distribution du paramètre dans le cas continu.

On se retrouve avec cette distribution de paramètres de copules dans le cas indépendant :

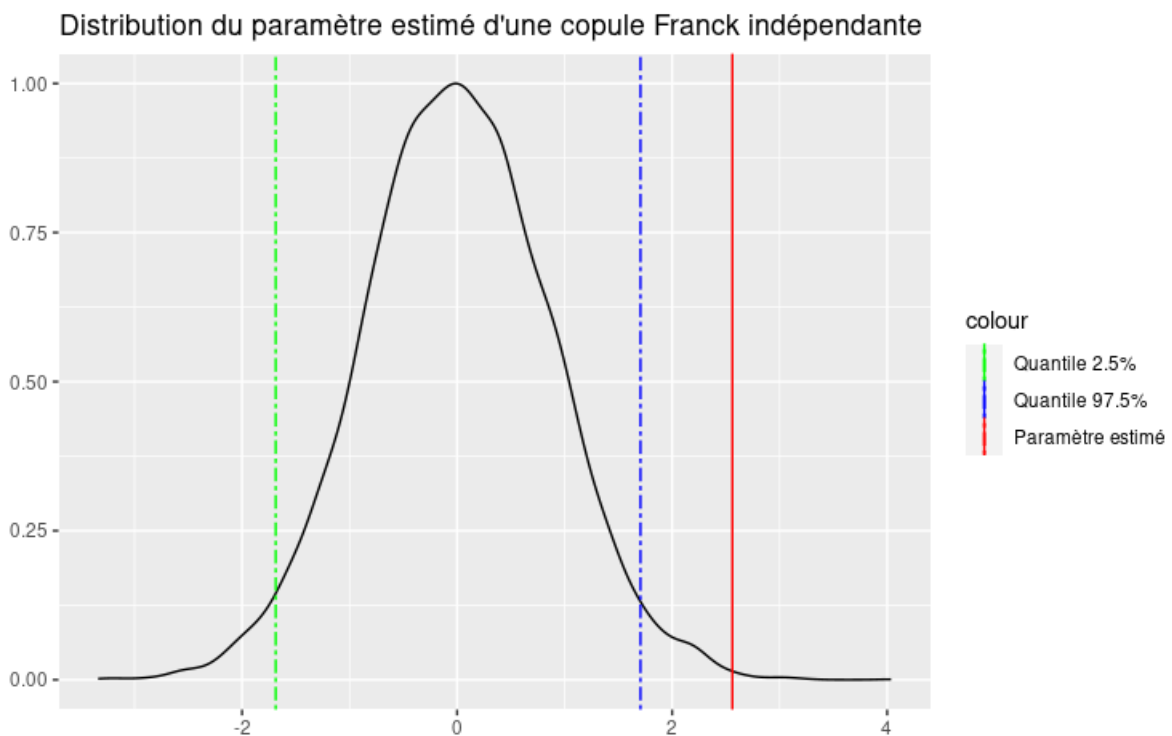


FIGURE 4.6 – Distribution du paramètre estimé d'une copule Franck indépendante avec $N = 55$ observations

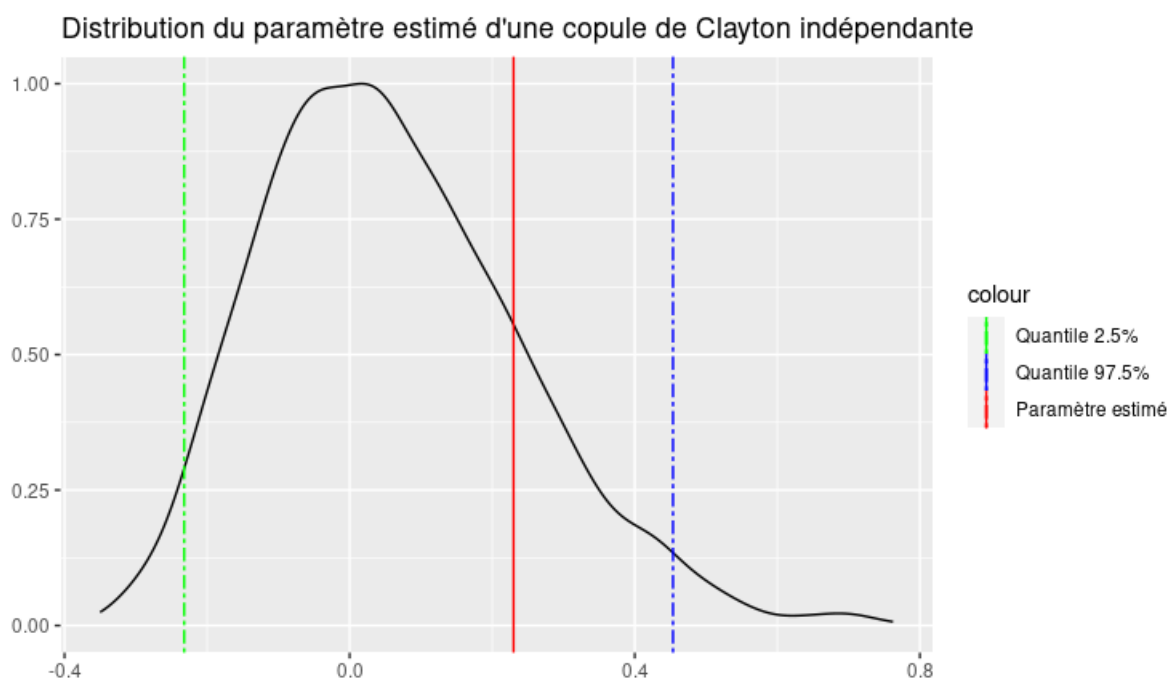


FIGURE 4.7 – Distribution du paramètre estimé d’une copule Clayton indépendante avec $N = 55$ observations

On voit bien que le paramètre estimé pour la copule Clayton pour modéliser la dépendance entre les périls du deuxième couple appartient à l’intervalle de confiance du paramètre d’une copule de Clayton dans le cas indépendant.

On rappelle que le rho de Spearman calculé pour le deuxième couple n’était pas significatif. On renonce ainsi à la modélisation de dépendance pour ce couple.

Pour le premier couple de périls TC et CS, on voit bien que le paramètre estimé de la copule Franck ne correspond pas au cas indépendant avec un seuil de confiance de 5%. On retient ainsi le paramètre de copule estimé et on procède à l’étude d’impact de l’insertion de cette dépendance sur la vue de risque.

4.4.3 Implémentation et étude d’impact

Afin de tester l’impact d’introduction de la dépendance estimée avec la copule Franck sur le couple de périls grêle (CS) et cyclone tropical (TC) aux États du groupe *South East/Mid Atlantic*, on dispose des YLTs par péril et par État avec des pertes qui représentent les expositions aux États-Unis. Ces YLTs sont fournis par le modélisateur RMS.

Pour ces deux périls, on commence par agréger les YLTs des États faisant partie du groupe *South East/Mid Atlantic* par année de simulation. On obtient ainsi deux catalogues d’évènements avec les pertes associées à chaque péril sur les $N = 100\,000$ années de simulation qu’on ordonne par ordre de perte décroissante.

On calcule le coefficient de corrélation linéaire et le rho de Spearman des pertes par année pour ce couple, et on vérifie bien qu’aucune dépendance ne peut être détectée (coefficients nuls ce qui montre qu’aucune dépendance n’était prise en compte dans la modélisation initiale).

L’objectif c’est de réattribuer les pertes aux années pour les deux YLTs de façon qu’ils respectent la dépendance qu’on a estimé avec la copule Franck. Pour ce faire [10] :

1. On simule $N = 100\,000$ d’observations d’un vecteur $Y = (X_1, X_2)$ tirées d’une copule de

Franck de paramètre $a = 2.5622$ en dimension 2.

```
param_a=2.562154
N=100000
frankCop=frankCopula(param_a, dim=2)
y=rCopula(N, copula=frankCop)
```

2. On obtient un tableau avec des nombres compris entre 0 et 1. Étant donné qu'on souhaite se rapporter à des rangs, on réarrange chaque colonne du vecteur d'observations de façon croissante et on attribue le rang (exemple dans le tableau 4.6).

observations X_1	observations X_1 réarrangées
2,1766E-05	2
3,6823E-05	4
2,1260E-06	1
5,8714E-05	7
2,7094E-05	3
1,1529E-04	10
6,5418E-05	8
4,2849E-05	5
7,0588E-05	9
5,8027E-05	6

TABLE 4.6 – Tableau qui illustre le réarrangement fait sur une colonne du vecteur y d'observations

3. On attribue un vecteur d'années de simulations $A = (1, \dots, 100\ 000)$ aléatoirement aux observations issues de la copule après réarrangement.

X_1	X_1 réarrangé	X_2	X_2 réarrangé	A
0,3077	30 934	0,4387	44 023	1
0,2576	25 877	0,6450	64 587	2
0,5523	55 457	0,2825	28 429	3
0,0563	5 628	0,8066	80 931	4
0,4685	47 163	0,1683	16 885	5
0,4837	48 680	0,6581	65 873	6
0,8124	81 177	0,3799	38 284	7
0,3703	37 246	0,2568	25 825	8
0,5465	54 876	0,5135	51 371	9
0,1702	17 010	0,4291	43 095	10

TABLE 4.7 – Réarrangement des observations et années de simulation

Ce tableau (4.7) est ensuite utilisé pour réarranger les pertes des YLTs initiaux tel que, dans l'année de simulation 1, on aura la perte de rang 30934 du premier péril et la perte de rang 44023 du deuxième péril.

4. On classe les pertes des YLTs marché des périls CS et TC de façon décroissante et on leur attribue un rang, tel que la plus grande perte ait le rang 1.
5. Pour chacun des YLTs initiaux, on remplace le vecteur année en fonction du rang de la perte de façon à ce que l'année de survenance de chaque perte respecte la structure de dépendance donnée par le vecteur y simulé avec la copule.

Exemple : la perte de rang 30934 dans le YLT du péril CS a lieu dans l'année de simulation numéro 1.

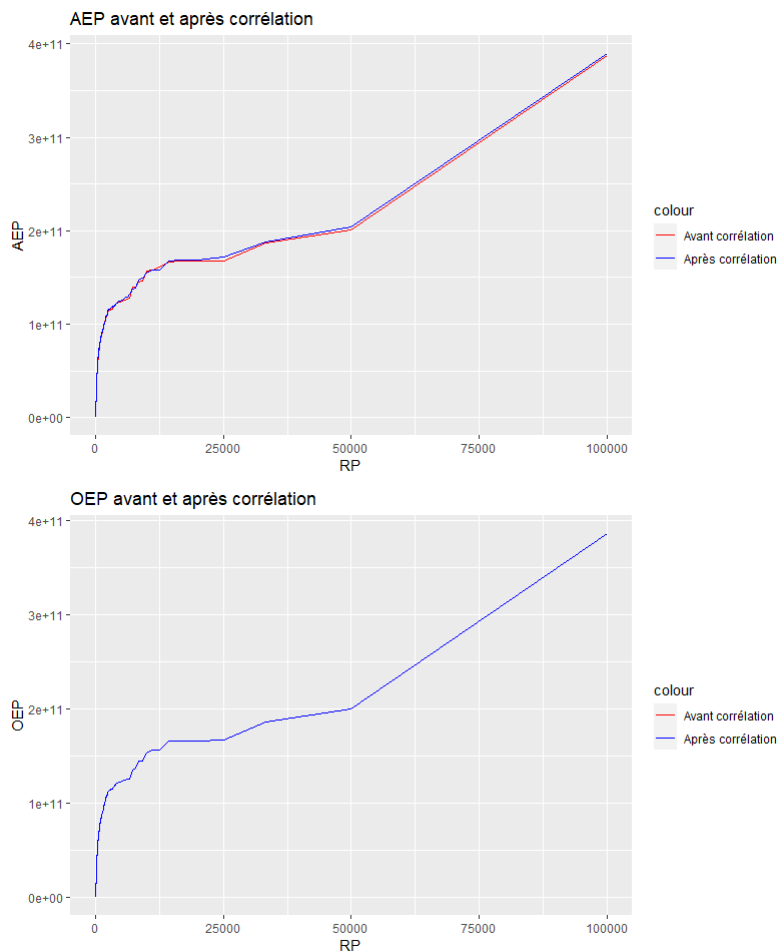
On obtient finalement deux nouveaux YLTs dont la distribution des pertes sur les années de simulation respecte la structure de dépendance d'une copule Franck de paramètre $a = 2.562154$.

Impact sur un contrat avant termes et conditions

Afin de visualiser l'impact d'introduire la dépendance entre ces deux périls TC et CS, on suppose qu'un contrat couvre les risques qui y sont liés dans la région du *South East/Mid Atlantic* aux États-Unis. On se sert des YLTs générés pour calculer la variation des mesures de risque.

On commence par agréger les pertes générées par les deux périls par année de simulation ensuite on calcule les périodes de retour pour chaque perte.

Suite à l'introduction de la dépendance, on n'observe pas de variation de l'AAL ni de l'OEP. Par contre, l'AEP présente des variations entre $[-13\%, 7\%]$ pour les différentes périodes de retour par rapport à l'AEP initiale. L'impact est donc visible sur la perte agrégée étant donné qu'après introduction de la dépendance estimée on peut introduire une simultanéité des pertes annuelles importantes ou l'inverse.



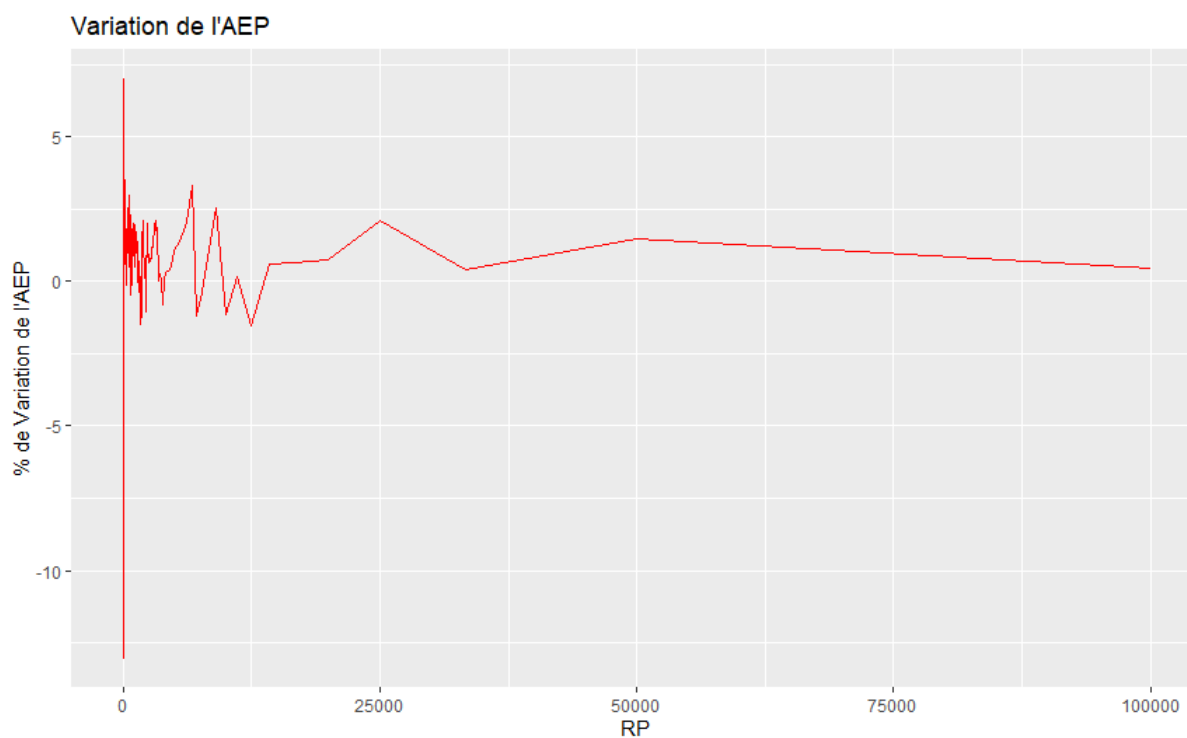


FIGURE 4.8 – Impact sur l’AEP et l’OEP après introduction de la dépendance entre les périls

Impact sur un contrat après termes et conditions

On introduit des termes et conditions pour la couverture de ces deux périls à savoir un contrat de type Cat XL : 59 000 000 000 XS 9 000 000 000 tel que la période de retour d’entrée est de 20 ans et celle de sortie s’élève à 500 années.

La perte moyenne de ce contrat augmente d’environ 6.7% en introduisant la dépendance. En effet, cette dépendance introduite a engendré plus d’occurrences simultanées de pertes importantes pour les deux périls dans la tranche couverte.

De plus, l’AEP de ce contrat après introduction de la dépendance entre les périls TC et CS a augmenté en moyenne de 16%.

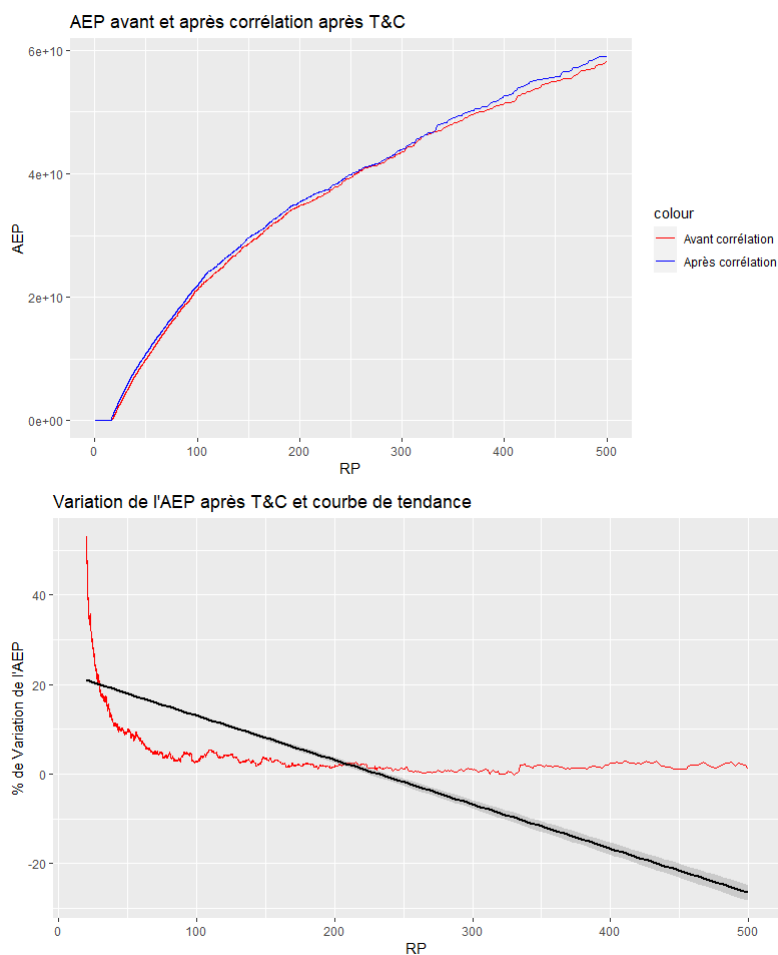


FIGURE 4.9 – Impact sur l'AEP après introduction de la dépendance entre les périls et après TC

Conclusion

L'hypothèse d'indépendance entre périls différents actuellement considérée dans la modélisation des risques catastrophes naturelles est une hypothèse très forte qui engendre une sous-estimation du risque. En effet, suite à notre exemple d'étude, la prise en considération de cette dépendance déduite des sinistres historiques augmente l'AEP d'un contrat et ainsi le risque.

La disponibilité de données historiques a limité cette recherche de dépendance entre d'autres *Region perils* dans d'autres zones géographiques, néanmoins la méthodologie reste pertinente et applicable éventuellement.

Conclusion

La modélisation proposée pour les *Region Perils* synthétiques chez SCOR permet de résoudre la problématique d'agrégation des pertes issues des *Region Perils* indépendamment de leur nature (modélisé ou synthétique).

Cette méthodologie déjà présente, a été renforcée par l'étude de la saisonnalité des *Region Perils* synthétiques étant donné son impact sur la vue de risque qui résulte en une amélioration de la qualité de la tarification et de l'accumulation.

Une étude de la dépendance observée historiquement entre périls différents dans des régions des États-Unis et grâce à la théorie des copules, a permis de voir l'impact de l'hypothèse d'indépendance entre *Region Perils* sur la tarification. La prise en considération de cette structure de dépendance permet ainsi d'éviter un cas de sous-estimation de risque.

Parmi les principaux défis de projet, on cite le caractère de faible fréquence des événements catastrophes naturelles. En effet, la disponibilité des données historiques a limité certains axes d'étude et a fait en sorte que la saisonnalité et la corrélation sont étudiées sur un niveau plus global que le niveau des *Region Perils*.

Les résultats obtenus dans ce mémoire sont fondés sur des données historiques. Cependant, nous sommes entrain d'étudier des catastrophes naturelles sujettes à de phénomènes climatiques fortement impactés par le changement climatique. On se demande ainsi : à quel point on peut se baser sur l'historique pour effectuer notre modélisation des catastrophes naturelles et sinon quelle serait l'alternative ?

Bibliographie

- [1] Article 1125-1 du code des assurances. https://www.legifrance.gouv.fr/codes/article_lc/LEGIARTI000018032948/. Version en vigueur au 18 novembre 2022.
- [2] Atlas magazine. <https://www.atlas-mag.net/>.
- [3] Center for research on the epidemiology of disasters (CRED). <https://www.cred.be/>.
- [4] Christophe Dutang Arthur Charpentier. *L'actuariat avec* : 2012.
- [5] Nicolas Baradel. *Assurance dommage*. 2022.
- [6] David Cadoux et Jean-Marc Loizeau. *Copules et dépendances : application pratique à la détermination du besoin en fonds propres d'un assureur non vie*.
- [7] Jean-David Fermanian. *Copules et applications financière*. Cours ENSAE, 2022.
- [8] Climate Change Infectious Diseases Group. *World maps of Koppen-Geiger Climate Classification*. 2019.
- [9] François GÉLAN. *Étude des procédés d'agrégation*. Mémoire d'actuariat, Institut des actuaires IA, 2015.
- [10] Marie Ait Mokhtar. *L'optimisation du portefeuille catastrophe naturelle face au défi des hypothèses*. Mémoire d'actuariat, Institut des actuaires IA, 2018.
- [11] SWISS RE. *Catastrophes naturelles et réassurance*. Peter Zimmerli, 2003.
- [12] SWISS RE. *Introduction à la réassurance*. 2003.
- [13] Gilbert Saporta. *Probabilités, analyse des données et statistique*. 2006.
- [14] Site web de SCOR. <https://www.scor.com/>.

Liste des abréviations

- **AAAL** : Annual Aggregated Loss
- **AEP** : Aggregate Exceedance Probability
- **cat-nat** : Catastrophes naturelles
- **EL** : Expected Loss
- **ELT** : Event Loss Table
- **LoL** : Loss on Line
- **L XS P** : Excédent de siniste de limite L et de priorité P.
- **OEP** : Occurrence Exceedance Probability
- **T&C** : Termes et conditions
- **YEQT** : Year Event Quantile Table
- **YLT** : Year Loss Table

Annexe A

Preuve du lemme 1

Lemme 2. Soit N une variable aléatoire de loi $\mathcal{P}(\lambda)$ avec $\lambda > 0$ et D une variable aléatoire telle que $D \mid N \sim \mathcal{B}(N, p)$ avec $p > 0$. Alors

$$N - D \sim \mathcal{P}(\lambda(1 - p)).$$

[5]

Démonstration. Soit $k \in \mathbb{N}$.

$$\begin{aligned} \mathbb{P}(N - D = k) &= \mathbb{P}\left(\{N - D = k\} \cap \left\{\bigcup_{n \in \mathbb{N}} \{N = n\}\right\}\right) \\ &= \sum_{n \in \mathbb{N}} \mathbb{P}(\{N - D = k\} \cap \{N = n\}) \\ &= \sum_{n \in \mathbb{N}} \mathbb{P}(\{D = n - k\} \mid \{N = n\}) \mathbb{P}(N = n) \\ &= \sum_{n \geq k} \frac{n!}{k!(n - k)!} p^{n-k} (1 - p)^k \frac{e^{-\lambda} \lambda^n}{n!} \\ &= \frac{e^{-\lambda} (1 - p)^k}{k!} \sum_{n \geq 0} \frac{p^n \lambda^{n+k}}{n!} \\ &= \frac{e^{-\lambda} [\lambda(1 - p)]^k}{k!} \sum_{n \geq 0} \frac{(p\lambda)^n}{n!} \\ &= \frac{e^{-\lambda(1-p)} [\lambda(1 - p)]^k}{k!}. \end{aligned}$$

□

Annexe B

Table des valeurs critiques du coefficient de corrélation de rangs de Spearman

Soit n la taille de l'échantillon d'observations et α le seuil.

Dans ce tableau, on regroupe les valeurs de r tels que $\mathbb{P}(|\rho_S| > r) = \alpha$.

n	Nominal α					
	0.10	0.05	0.025	0.01	0.005	0.001
4	1.000	1.000	—	—	—	—
5	0.800	0.900	1.000	1.000	—	—
6	0.657	0.829	0.886	0.943	1.000	—
7	0.571	0.714	0.786	0.893	0.929	1.000
8	0.524	0.643	0.738	0.833	0.881	0.952
9	0.483	0.600	0.700	0.783	0.833	0.917
10	0.455	0.564	0.648	0.745	0.794	0.879
11	0.427	0.536	0.618	0.709	0.755	0.845
12	0.406	0.503	0.587	0.678	0.727	0.818
13	0.385	0.484	0.560	0.648	0.703	0.791
14	0.367	0.464	0.538	0.626	0.679	0.771
15	0.354	0.446	0.521	0.604	0.654	0.750
16	0.341	0.429	0.503	0.582	0.635	0.729
17	0.328	0.414	0.488	0.566	0.618	0.711
18	0.317	0.401	0.472	0.550	0.600	0.692
19	0.309	0.391	0.460	0.535	0.584	0.675
20	0.299	0.380	0.447	0.522	0.570	0.662
21	0.292	0.370	0.436	0.509	0.556	0.647
22	0.284	0.361	0.425	0.497	0.544	0.633
23	0.278	0.353	0.416	0.486	0.532	0.621
24	0.271	0.344	0.407	0.476	0.521	0.609
25	0.265	0.337	0.398	0.466	0.511	0.597
26	0.259	0.331	0.390	0.457	0.501	0.586
27	0.255	0.324	0.383	0.449	0.492	0.576
28	0.250	0.318	0.375	0.441	0.483	0.567
29	0.245	0.312	0.368	0.433	0.475	0.558

	Nominal α					
30	0.240	0.306	0.362	0.425	0.467	0.549
31	0.236	0.301	0.356	0.419	0.459	0.540
32	0.232	0.296	0.350	0.412	0.452	0.532
33	0.229	0.291	0.345	0.405	0.446	0.525
34	0.225	0.287	0.340	0.400	0.439	0.517
35	0.222	0.283	0.335	0.394	0.433	0.510
36	0.219	0.279	0.330	0.388	0.427	0.503
37	0.215	0.275	0.325	0.383	0.421	0.497
38	0.212	0.271	0.321	0.378	0.415	0.491
39	0.210	0.267	0.317	0.373	0.410	0.485
40	0.207	0.264	0.313	0.368	0.405	0.479
41	0.204	0.261	0.309	0.364	0.400	0.473
42	0.202	0.257	0.305	0.359	0.396	0.468
43	0.199	0.254	0.301	0.355	0.391	0.462
44	0.197	0.251	0.298	0.351	0.386	0.457
45	0.194	0.248	0.294	0.347	0.382	0.452
46	0.192	0.246	0.291	0.343	0.378	0.448
47	0.190	0.243	0.288	0.340	0.374	0.443
48	0.188	0.240	0.285	0.336	0.370	0.439
49	0.186	0.238	0.282	0.333	0.366	0.434
50	0.184	0.235	0.279	0.329	0.363	0.430
51	0.182	0.233	0.276	0.326	0.359	0.426
52	0.180	0.231	0.274	0.323	0.356	0.422
53	0.179	0.228	0.271	0.320	0.352	0.418
54	0.177	0.226	0.268	0.317	0.349	0.414
55	0.175	0.224	0.266	0.314	0.346	0.411
56	0.174	0.222	0.264	0.311	0.343	0.407
57	0.172	0.220	0.261	0.308	0.340	0.404
58	0.171	0.218	0.259	0.306	0.337	0.400
59	0.169	0.216	0.257	0.303	0.334	0.397
60	0.168	0.214	0.255	0.301	0.331	0.394

Annexe C

Autres méthodes d'estimation de copules

Méthode des moments

Cette méthode est basée sur une mesure de dépendance telle que le Rho de Spearman ou le tau de Kendall et à l'estimation de la valeur du paramètre a de la copule C qui égalise la valeur théorique de cette mesure et la valeur estimée.

Elle n'assure aucune robustesse de l'estimateur. Dans la pratique, l'estimateur empirique du tau de Kendall est le plus souvent utilisé comme mesure de dépendance en raison de sa simplicité de calcul.

Inférence sur les marginales

La méthode IFM ou méthode des fonctions d'inférence des marginales repose sur le fait que la modélisation avec les copules permet de séparer les paramètres spécifiques des distributions marginales de ceux de la structure de dépendance. Ainsi, la log-vraisemblance peut s'écrire de la manière suivante :

$$l(\theta) = \sum_{t=1}^n \ln c(F_1(x_1^t, \theta_1), \dots, F_d(x_d^t, \theta_d); a) + \sum_{t=1}^n \sum_{k=1}^d \ln f_k(x_k^t; \theta_k)$$

où $\theta = (\theta_1, \dots, \theta_d, a)$ le vecteur des paramètres contient les paramètres θ_k de chaque marginale et les paramètres a de la copule.

La maximisation s'effectue sur deux étapes : on estime les paramètres de chaque marginale

$$\begin{aligned} \hat{\theta}_k &= \arg \max l_k(\theta_k) \\ &= \arg \max \sum_{t=1}^d \ln f_k(x_k^t; \theta_k) \end{aligned}$$

on estime a à partir des estimateurs précédents

$$\begin{aligned} \hat{a} &= \arg \max l_c(a) \\ &= \arg \max \sum_{t=1}^d \ln c(F_1(x_1^t, \hat{\theta}_1), \dots, F_n(x_n^t, \hat{\theta}_n); a). \end{aligned}$$

L'estimateur $\hat{\theta}_{IFM} = (\hat{\theta}_1, \dots, \hat{\theta}_d, \hat{a})$ vérifie la propriété de normalité asymptotique et on a $\sqrt{T}(\hat{\theta}_{IFM} - \theta_0) \rightarrow N(0, v^{-1}(\theta_0))$ avec $v(\theta_0)$ la matrice d'information de Godambe. Si on

définit $g(\theta) = \left(\frac{\partial}{\partial \theta_1} l_1, \dots, \frac{\partial}{\partial \theta_d} l_d \right)$ alors la matrice de Godambe s'écrit :

$$v(\theta_0) = D^{-1} M (D^{-1})^T \text{ où } D = E \left[\frac{\partial}{\partial \theta} (g(\theta)^T) \right] \text{ et } M = E [g(\theta)^T g(\theta)] .$$

Cette méthode peut avoir l'avantage d'être basée sur des calculs plus légers que ceux générés par la méthode du maximum de vraisemblance, mais la détermination de la matrice de Godambe peut être très difficile en raison des multiples calculs de dérivées. Ainsi, elle est sensible à une définition incorrecte des marges pour la même raison que celle évoquée ci-dessus.